



HAL
open science

DeepStim Project. Modeling states of consciousness and their modulation by electrical Deep Brain Stimulation : from experimental data to computational models

Chloé Gomez

► To cite this version:

Chloé Gomez. DeepStim Project. Modeling states of consciousness and their modulation by electrical Deep Brain Stimulation : from experimental data to computational models. Medical Imaging. Université Paris-Saclay, 2024. English. NNT : 2024UPASL027 . tel-04736047

HAL Id: tel-04736047

<https://theses.hal.science/tel-04736047v1>

Submitted on 14 Oct 2024

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

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

DeepStim Project. Modeling states of
consciousness and their modulation by
electrical Deep Brain Stimulation: from
experimental data to computational
models

*Projet DeepStim. Modélisation des états de conscience et
de leur modulation par la stimulation cérébrale profonde :
des données expérimentales aux modèles computationnels*

Thèse de doctorat de l'université Paris-Saclay

École doctorale n°568 Signalisations et réseaux intégratifs en biologie
(BioSigne)
Spécialité de doctorat: Sciences de la vie et de la santé
Graduate School : Sciences de la vie et santé, Référent : Faculté de
médecine

Thèse préparée dans les unités de recherche **Neuroimagerie cognitive**
(Université Paris-Saclay, Inserm, CEA) et **BAOBAB** (Université Paris-Saclay,
CEA, CNRS), sous la direction de **Béchir JARRAYA**, professeur des
universités-praticien hospitalier (PU-PH), la co-direction de **Antoine GRIGIS**,
ingénieur de recherche

Thèse soutenue à Paris-Saclay, le 24 mai 2024, par

Chloé GOMEZ

Composition du jury

M. Demian BATTAGLIA Senior researcher, Université de Strasbourg	Président & Rapporteur
M. Jacobo SITT Directeur de recherche, Paris Brain Institute (ICM)	Rapporteur & Examineur
Mme Joana CABRAL Group Leader, University of Minho, Portugal	Examinatrice
M. Gustavo DECO Full professor, University Pompeu Fabra, Spain	Examineur

Titre: Projet DeepStim. Modélisation des états de conscience et de leur modulation par la stimulation cérébrale profonde : des données expérimentales aux modèles computationnels

Mots clés: Apprentissage profond ; IRM fonctionnelle ; Troubles de la conscience ; Connectivité dynamique fonctionnelle ; Stimulation cérébrale profonde ; Réseau de neurones artificiels

Résumé: Le diagnostic des patients dans le coma est souvent difficile. Les examens cérébraux renseignent les médecins sur l'étendue des lésions cérébrales mais ne permettent pas de déterminer avec précision l'état de conscience du patient. De plus, aucune approche thérapeutique ne permet une restauration systématique de la conscience. Des études pionnières menées sur des patients et des Primates Non Humains (PNH) ont montré que la Stimulation Cérébrale Profonde (SCP) des noyaux intralaminaires du thalamus pouvait restaurer ou améliorer la conscience lorsqu'elle est altérée. Cependant, les conséquences corticales associées à la SCP restent largement inconnues et imprévisibles. Les techniques d'imagerie fonctionnelle, telles que l'Imagerie par Résonance Magnétique fonctionnelle de repos (IRMf de repos), peuvent aider à identifier des signatures de la conscience. L'activité cérébrale au repos, organisée en réseaux, peut être modélisée à l'aide de la connectivité fonctionnelle. Cette thèse vise à disséquer, à l'aide du modèle PNH, les effets sur la connectivité fonctionnelle d'une modulation de la conscience induite par des agents anesthésiques ou de la SCP à l'échelle du cerveau entier. Cela nécessite le développement de modèles interprétables et prédictifs des effets d'une telle modulation sur la fonction cérébrale globale. Pour identifier les schémas récurrents dominants (c'est-à-dire les différents états du cerveau) à partir de la connectivité fonctionnelle, une technique d'apprentissage automatique non supervisée (K-Means) a été proposée précédemment. Dans le cadre de cette thèse, nous développons de nouveaux outils d'analyse en tirant parti des avancées des techniques d'apprentissage profond auto-supervisé. Nous émettons l'hypothèse que l'identification de variables latentes dans les signaux IRMf de repos peut nous informer sur la

modulation des états de conscience. Tout d'abord, nous cherchons à identifier une signature spatiale, moyennée temporellement, de la conscience à la fois dans l'état éveillé et sous anesthésie. Nous utilisons une méthode de variables latentes qui décompose les signaux IRMf de repos en réseaux fonctionnels associés à l'accès conscient. Afin d'étudier la restauration de la conscience, nous étendons cette analyse aux PNH éveillés ou réveillés par DBS du thalamus central. Notre modèle suggère de manière automatique que le cortex antérieur et le cortex postérieur contribuent tous deux à la conscience, un sujet qui fait débat au sein de la communauté scientifique. En outre, il souligne l'importance des régions clés au sein de l'espace de travail neuronal global, une théorie importante concernant l'accès à la conscience. Suite à cette analyse moyennée temporellement, reconnaissant l'importance de la dynamique temporelle dans l'analyse de la conscience, nous proposons de remettre en question les méthodes conventionnelles de connectivité fonctionnelle dynamique. Nous utilisons un modèle d'apprentissage profond contrastif pour prédire les schémas cérébraux caractéristiques de différents états de conscience. Les expériences démontrent que les prédictions du modèle basées sur la connectivité fonctionnelle dynamique mettent en avant des transitions entre les schémas cérébraux. Enfin, pour mieux comprendre la dynamique des états de conscience, nous nous écartons du cadre conventionnel de classification en sous-groupes et introduisons une méthode de réduction de dimensions. Cette approche vise à condenser ces états en un nombre limité de variables interprétables et explicables. Nos résultats indiquent que l'approche catégorielle traditionnelle ne permet pas de saisir de manière adéquate le continuum de la dynamique des états de conscience.

Title: DeepStim Project. Modeling states of consciousness and their modulation by electrical Deep Brain Stimulation: from experimental data to computational models

Keywords: Deep learning ; Functional MRI ; Disorders of consciousness ; Dynamic functional connectivity ; Deep brain stimulation ; Artificial neural network

Abstract: Diagnosis of patients with coma is often difficult. Brain examinations inform physicians about the extent of brain damage but do not accurately determine the patient's level of consciousness. Moreover, no therapeutic approach allows a systematic restoration of consciousness. Pioneering studies in patients and Non-Human Primates (NHP) have shown that Deep Brain Stimulation (DBS) of the intralaminar nuclei of the thalamus could restore or improve consciousness when it is impaired. However, the cortical consequences associated with DBS remain largely unknown and unpredictable. Functional imaging techniques, such as Resting-State functional Magnetic Resonance Imaging (RS-fMRI), can help identify signatures of consciousness. Brain activity at rest, organized into networks, can be modeled using functional connectivity. This thesis aims to dissect, using the NHP model, the effects on functional connectivity of a modulation of consciousness induced by anesthetic agents or DBS on a whole-brain scale. This requires the development of interpretable and predictive models of the effects of such modulation on global brain function. To identify dominant recurrent patterns (i.e., different brain states) from functional connectivity, an unsupervised machine learning technique (K-Means) has been previously proposed. As part of this thesis, we develop new analysis tools by taking advantage of the advances in self-supervised deep learning techniques. We hypothesized that identifying latent variables in RS-fMRI signals can inform us about the modulation of states of consciousness. First, we aim to identify a time-averaged spatial signa-

ture of consciousness in both the awake state and under anesthesia. This is achieved through a latent variables method that decomposes resting-state fMRI signals based on functional networks associated with conscious access. In a translational effort to investigate consciousness restoration, we extend this analysis to awake or awakened NHPs by DBS of the central thalamus. Our model autonomously suggests that both the anterior and posterior cortex contribute to consciousness, a debatable topic in the scientific community. Additionally, it underscores the significance of key regions within the global neuronal workspace, a prominent theory regarding conscious access. Following this time-averaged analysis, recognizing the critical importance of temporal integration in consciousness analysis, we propose to challenge conventional dynamic functional connectivity methods. We employ a contrastive deep learning model to predict brain patterns characteristic of various consciousness states. Experiments demonstrate that the model predictions based on dynamic functional connectivity facilitate the examination of different transient brain states. Lastly, to gain a deeper understanding of the dynamics of consciousness states, we diverge from the conventional subgroup classification framework and introduce a dimension-reduction method. This approach aims to condense these states into a limited number of interpretable and explicable variables. Our findings indicate that the traditional categorical approach inadequately captures the continuum of consciousness state dynamics.

N'oublie jamais, celui qui croit savoir n'apprend plus.

Pierre Bottero

*La Nature est un temple où de vivants piliers
Laissent parfois sortir de confuses paroles ;
L'homme y passe à travers des forêts de symboles
Qui l'observent avec des regards familiers.*

*Comme de longs échos qui de loin se confondent
Dans une ténébreuse et profonde unité,
Vaste comme la nuit et comme la clarté,
Les parfums, les couleurs et les sons se répondent.*

Extrait de "Correspondances", des *Fleurs du Mal*,
Charles Baudelaire

Acknowledgements

Il y a trois ans et demi, j'ai embarqué dans l'aventure qu'est la thèse sans trop savoir où je mettais les pieds, mais puisqu'on m'avait parlé de voyage solitaire, je m'étais imaginée la Solitaire du Figaro, la fameuse course de voiliers en solo, sans assistance. Aujourd'hui, l'aventure se termine, et j'ai dû me tromper de ponton au départ parce que je n'ai jamais navigué seule. J'ai eu l'immense plaisir d'embarquer avec un équipage formidable et sans lui, je ne serais certainement pas parvenue sur la ligne d'arrivée. Je tiens d'abord à remercier, le chef de bord, mon directeur de thèse, Béchir de m'avoir laissé monter à bord. Tu m'as fait confiance pour travailler sur ce projet qui me tenait beaucoup à cœur. Je sais que ça n'a pas toujours été simple de faire naviguer ce bateau dans la direction que tu souhaitais au départ, nous étions une équipe avec des intérêts différents et concilier les envies et les affinités de chacun n'était pas une mince affaire. Merci de m'avoir confié la barre. Évidemment, cette aventure n'aurait pas été possible sans le bras droit du capitaine, mon encadrant, Antoine. Sur un bateau, le second capitaine est chargé de la stabilité et de la sécurité du navire. Antoine, tu as rempli ces rôles à la perfection. Ton soutien a été sans faille. Un sale temps, nommé COVID, s'est abattu pendant une longue partie de la traversée. Il nous a forcé à changer nos méthodes de travail, à réduire les interactions et malgré ça, jamais, je ne me suis sentie seule à bord. Merci pour les appels réguliers. Merci pour les points hebdomadaires, pour les relectures, les re-relectures, pour les corrections, les réécritures. Pour les réponses aux questions, parfois toujours les mêmes, pour ta patience. Ton investissement sur le projet m'a toujours fascinée, j'ai eu l'impression que tu n'étais que sur mon bateau alors que tu faisais la traversée avec plein d'autres en parallèle. Tu m'as redonné confiance quand celle-ci faiblissait, tu n'as jamais abandonné.

Marianne, ma mentor du programme Femmes et Science, merci de m'avoir rappelé qu'il faut d'abord s'écouter et qu'on n'a rien à perdre en essayant. À l'époque, quand on me demandait ce que je voulais faire plus tard, je disais que je voulais être chercheur. Pas chercheuse. Chercheur, comme un homme, chercheur pour de vrai, pas comme une fille. Merci de m'avoir montré qu'on n'avait pas à rougir de vouloir être chercheuse. Merci à tous les autres PIs qui m'ont accompagnée, pour leurs conseils avisés tout au long de l'aventure : Vincent, Cathy, Edouard, Jean-François, Guylaine, Rodrigo. Merci également à mon CSI, Gustavo Deco et Alexandre Gramfort de m'avoir permis de garder un cap. Merci à mon jury, Damian Battaglia, Jacobo Sitt, Joana Cabral et Gustavo Deco d'avoir accepté de relire ma thèse et d'assister à ma soutenance. Merci pour vos questions, vos remarques, elles ont considérablement agrandi l'horizon des possibles et donnent envie de continuer la traversée.

Vient la partie la plus complexe. J'ai du mal avec les catégories, les petites boîtes bien rangées, les étiquettes et les frontières arbitraires qu'on définit. Alors en vrac, merci à tout l'équipage, les collègues de Neurospin (dont les stagiaires), les anciens collègues, la team colombia. La plupart d'entre vous portent plusieurs étiquettes et c'est bien plus rigolo comme ça. Je ne vais pas citer ce que chacun d'entre vous m'a apporté, sachez juste, que vous êtes mes ailes, vous avez rendu cette traversée si légère, vous avez transformé ce bateau en foyer. Notre engouement mutuel pour le sport (l'escalade, la piscine, la rando, le ski, les danses latines, et évidemment le lever de coude) a scellé une amitié indéfectible. Quelles que soient mes destinations futures, vous ferez toujours partie du voyage. Merci à mes amies d'enfance et de prépa (jusqu'à quel âge est-on enfant ?) Marie, Virginie, Camille, Xaxou, Pauline. Je vous compte sur les doigts d'une main, mais c'est une main qui m'est sacrément indispensable. Merci à mes ami.e.s d'école d'ingénieur

d'être venues des quatre coins de France pour me soutenir le jour J. Vous êtes mes phares, loin des yeux, près du cœur, toujours. Merci à ma coloc, Agathe, d'être entrée dans ma vie comme une bourrasque, de m'avoir ouvert les yeux sur une autre façon de vivre sa vie et d'être toujours partante pour tout, n'importe quand, surtout n'importe quand ! A veces algunos se quedan en tierra firme mientras otros continúan la aventura. Cristobal, navegaste a mi lado durante mucho tiempo, demasiado para que olvide tu apoyo. A pesar del mareo, gracias por venir a bordo, tan lejos de casa, y por estar ahí cada vez que te necesito.

Merci enfin, à ma famille. À mes oncles et tantes, à mes cousins cousines, à mes grands-parents, fidèles supporters qui s'intéressent toujours à ce que je fais, qui me posent mille questions auxquelles je n'ai souvent pas les réponses. À mes parents, merci, de m'avoir laissée faire mes choix, de m'avoir offert cette possibilité folle d'explorer tous les horizons. Ne n'avoir jamais réfréné mes envies, même quand la peur a dû vous assaillir. Quand de mon pays basque presque natal, j'ai tracé la plus grande des diagonales en France et que je suis partie à Strasbourg, puis à Berlin, puis au Chili, sans billet retour. Merci de m'avoir soutenue moralement, de m'avoir encouragée même si vous ne compreniez pas tout à fait le titre de ma thèse, et de m'avoir accueillie à la maison toutes les fois où j'en ai eu besoin. Papa, ta détermination, cette capacité folle de te réinventer, de suivre tes rêves, tes projets qu'ils soient professionnels ou personnels est un puits d'inspiration sans fin. Maman, je n'aurais probablement jamais écrit cette thèse aujourd'hui si tu ne m'avais pas insufflé le goût de la lecture, et celui de l'écriture. Celui aussi, du travail acharné, parce que finalement, quand on aime, on ne compte pas. Vous m'avez appris la résilience, la persévérance, à se relever toujours et à avancer vers ce en quoi l'on croit. A ma petite sœur, enfin, qui a parcouru bien plus de milles que moi dans la vraie vie (et écrit bien plus d'articles aussi !!), dont j'admire les choix audacieux, qui m'inspire à faire bouger les lignes et sur qui je peux compter à chaque étape de ma vie.

Enfin, un mot à mes professeurs, ceux qui y ont cru plus que moi, qui m'ont poussé à continuer même quand cela faisait longtemps que j'avais quitté leur classe. Ceux qui m'ont donné envie de chercher, d'apprendre, de savoir et de douter, sans arrêt.

Contents

List of abbreviations	1
Introduction	5
I Background - The crystal consciousness	7
1 To be or not to be... conscious	9
1.1 Mirror mirror, on the wall, consciousness...	9
1.1.1 ... what is it ?	9
1.1.2 ... how to study it ?	13
1.1.3 ... why studying it ?	25
1.2 Losing consciousness	27
1.2.1 Disorders of consciousness (DoCs)	27
1.2.2 Anesthesia-induced loss of consciousness	32
1.3 Restoring consciousness: from science fiction to science	35
1.3.1 Pharmacological treatments	35
1.3.2 Neurostimulation to restore consciousness	36
1.3.3 Challenges	41
2 fMRI: the brain's cloak of visibility	45
2.1 Theory	45
2.1.1 Magnetic Resonance Imaging : from magnet to image	46
2.1.2 Functional MRI (fMRI)	47
2.1.3 fMRI parameters	48
2.1.4 Contrast agents in fMRI	50
2.1.5 Resting-state fMRI	51
2.1.6 Relation with neuronal activity	52
2.1.7 Benefits, drawbacks and comparison	52
2.2 Datasets	54
2.2.1 Anesthesia dataset	54
2.2.2 DBS dataset	57
2.2.3 External resources: ROI template & reference anatomical connectivity	62
2.2.4 Functional connectivity (FC)	63
2.3 Limits and associated challenges	67
2.3.1 The limits of acquisition	68
2.3.2 The limits of preprocessing	69
2.3.3 The limits of connectivity computation	71

Motivations and aims of the thesis	73
------------------------------------	----

II Seeing patterns, deciphering messages: the spatial signatures of consciousness **77**

3 Functional connectivity at rest: studying conscious networks	79
3.1 Introduction	80
3.2 Related works to identify stationary networks	81
3.2.1 Multivariate decomposition	82
3.2.2 Graph	86
3.3 Material and Method	88
3.3.1 Datasets and atlas choice	88
3.3.2 The Modular Hierarchical Analysis (MHA)	89
3.3.3 Decoding Brain Network Activities (BNAs)	91
3.4 Results on Anesthesia Dataset	92
3.4.1 Consciousness connectivity can be decomposed into few consistent brain networks	92
3.4.2 Brain network 1 intersects the Global Neuronal Workspace (GNW)	93
3.4.3 Which network best predicts the depth of anesthesia from BNAs?	94
3.4.4 Influence of time window size on predictions: sensitivity study	96
3.4.5 Conclusion	96
3.5 Results on DBS dataset	97
3.5.1 Consciousness connectivity can be decomposed into few consistent brain networks	97
3.5.2 Identified networks	97
3.5.3 Reconciling the front and back of the brain in the processing of conscious information	98
3.5.4 Networks capture differences induced by the effective DBS condition	98
3.5.5 Conclusion	98
3.6 Discussion	99

III Seeing beyond reflection: latent variable models for studying states of consciousness **107**

4 From networks to dynamic functional connectivity analysis	109
4.1 Introduction	109
4.1.1 Material and Method	110
4.1.2 Application to consciousness : state-of-the-art	111
4.1.3 Overcoming the limitations of K-means clustering	114
4.2 Material and Method	118
4.2.1 Dataset and data partitioning	118
4.2.2 Model architecture	118
4.2.3 Maps of Predictive Connections	120
4.3 Results on Anesthesia dataset	120

4.3.1	Performance of the classifier	120
4.3.2	Towards modeling the brain patterns dynamic	123
4.3.3	Maps of predictive connections	124
4.4	Conclusion	126
4.5	Discussion	126
5	Unravelling brain dynamics: deciphering brain states	129
5.1	Introduction	130
5.2	Material and method	131
5.2.1	Dataset	131
5.2.2	Low-dimensional generative models	131
5.2.3	Model evaluation	134
5.2.4	Latent space exploration	135
5.2.5	Connection-wise simulations	136
5.3	Results on Anesthesia dataset	138
5.3.1	Model evaluation	138
5.3.2	Latent space exploration	140
5.3.3	Connection-wise simulations	142
5.4	Discussion	144
	Conclusion and Perspectives	149
	Appendix	171
	Bibliography	173

List of abbreviations

- AI** Artificial Intelligence. 18
- ARAS** Ascending Reticular Activating System. 43
- AUC** Area Under the Curve. 123
- BN** Brain Network. 156
- BNA** Brain Network Activities. 89–95, 98–100
- BOLD** Blood Oxygenation Level Dependent. 47–52
- BP** Brain Pattern. 110, 111, 113, 134, 150
- CBV** Cerebral Blood Volume. 47, 50
- CM** Centro-Median. 59
- CoCoMac** . 62–64, 72, 88, 92–95, 97, 99–101
- CRS-R** Coma Recovery Scale-Revised. 28
- CT** Central Thalamus. 39, 41, 57, 80
- DAN** Dorsal Attentional Network. 102
- DBS** Deep Brain Stimulation. 5, 6, 36–39, 41–43, 54, 57–62, 64, 66–69, 71, 73, 75, 148
- dFC** Dynamic Functional Connectivity. 63, 64, 75, 106, 110, 126, 132
- DL** Dictionary Learning. 82, 85
- DMN** Default Mode Network. 80
- DoC** Disorder of Consciousness. 11, 25, 27–31, 35–42, 113
- EEG** Electroencephalography. 22, 25, 28, 29, 34, 41, 53–55, 58, 59, 65–67
- EPI** Echo Planar Imaging. 48, 61
- FC** Functional Connectivity. 63
- FEF** Frontal Eye Field. 94, 102
- FID** Free Induction Decay. 49

fMRI Functional Magnetic Resonance Imaging. 3, 5, 6, 29–31, 45–53, 56–59, 61, 62, 66, 69, 71

GABA Gamma-Amino-Butyric-Acid. 32

GCS Glasgow Coma Scale. 28

GNW Global Neuronal Workspace. 19, 20, 96, 149

GWT Global Workspace Theory. 18, 20, 21

HRF Hemodynamic Response Function. 48, 68

IC Independent Component. 101

ICA Independent Component Analysis. 82, 88, 89, 101

IIT Integrated Information Theory. 18, 20–22

IPG Implantable Pulse Generator. 38

KL Kulback-Leibler. 131

LOC Loss of consciousness. 32, 33, 43

MCS Minimally Conscious State. 27–30, 40, 41, 43, 101

MD Mean Diffusivity. 143, 144

MEG Magnetoencephalography. 53, 54

MHA Modular Hierarchical Analysis. 8, 79, 81, 88–90, 93, 97, 98, 100

MION Monocrystalline Iron Oxide Nanoparticles. 50, 51, 56, 68

MLP Multi-Layer Perceptron. 130

MNI Montreal Neurologic Institute. 57, 62

MPRAGE Magnetization Prepared-Rapid Gradient Echo. 59

MRI Magnetic Resonance Imaging. 5, 46, 47, 59–61, 64, 66, 71, 72

MSE Mean Squared Error. 131, 132, 134

NCC Neural Correlates of Consciousness. 21

NHP Non-Human Primate. 5, 6, 31, 32, 54, 55, 75, 81, 85

NMR Nuclear Magnetic Resonance. 49

NREM Non Rapid Eye Movements. 24, 25

NSM NeuroSpin Monkey. 57

PCA Principal Component Analysis. 83, 88, 89

PET Positron emission tomography. 30, 53

PFC Pre-Frontal Cortex. 17, 20, 23, 24

PPCA Probabilistic PCA. 133, 134, 138

Prime-DE Primate neuroimaging Data-Exchange. 71, 104

QC Quality Control. 71

RBF Radial Basis Function. 91

REM Rapid Eye Movements. 3, 24, 25, 114

RF Receptive Field. 136, 137, 143, 145, 148

ROC Receiver Operating Characteristic. 122, 123

ROI Regions of Interest. 62, 118

RS-fMRI Resting-State fMRI. 51, 54, 57, 61, 63, 80, 81, 87, 88, 99, 101

RSN Resting State Network. 51, 80, 99

SN Salience Network. 102, 103

SNR Signal-to-Noise Ratio. 68

SSIM Structural SIMilarity. 134, 135, 138, 139

SVC Support Vector Classifier. 121, 126, 150

SVM Support Vector Machines. 91

TBI Traumatic Brain Injury. 35, 40, 41

tDCS transcranial Direct Current Stimulation. 36, 80

TE Echo Time. 48, 49

TI Inversion Time. 59

TMS Transcranial Magnetic Stimulation. 21, 36

TR Repetition Time. 48, 49, 68

USPIO Ultrasmall Superparamagnetic Iron Oxide. 50

UWS Unresponsive Wakefulness Syndrome. 27–31, 40, 41, 43, 101

VAE Variational AutoEncoder. 130

VL VentroLateral. 57, 59

VS Vegetative State. 27

WCSS Within-Cluster Sum of Squared. 114

Introduction

Electrical brain stimulation stands as a pivotal technology in neuroscience. Within fundamental neuroscience, precisely targeted electrical stimulation of specific brain regions aids in defining their functions and elucidating the causal relationship between neural activity and behavior. In translational and clinical neuroscience, Deep Brain Stimulation (DBS) has revolutionized the treatment landscape for Parkinson's disease, primarily through insights garnered from Non-Human Primate (NHP) models. Furthermore, DBS holds promise for addressing various neurological and psychiatric disorders, including obsessive-compulsive disorder, severe depression, and, particularly at stake in this thesis, disorders of consciousness. However, despite these advancements, patient care and identifying novel therapeutic targets remain largely empirical due to the absence of a predictive model of the neuronal effects of DBS within specific brain regions. Indeed, the neural mechanisms underlying DBS and its broader implications on neural circuits throughout the brain remain largely uncharted territory and lack predictability.

Advances in neuroimaging, particularly the development of Functional Magnetic Resonance Imaging (fMRI), provide vital information on brain activity, both during a task and at rest. Being able to "see the brain think", as Denis Le Bihan wrote [19], is a way to make the global consequences of neuromodulation on the brain visible. In his book entitled "Le Cerveau de Cristal" ("The Crystal Brain"), subtitled "Ce que nous révèle la neuro-imagerie" ("What neuroimaging reveals") [19], Le Bihan describes how MRI makes this once impenetrable brain, transparent. Neuroimaging makes it possible to observe, dissect and analyze the brain without opening it up. Le Bihan probably didn't see the glass half-empty when he used the crystal metaphor to transform the brain, 80% of whose mass comes from water, into a glass of water. Finally, to conclude his book, Le Bihan takes us back to one of the major research areas for the 21st century: digital brain simulation. Projects such as the Human Brain Project and the Blue Brain Project have led to considerable advances in brain modeling. At the heart of these "in silico" experiments is silicon, an essential component of glass and our computers. The glass metaphor is complete!

Yet, ten years after the initial reading of this book, one question remains. Le Bihan did not choose to title his book "The Glass Brain". I dare to take up his image for myself and add the magical, esoteric, divinatory dimension I've always seen in it. Is the brain, that organ that is still so little understood, a crystal ball that can be interrogated? A crystal ball that would dictate our future, allow us to see our past, our emotions, our thoughts? Would it reveal our consciousness, that mysterious object that seems to emanate from this well-protected organ? Perhaps even give us access to our unconscious?

The DeepStim project endeavors to investigate, employing the NHP model, the impact of consciousness modulation induced by anesthesia and DBS on a whole-brain scale. To achieve this goal, we propose several computational models designed to analyze and forecast various states of consciousness and the global cerebral repercussions of consciousness modulation. The project objective is to delineate the cerebral implications, utilizing fMRI mapping, resulting from consciousness modulation. This seeks to enhance our comprehension of the alterations in local, deep, and global cerebral functional connectivity arising from anesthesia and DBS. For that, we will employ latent variable models for predicting the overall cerebral implications of anesthesia and DBS. This innovative approach to therapeutic brain stimulation, grounded in global brain modeling, holds the potential to rationalize the targets of brain stimulation and ascertain the cortical networks influenced by DBS, thereby laying the groundwork for personalized medicine.

This thesis is organized into three main parts.

- In part I, we introduce this metaphorical crystal ball with translucent surfaces, emitting elusive wisps of smoke that defy capture. We provide, first, an extensive overview of consciousness, exploring its definition and the processes involved in its loss and recovery with DBS. Subsequently, we delve into the evolution of neuroimaging techniques, illustrating how the once opaque surfaces of the crystal ball have transformed into transparent mirrors, allowing for the visualization of the intricacies of consciousness. We outline the dataset at our disposal, which will be subject to interpretation in later sections.
- In part II, we leverage this dataset to investigate the spatial manifestations of consciousness, both during wakefulness, unconsciousness and resurgence by DBS. By peering into the metaphorical crystal ball, we scrutinize the shapes, patterns, and colors that manifest within it.
- In part III, we propose incorporating a dynamic dimension, examining the evolving patterns and transitions over time. Our objective is to characterize consciousness's dynamics, unraveling its manifestations' temporal subtleties. To simplify the analysis of these high-dimensional datasets, we reduce the dimensionality by using deep learning methods. Transformed into symbols, reduced to their essence, the patterns can reveal themselves and enlighten us.

All illustrative images without captions have been generated with the assistance of an AI, Dall-E 3, without modification.

Part I

Background - The crystal consciousness



1 - To be or not to be... conscious

Contents

1.1	Mirror mirror, on the wall, consciousness...	9
1.1.1	... what is it ?	9
1.1.2	... how to study it ?	13
1.1.3	... why studying it ?	25
1.2	Losing consciousness	27
1.2.1	Disorders of consciousness (DoCs)	27
1.2.2	Anesthesia-induced loss of consciousness	32
1.3	Restoring consciousness: from science fiction to science	35
1.3.1	Pharmacological treatments	35
1.3.2	Neurostimulation to restore consciousness	36
1.3.3	Challenges	41

1.1 . Mirror mirror, on the wall, consciousness...

1.1.1 what is it ?



By Olivier Bonhomme from [170]

Human consciousness, though central to our experience and understanding of the universe around us, remains one of the most enigmatic phenomena. For centuries, philosophers, scientists and thinkers have attempted to pierce the veil of consciousness, seeking to elucidate its origins, nature and implications. This thesis explores consciousness from a multidimensional perspective, examining recent advances in the fields of neurobiology and the science of artificial intelligence. Consciousness, as the cognitive faculty that enables us to perceive, feel, reflect and act as thinking beings, raises fundamental questions that affect our understanding of reality, subjectivity and the nature of mind. How does consciousness emerge from a complex network of neural processes? What are the links between consciousness and sensory perception, memory, emotion and decision-making? The recent explosion in technological advances and interdisciplinary research methods has opened up new perspectives for exploring consciousness. State-of-the-art brain imaging, advanced computational models and innovative philosophical approaches are converging to shape a deeper, more nuanced understanding of this complex phenomenon. At the same time, discussions about artificial consciousness are raising fundamental ethical and metaphysical questions about the nature of consciousness and what distinguishes conscious beings from purely computational systems.

A few years ago, it would have been unthinkable to imagine that these few introductory lines could have been written by a being devoid of conscience. But I must give credit where credit is due: these words are not mine. A computer program, an artificial intelligence, ChatGPT wrote them. The text is coherent, the syntax and spelling are perfect. Semantically, there's nothing to complain about; it makes sense and responds well to my request, "Write an introduction on consciousness as part of a thesis". Had you guessed that it wasn't the fruit of a brain? Even for the creators of this AI, distinguishing between the work of their tool and that of a human has become an arduous task. In January 2023, they proposed using a machine learning algorithm to distinguish between text written by an AI and text written by a human. A few months later, they published the following note on their site: "As of July 20, 2023, the AI classifier is no longer available due to its low rate of accuracy" [179]. Lacking confidence in their results, they abandoned the project. If I'd had more writings put online before 2021, when the training of this artificial intelligence ended, it would have been able to reproduce even my writing style, as it did in the preface to *Sapiens* (2022 edition) by Yuval Noah Harari [93]. It raises new questions on a subject that continues to be debated: what is consciousness?

Defining consciousness

Consciousness. Latin *conscientia*, from *scire*, to know. Immediate intuitive or reflexive knowledge of one's own existence and that of the outside world [130]. The human faculty of knowing and judging one's own reality [204]. [In man, unlike other animate beings] Organization of his psyche which, by enabling him to be aware of his states, his acts and their moral value, enables him to feel that he exists, to be present to himself [43].

These three definitions, taken from leading French dictionaries and translated, illustrate how polysemous the term conscience is. While some characteristics overlap in the various definitions I have come across, it is also easy to find a definition that matches what you want to say about conscience. So there's no need to start a debate on consciousness with your friends and hope to find an answer in the dictionary that will make everyone agree. You'll find as many definitions as there are opinions on the subject. As many as there are languages, too, since the nuances of the translation of this term are not precisely the same from one language to another. Scientific studies on consciousness are often criticized for failing to define precisely what consciousness is [78]. Part of the problem is that consciousness is a term that comes from everyday, non-scientific language. It covers different realities and is used to designate different concepts.

The polysemy of consciousness in scientific literature: a bibliometric analysis

This plurality of definitions is pretty self-explanatory. ChatGPT placed consciousness at the interface of several domains, and it is not wrong. A bibliometric analysis of "consciousness" in the literature quickly confirms this. The visualization proposed by "Web Of Science" places consciousness at the frontier of neuroscience, philosophy, clinical neurology, psychology and the social sciences (see Fig. 1.1). With so many different disciplines taking consciousness as their object, it is understandable that the term does not cover the same reality from one field to another. To understand these different nuances, a visualization of co-occurrences is proposed, via the VOSviewer software (v1.6.19) [247]. The terms present in the titles and abstracts of the 10,000 most relevant articles on the "Web of Science" site and containing the term "consciousness" are used. The results of this lexical landscape reveal three clusters (see Fig. 1.2). The first, in red, highlights the terms "theory", "experience", "mind" and "philosophy". We also find "phenomenal consciousness" and "content". All these words are widely associated with theories of consciousness and cut across the fields of philosophy and psychology. The second large cluster, in green, highlights the word "patient". Around it, three sub-clusters gravitate: the field of Disorder of Consciousness (DoC): "recovery", "brain injury", "DoC"; that of the connectivity of conscious networks: "network", "connectivity",

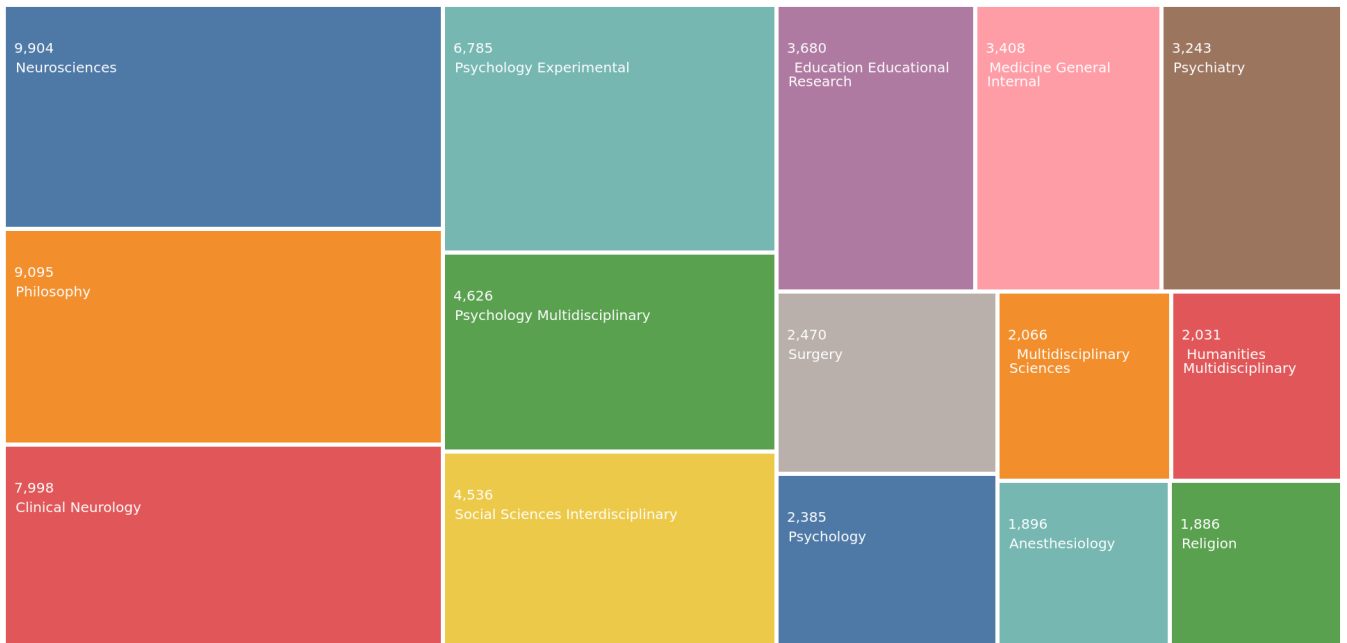
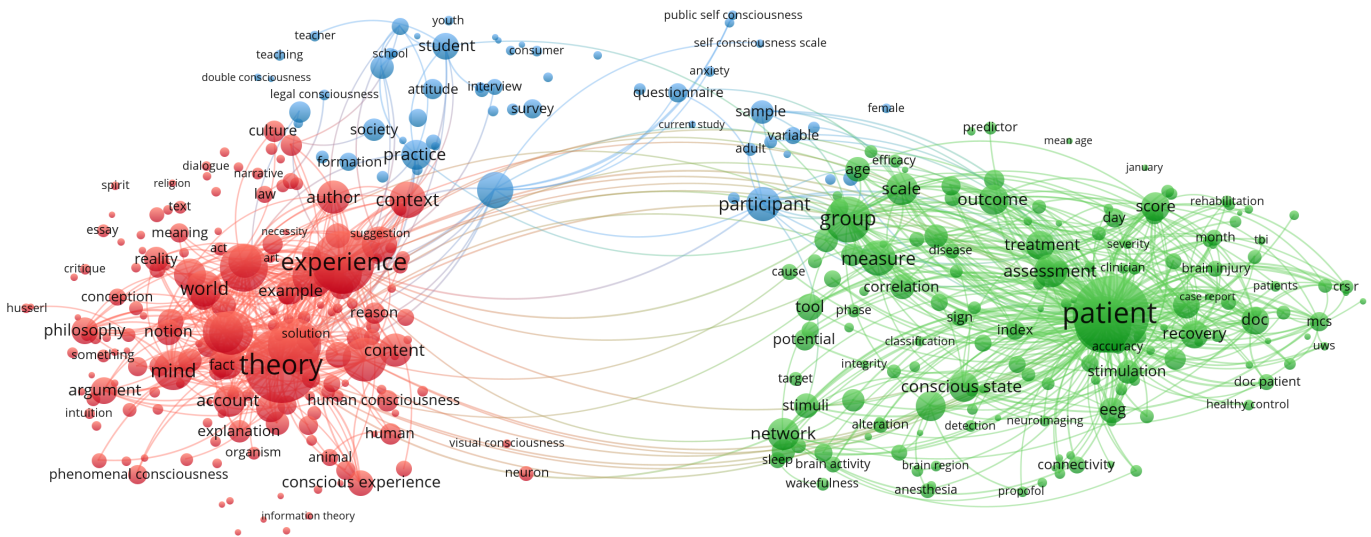


Figure 1.1: The visualization proposed by "Web Of Science" places consciousness at the frontier of neuroscience, philosophy, clinical neurology, psychology and the social sciences.

"conscious state", "neuroimaging"; a final, more methodological cluster: "group", "measure", "scale", which is also close to the third cluster. This last blurrier cluster, in blue, highlights the words "participant", "questionnaire" and "society". It seems to combine an experimental science aspect, where the presence of participants for experiments, in cognitive science, for example, is essential, with a societal aspect that's harder to pin down. Looking at the "historical" evolution of these terms (Fig. 1.3), we see that before 2015, consciousness literature revolved a great deal around the various theories. Then, there is a gradual shift towards clinical neuroscience and neurology, where studies focus on the patient.

The components of consciousness

Starting from a definition of consciousness based on common language, using definitions for the general public, we then studied the meanings of this word in scientific and technical literature. This led to the emergence of sub-groups, and within these sub-groups, different concepts linked to the word "consciousness". These components of consciousness are further described in the literature review proposed by *Dehaene and Changeux (2011)* [58]. Indeed, it is reported that the word "conscious" is ambiguous. Used intransitively, as in the sentence "he was still conscious when the firemen arrived", it refers to the state or level of consciousness, also called vigilance, wakefulness or arousal. Used intransitively, as in the sentence



VOSviewer

Figure 1.2: Co-occurrence map (VOSviewer (v1.6.19) [247]) based on the titles and abstracts of the 10,000 most relevant articles containing the term "consciousness", on the "Web of Science" site. The results of this lexical landscape highlight three clusters: the first, in red, is mostly associated with theories of consciousness; the second, in green, is centered around the clinic; the third, in blue, remains difficult to define (societal aspect, cognitive sciences).

"I was not aware of the danger", it refers to conscious access to specific information, the conscious content, also known as awareness. Consciousness can be understood as a continuum, with varying levels of wakefulness and awareness [231]. These two perspectives help us to define more clearly what we mean when we use the word consciousness. They are also essential in the clinical field, enabling us to situate the various states and pathologies linked to disorders of consciousness (Fig 1.4).

Our subject of research has just been dissected in an attempt to understand it better. But we still don't know how to study it. What forms does the study of consciousness take, and how is it investigated?

1.1.2 how to study it ?

Our bibliometric analysis of the literature has been telling us about consciousness over the last fifty years. However, its evocation goes back much further. To better understand what consciousness means today, it's worth delving into its past. This will give us a deeper understanding of what it has become.

Consciousness history

Descartes

The first systematic approach to consciousness is attributed to René Descartes

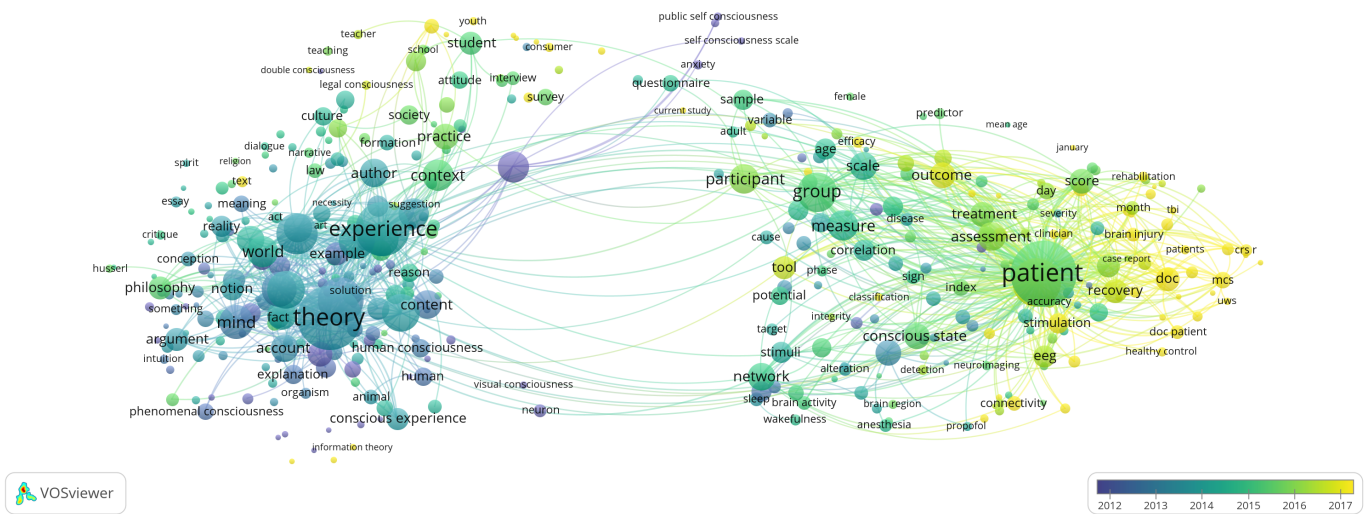


Figure 1.3: Same co-occurrence map as Fig.reffig:2-1:b, colored according to the temporal evolution of these terms. Before 2015, the consciousness literature focused on theories. Gradually, the patient becomes central, and consciousness moves into the domain of clinical neuroscience and neurology.

(1596-1650) [78]. He is known for his dualism: the distinction he makes between the physical body and the mental, the soul. For Descartes, the brain, while playing an important role in sensory input and motor output, is not the basis of the mind. According to him, the brain plays a linking role between matter and mind, in particular, the *pineal gland* (Fig 1.5) [78]. We now know that the pineal gland is not the seat of the soul, and most scientists reject the idea of dualism, believing that the mind emerges from the physical properties of the brain. However, the concept of a mind/body distinction remains, which means that consciousness today is still a complex problem.

Pineal gland (or epiphysis).

Small endocrine gland in the epithalamus of the vertebrate brain. It plays a central role in regulating biological rhythms (sleep/wake and seasonal) by secreting melatonin. In the human species, the pineal gland is shaped like a pine nut (hence the adjective pineal). Its role was poorly understood for a long time, giving rise to several speculations, such as its supposedly central role in thought. Descartes, in particular, based his theory on the fact that the pineal gland was the only head organ not to be conjugated, i.e., not to appear as a pair of symmetrical organs located on either side of the sagittal plane. Moreover, its central position and the fact that it lies just above the aqueduct of Sylvius, which Descartes believed guided what he called the "animal spirits" that were supposed to give rise to sensations in the soul by striking the pineal gland, contributed to the confusion. Today, thanks to histological studies, we know that the pineal gland is indeed a conjugated organ, but the two hemispheres that make it up are almost fused.

After Descartes

Several philosophers and scientists took up the question after Descartes and tried to unite body and mind (Baruch Spinoza (1632-77), Gottfried Leibniz (1646-1716)). John Locke (1632-1704) differentiates between the outer sense, the thought exper-

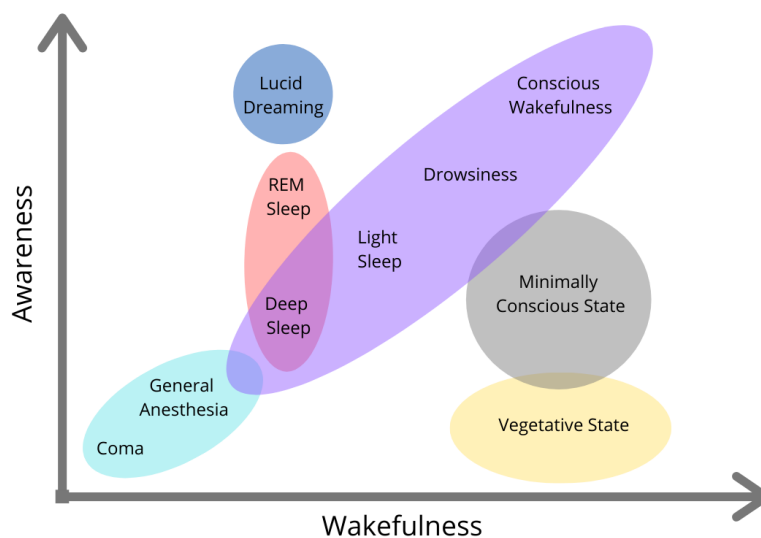


Figure 1.4: Level (wakefulness) and contents (awareness) of consciousness. Consciousness can be understood as a continuum, with varying levels of wakefulness and awareness [231].

rience, and the inner sense, the reflective capacity around the thought experience. His ideas, and those of other empiricist philosophers after him, helped to found a science of psychology. For Immanuel Kant (1724-1804), on the other hand, such a science is not possible, since it uses neither mathematics nor experimentation to study the mind. For a long time, the idea persisted that psychology could not be a scientific object, especially when studying subjective experience. In the 19th century, the belief that mental phenomena cannot be scientific objects was countered by psychophysical methods, and experimental psychology developed. In particular, Gustav Fechner (1801-87) showed that the relationship between stimulus intensity and subjective sensation is logarithmic (the Weber-Fechner law). He thus showed that what is mental can be measured and has a strong link with the physical body [78].

At the same time, significant progress is made in understanding the nervous system. The myograph was developed by Hermann von Helmholtz (1821-94), enabling the speed of nerve impulses to be measured. Contrary to the beliefs of the mainly vitalist biologists of the time, life no longer depended on a vital force that could not be calculated. By the end of the 19th century, much progress had been made on the brain. Ramon y Cajal proposed the idea that the basic unit of the nervous system is the neuron [113], motor and sensory regions were identified by Ferrier, a pioneer in animal electrical stimulation studies [71] and Brodmann began to identify various areas that still bear his name today [26]. At the same time, psychology became a scientific discipline in its own right, and an essential technique



Figure 1.5: *Diagram of the pineal gland as seen by Descartes in his treatise L'Homme (published in the 1664 edition) [63].*

for measuring the duration of mental events was proposed by Donders (1818-89): reaction time. Thanks to this measurement, cognitive-process correlates could be isolated for the first time. The dominant figure in late 19th-century psychology was William James [109]. He equated the flow of thoughts with consciousness and recognized the importance of attention and unconscious processes. In 1913, John Watson proposed that scientific psychology should be based on observable events (stimuli and responses) rather than on hypotheses about mental states [251]. The behaviorist movement was born of this idea, and for the first half of the 20th century, experimental psychology excluded the possibility of studying subjective experiences [135].

The early 20th century is often considered a desert in the field of consciousness. However, additional building blocks were laid, particularly in the study of visual illusions, in which subjective experience is decoupled from physical stimuli. Although these illusions had already been described, they are now being brought to the fore to understand the mechanisms of perception. In addition, the introduction of information theory by Hartley (1928) [94] and Shannon and Weaver (1949) [222] marked the first step towards a mathematical approach to cognition. The brain can now be seen as a communication system that processes and transmits information, rather than movement or energy. The neuron is no longer merely the anatomical central unit, but the information-processing unit. The medical sciences were not to be outdone, particularly by Penfield's research on humans. In epileptic patients, in preparation for operations to remove epileptic areas, he applied electri-

cal stimulation to the brain to locate the regions involved in language and thoughts, to avoid damaging them [190]. This enabled him to obtain verbal reports from patients about their subjective experience following stimulation, demonstrating the importance of the cerebral cortex in conscious experience [70].

The last fifty years

The most significant advance in the last fifty years has been, paradoxically, the study of the absence of consciousness. Demonstrating the unconscious of automatic psychological processes in perception, memory and action, are indeed major advances. Neuropsychology in animals, particularly in monkeys, made it possible to study the behavioral response to a cerebral modification [120, 121]. Other works highlight the role of specific regions of the Pre-Frontal Cortex (PFC) in tasks involving short-term memory [164]. In current theories of consciousness, short-term memory (also working memory) and the PFC are always central [135]. In humans, subliminal perception is being studied: this modifies a subject's behavior even though the stimulus has not been consciously perceived [125]. The study of brain-damaged patients provides an even better understanding of unconscious psychological processes. Some, for example, can guess the properties of a visual stimulus they can't see [252]. Others with severe amnesia, however, can retain information about a stimulus they have no memory of having seen before [250]. But studies of the unconscious raise a new problem: how can we verify that a subject who claims not to have seen a stimulus has not, despite everything, processed it? Conversely, if he correctly detects or discriminates a stimulus, this does not mean that he was aware of it. Underlying unconscious processes may be behind the choice. The development of brain exploration methods, including neuroimaging, is helping to accumulate additional markers of these unconscious processes. It offers the possibility of associating specific patterns of activity with unconscious processes.

In the 21st century, while we now know that life does not depend on a vital essence, consciousness remains a mystery. Science and philosophy continue to intertwine when it comes to consciousness. Consciousness does not pass from a philosophical past to a scientific present. Throughout history, it has been and continues to be, studied by philosophers and scientists. Neither discipline has been neglected as the understanding of phenomena has improved. The two are virtually indissociable, since the questions one poses, raise new ones in the other. Where does consciousness emerge from? What is its role? If, as Descartes thought, it is no longer necessary for rational thought and decision-making, what is it good for? How is the presence of consciousness determined? Are there biological markers of consciousness? Over the years, we have seen many different pairs of glasses put on to try and see things more clearly. Consciousness is multi-faceted, and like the visual illusions we use to study it, it's almost as simple as blinking an eye to observe it differently.

Access consciousness and phenomenal consciousness

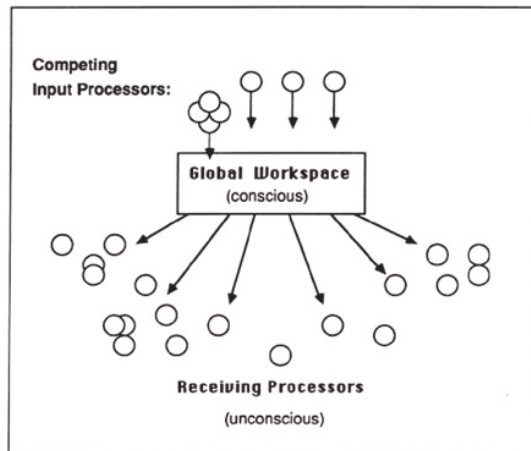
Consciousness is, therefore, a polysemous object, and its study has evolved considerably over time. In modern literature, two facets of consciousness are often developed. We present here a summary, based on Kriegel's (2007) chapter [127], which explains this duality. Philosopher Ned Block, in particular, distinguishes between access consciousness and phenomenal consciousness [21]. The former concerns what enables us to act, i.e., information that is accessible to the cognitive system to speak, reason and control high-level actions. The latter relates to subjective experience and the feelings associated with that experience. These two aspects of consciousness may not be present simultaneously. For Block, these two concepts show different properties of consciousness. For him, phenomenal consciousness is the most difficult to study and understand, but its research is overshadowed by mainstream work on conscious access. Some thinkers and scientists disagree with this view and consider phenomenal consciousness and access consciousness to be linked (see [62, 42, 127]). Lionel Naccache goes even further, arguing that conscious access alone is sufficient to explain consciousness, and demonstrating some of the limitations of Block's definition, particularly on the theme of subjective report [174]. He proposes a definition in which phenomenal consciousness is subsumed within conscious access [174]. These distinct visions of consciousness consequently promote associated theories that are sometimes difficult to reconcile.

Major modern theories

The last thirty years have seen the flowering of theories of consciousness. This explosion of attempts to explain the phenomenon stems from the need to understand its biological and physical foundations and to make the causal link between neural mechanisms and consciousness. The resulting landscape is dense and heterogeneous, with over twenty theories proposed [219]. Bringing the theories together and putting them to the test of experience still remains a challenge. Seth and Bayne (2022) [219] propose a review of the four most important theories. We are going to look at two of them, which are also the subject of an opposing collaboration, to try to decide between them [47]: Global Workspace Theory (GWT) and the Integrated Information Theory (IIT).

Global Workspace Theory (GWT)

The cognitive theory of GWT, developed by Baars [9], has its roots in the theory of "blackboard systems" in Artificial Intelligence (AI), where the blackboard is a central resource that receives and transmits information from specialized processors. In the GWT of consciousness, so-called "conscious" mental states are those that can interact with a number of other automatic, non-conscious processes, such as memory, attention and reporting. These mental states are said to be "globally accessible" (Fig. 1.6).



Baars 1989

Figure 1.6: As per Baars, conscious access occurs when information gains entry to a global workspace, which subsequently disseminates it to various other processing units. From [58].

This accessibility of information is reflected in the neural theory of consciousness, the Global Neuronal Workspace (GNW) theory, later developed by Stanislas Dehaene. Sensory information becomes conscious when disseminated to an anatomically localized workspace, particularly at the fronto-cingulo-parietal level. Dehaene and Changeux's proposal [58] is that a subset of cortical pyramidal cells with excitatory long-range axons that are particularly dense in the prefrontal, cingulate and parietal regions, together with precise thalamocortical loops form the neuronal workspace (Fig. 1.7) [58, 158].

First, a stimulus is perceived unconsciously; then, in a second step, at around 250-300 ms, this representation accesses the global neuronal workspace. One criticism of this theory is that it fails to consider the phenomenal difference between different types of experience. This theory is more concerned with what makes a representation conscious (the why) than the experience of the phenomenon itself. For example, in the *binocular rivalry* task (Fig 1.8), this theory seeks to understand why, at a given moment in time, it is the house that becomes conscious in our mind rather than the face. It is not concerned with the experience of seeing a house or a face. Its focus is on conscious access: why some representations are accessible to the workspace while others are not. This theory argues that different states of consciousness can be observed when the workspace is functionally altered. For example, among the signatures of consciousness identified, altered functional connectivity is visible in fronto-parietal regions (core regions in the theory) during a loss of consciousness. Moreover, functional connectivity patterns become very similar to structural connectivity, as if constrained by it [14, 245].

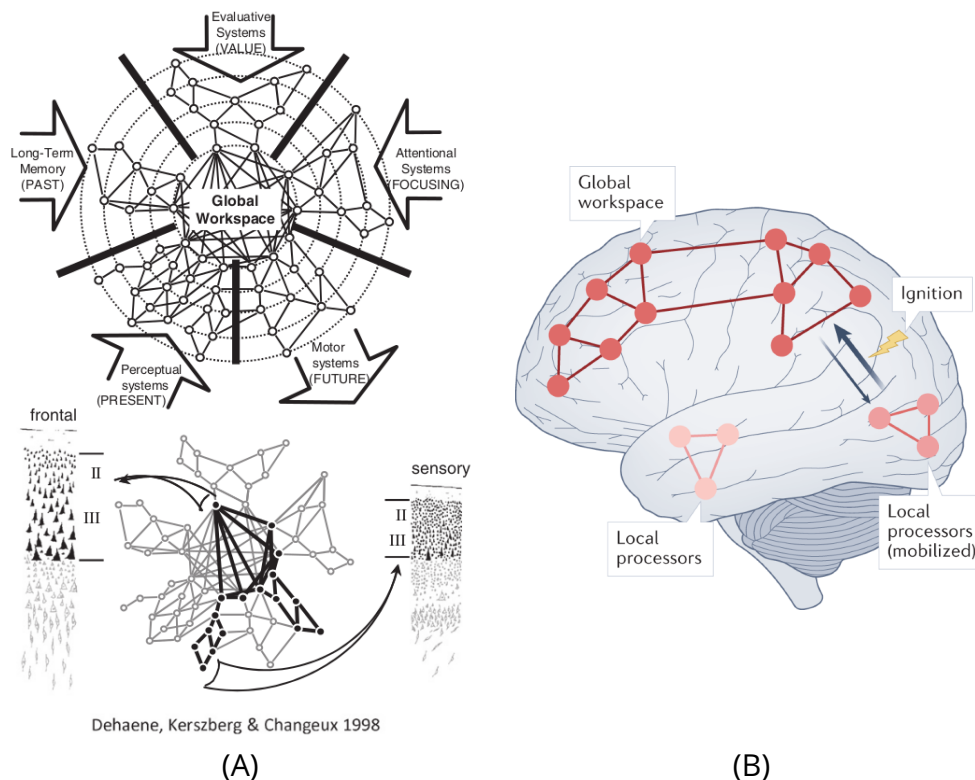


Figure 1.7: GWT (A) The GNW hypothesis suggests that various associative functions, including perception, motor control, attention, memory, and value assessment areas, interconnect to create a higher-level unified space where information is widely shared and relayed back to lower-level processing units. The defining feature of the GNW is its extensive connectivity, facilitated by layers comprising large pyramidal cells that send long-distance cortico-cortical axons, particularly concentrated in the PFC. From [58]. (B) GWTs of consciousness posit that mental states achieve consciousness by being broadcasted within a global workspace, with fronto-parietal networks serving as a central hub. During ignition, activity in specific local processors, such as sensory regions, is temporarily integrated into the workspace. From [219].

Integrated Information Theory (IIT)

IIT posits that consciousness can be delineated by a system's inherent structural capacity to generate irreducible, integrated information. This concept is quantified by a singular and measurable parameter known as " ϕ " [239, 240, 76]. From an anatomical perspective, the theory associates consciousness with posterior cortical areas, often referred to as the "posterior hot zone," which includes parietal, temporal, and occipital regions. Within these regions, neuroanatomical properties are finely arranged to generate high levels of integrated information ϕ [219]. The manifestation of consciousness hinges upon physical and functional information summation within the interconnected neuronal circuits [238]. Subjective conscious



Figure 1.8: Binocular rivalry. Visual occurrence that arises when dissimilar monocular stimuli are presented to corresponding retinal locations in each eye. Instead of perceiving a steady, unified amalgamation of the two stimuli, individuals undergo fluctuations in perceptual awareness over time as the competing stimuli vie for dominance. This phenomenon serves as a fascinating illustration of multistable perception that have been effectively employed to investigate visual processing beyond conscious awareness. From [258].

experiences are underpinned by three neural processing domains: (i) perceiving sensory input from external and internal stimuli, (ii) making decisions and preparing for actions, and (iii) intentionally controlling emotions, thoughts, and actions. Consequently, these neural domains encode various aspects of an experience that necessitate integration to form a unified and cohesive conscious representation of our surroundings [141]. One of the limitations of this theory is that it says little about its link with other aspects of thinking, such as memory, attention or learning. It is also difficult to test empirically. A proxy measure of ϕ is the Perturbational Complexity Index (PCI), which measures the complexity of brain responses to Transcranial Magnetic Stimulation (TMS). This index has been used to diagnose and predict the level of consciousness in patients with disorders of consciousness [36]. However, these measures are not incompatible with other theories of consciousness either. Since the condition for a system's consciousness is that it generates integrated information independently of the substrate, this theory suggests that there can be consciousness in any material system. This, coupled with the fact that the theory is only partially empirically proven, led several researchers in September 2023 to publish an article reminding us of the dangers of taking an experimentally unvalidated theory at face value. In particular, they warned of the ethical consequences this could have on the status of organoids and the embryo [72].

The time for a unified theory of consciousness does not seem to be just around the corner, given the divergent nature of the various theories. Consciousness always shows many faces, which complicates the task of painting a clear picture. Although the mechanisms underlying the concepts of GWT and IIT appear distinct, they converge on a common notion: the necessity for various domains of neural information to interact in facilitating coherent conscious processing. These theories aim to elucidate consciousness through neural activity within the brain, essentially focusing on the Neural Correlates of Consciousness (NCC). Consequently, concepts of consciousness based on NCC agree that conscious processing is both

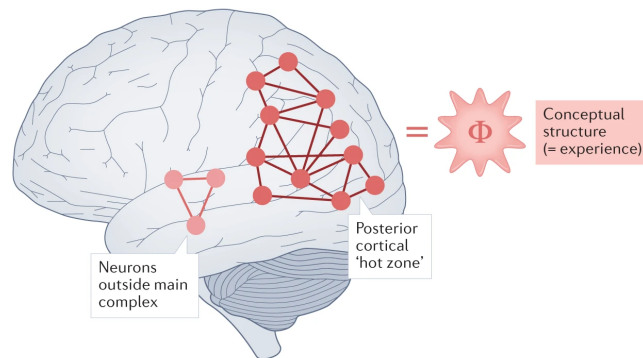


Figure 1.9: Integrated Information Theory (IIT). IIT asserts that consciousness is fundamentally tied to the cause-effect structure of a physical system, particularly one that delineates the maximum irreducible integrated information. The content of consciousness is intricately linked to the form of this cause-effect structure, while the degree of consciousness is determined by its irreducibility, measured quantitatively by the parameter ϕ . Anatomically, IIT is linked to a posterior cortical 'hot zone.' However, validating this core claim empirically poses challenges, mainly due to the difficulty of measuring ϕ , especially in complex systems, limiting assessments primarily to simple model systems. From [219].

unifying and integrative [141]. The many theories help us to apprehend its various aspects, and, little by little, consciousness reveals itself. What remains to be understood is where it comes from: what is its substrate? What is the neuroanatomy of consciousness? And what are its neurophysiological bases?

Consciousness and its interaction with other high-level functions

Consciousness is a fully-fledged brain function, separable from other brain functions such as attention, language, reflection, ... [133]. For example, as regards the link between consciousness and sensory stimuli, dreaming and *locked-in syndrome* show that there is no need for interaction with the environment or motor action to be conscious. It would seem that relying on the subject's physical state to conclude about the presence or absence of consciousness is not appropriate. Instead, it all depends on what the brain is doing, and how and where the information is processed. In the case of dreaming, for example, Electroencephalography (EEG) and neuroimaging studies show that the cortico-thalamic system continues to function in a similar way to the waking brain [133]. Similarly, numerous studies teach us that consciousness needs neither language to exist - patients suffering from a neurological symptom such as aphasia, a communication disorder that can manifest itself in oral or written expression and/or comprehension, are conscious even if their consciousness is altered - nor introspection - patients suffering from a neurological symptom such as aphasia, a communication disorder that can manifest itself in oral or written expression and/or comprehension, are conscious even if their consciousness is altered, nor introspection - we don't need to be aware that we're

conscious to be conscious, nor memory - patients with lesions in the PFC whose working memory is severely damaged, or those with impaired episodic memory, are also conscious. On the other hand, when amnesia is very severe, consciousness is impaired (confusion, loss of reference points). Attention is more controversial: for some, attention is an indispensable prerequisite for consciousness, while for others it is possible to be aware of something without paying attention to it. Neuroimaging studies seem to suggest that the neural correlates of consciousness with and without attention are different (PFC mainly, with some interactions with posterior cortex vs. rather only interactions in posterior cortex) [133].

Locked-in syndrome

Rare disorder of the nervous system characterized by paralysis, except the muscles controlling eye movement. Individuals with locked-in syndrome remain conscious (aware). They still have the ability to think and reason, yet are unable to move or speak. However, communication may be possible through eye movements, such as blinking [176].

The neural substrate of consciousness

According to Laureys and Tononi (2009) [133], we still don't know exactly what the basic unit of consciousness is: are we talking about groupings of neurons, individual neurons, neurons located in particular layers or of a certain type? More generally, it's not clear what set of brain regions is considered minimally necessary and sufficient for the emergence of consciousness. On the other hand, it is known that inactivations or lesions of the corticothalamic system result in loss of consciousness, whereas lesions in other regions have no influence, such as the spinal cord or cerebellum [133]. If we know that a brainstem lesion has a very high probability of putting a subject into a coma, the study of minimally conscious patients shows that if the brainstem is functioning but the corticothalamic system is not, then the patient may be awake but lacks the conscious part of experience. Other regions, such as the claustrum, a thin layer of gray matter located between the insula and the putamen, and the basal ganglia, seem to be involved in consciousness, but it is not clear whether they are necessary for the emergence of consciousness. As for the thalamus, while its crucial role in consciousness is generally accepted, for some, it has a central role in consciousness, particularly the intralaminar nucleus, while for others, it has an indirect role, as a kind of antenna facilitating cortico-cortical interactions. At the same time, some data suggest links between the cortex and consciousness without necessarily involving the thalamus [133]. After arguing that cortex and consciousness are undeniably linked, the authors of "The Neurology of consciousness" (2009) [133] ask where consciousness is located: at the front or back of the cortex? On the sides or in the middle? In the right or left hemisphere? To answer these questions, the most convincing results come from the consequences of lesions in patients. Studies of patients with lesions in the PFC show they do not lose consciousness. However, patients with lesions in both hemispheres are rare, as are lesions in specific regions

of the PFC, making it impossible to study the role of some regions. In comparison, studies of patients with lesions in the posterior cortex appear to lose awareness while remaining awake. The authors support the view that awareness is more likely to reside in the posterior cortex but concede that this does not detract from the fact that the PFC also plays a direct role. As for the involvement of the lateral vs. medial regions, both patient observations and neuroimaging studies suggest that both regions are involved in conscious experience and form a highly interconnected network. Finally, the study of *split-brain* patients, or patients who have undergone a right *hemispherectomy*, is rich in information as to whether consciousness resides in the right or left hemisphere. These studies show that the left hemisphere alone can support self-awareness similarly to a whole brain. The right hemisphere, on the other hand, is limited in most people when it comes to language and reasoning functions and is largely dominated by the left hemisphere. However, rarer cases of left hemispherectomy indicate that consciousness and self can also exist in the right hemisphere. The authors hypothesize that "right" consciousness is comparable to that of certain primates, conscious but deprived of language.

Split-brain

A disconnection syndrome that arises when there is a partial cut of the corpus callosum, the structure linking the two brain hemispheres.

Hemispherectomy

Surgical operation in which one cerebral hemisphere is removed or disconnected from the other.

The neural correlates of consciousness

The work summarized in "The Neurology of consciousness" (2009) [133] shows that changes in neural activity and conscious experience are not necessarily correlated. We also know that while the cortex is active during *sleep Non Rapid Eye Movements (NREM)* and anesthesia, subjects are not necessarily conscious. These few elements suggest that cortical activity is not enough to generate consciousness. This activity must possess certain characteristics, particularly dynamic ones, for conscious experience to be present. In particular, the authors question the need for activity to be prolonged, or per phase, reentrant or feed-forward, and synchronous or oscillatory. One of the most likely ideas is that neural activity participates in consciousness if, and only if, it is maintained for a certain length of time, around a few hundred milliseconds. Experiments using *the attentional blink* (Fig. 1.10), show that around 200-400 ms, target detection is impaired, which is less the case directly after the first target. This experiment, and others following it, indicate that consciousness seems to require neural activity in the appropriate brain structures that lasts for a minimum time, perhaps the time required for interaction between several areas. Other data suggest that it does not require continuous cortical activity, but rather neuronal activity in phases: neuronal discharges would not cause consciousness directly, but would activate areas that directly support

consciousness. Another hypothesis takes a closer look at how the neural wave propagates. Here, it's not so much a question of whether it's a continuous or discrete stimulus that triggers awareness, but rather whether there's a "returning" wave of activity (also known as recursive, recurrent or reverberating) from high-level to lower-level areas.

Sleep (Non Rapid Eye Movements (NREM)/Rapid Eye Movements (REM)). There are two types of sleep: NREM sleep and REM sleep. NREM sleep is further categorized into stages 1, 2, and 3, representing a spectrum of increasing depth. The identification of sleep cycles and stages has been made possible through EEG, which captures the electrical patterns of brain activity. Each of the three stages of NREM sleep is characterized by distinct brain activity and physiological features. REM sleep is characterized by desynchronized brain wave activity (low-voltage, mixed-frequency), muscle atonia (loss of muscle tone), and intermittent REM. Dreaming is predominantly associated with REM sleep. The loss of muscle tone and reflexes during this phase is thought to serve a crucial function by preventing individuals from physically acting out their dreams or nightmares while asleep [45, 216].

1.1.3 why studying it ?

As we saw earlier, consciousness has fascinated, intrigued and questioned us for generations. It raises philosophical and scientific questions, is also an object of study from the point of view of pure knowledge. The more we seem to want to lock it into a box, confine it to a single definition, explain it with a theory, the more it eludes them. But it's not just a mystery, a puzzle to be solved. Behind the understanding of consciousness, as the co-occurrence map clearly showed, lies a darker reality, that of patients suffering from DoCs. A better understanding of how consciousness works could lead to major clinical advances in this area.

Indeed, diagnosis is not straightforward and requires extensive testing to assess levels of wakefulness and awareness. The proposed treatments aim to support the patient (feeding and drinking, movement, hygiene, auditory or visual stimulation) but can't ensure recovery from impaired consciousness. Predicting the chances of improvement of someone in a state of impaired consciousness [218] is impossible. In the remainder of this introductory section on consciousness, we will look at disorders of consciousness and the proxy used to study them.

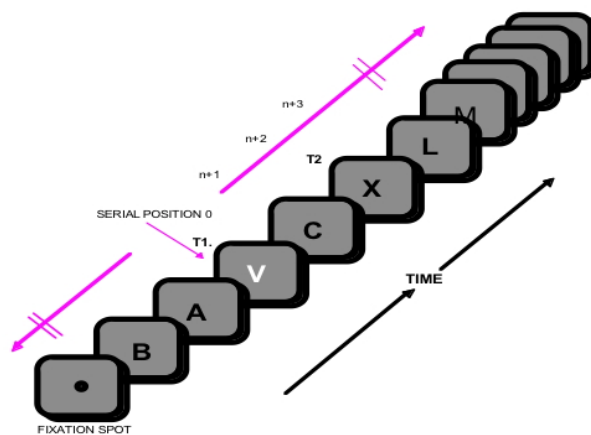


Figure 1.10: Attentional blink. A phenomenon where the second of two targets becomes difficult to detect or identify when it closely follows the first [223]. The experimental paradigm generally employed involves rapid serial visual presentation, where stimuli such as letters, digits, or pictures are presented sequentially at a single location at rates ranging from 6 to 20 items per second. The methodology introduced by [201] involves participants identifying the only white letter (first target; T1) in a rapid stream of black letters (non-targets or distractors) presented at a rate of 10 items per second. Following this, participants are tasked with reporting whether the letter 'X' (second target; T2) appears in the subsequent letter stream. T2 is presented in only 50% of the trials, and when presented, it occurs with an interval between the two targets ranging from 100 to 800 milliseconds. Participants are required to report both targets after the stimulus stream concludes. The attentional blink is said to occur when T1 is reported correctly, but the report of T2 is inaccurate at short T1/T2 intervals, typically ranging between ≈ 100 to 500 milliseconds and accuracy in reporting T2 recovers to the baseline level at longer intervals. From [223]

1.2 . Losing consciousness

1.2.1 . Disorders of consciousness (DoCs)

Coma and other DoCs remain one of medicine's most important challenges. They are characterized as a state of prolonged altered consciousness, which can be categorized into coma, Unresponsive Wakefulness Syndrome (UWS) (previously Vegetative State (VS)) or Minimally Conscious State (MCS) based on neurobehavioral function [18, 180, 236]. It is difficult to know exactly the level of consciousness in these patients and study their states of consciousness because of very reduced or inexistent movements. Misdiagnosis, in particular, is one of the most serious problems as it impacts medical decision-making, and patients' well-being [68].

Main clinical entities of DoC and causes

Among the DoCs, three main clinical entities are defined according to the patients' level of arousal and awareness following a behavioral examination (cf. Fig 1.4).

Coma

Coma arises from severe brain injury, characterized by an absence of arousal (e.g., eyes remain closed even when stimulated) and an absence of self-awareness or awareness of the environment. This state is typically transient, lasting less than two to four weeks, after which patients may progress to brain death or exhibit partial or full recovery [236]. Typically, coma results from the suppression of corticothalamic function due to drugs, toxins, or internal metabolic imbalances. Other contributors to coma include traumatic brain injuries like severe head trauma, non-traumatic brain injuries such as strokes, or hypoxia resulting from heart failure, all of which lead to widespread disruption of corticothalamic circuits (Figure 1.11). Gradual onset of coma can also occur in conditions like Alzheimer's disease, characterized by progressive brain deterioration. Additionally, smaller lesions affecting the reticular activating system can induce unconsciousness by indirectly deactivating the corticothalamic system [133].

UWS

Patients with UWS demonstrate no signs of awareness but may display reflexive movements like teeth grinding, yawning, or groaning. This condition can be transient, prolonged, or permanent [236]. Postmortem analysis of patients with UWS reveals that the brainstem, hypothalamus, and specifically the reticular activating system remain largely intact, which accounts for why patients appear awake despite their unresponsive state (Figure 1.11). Typically, UWS results from extensive lesions in the gray matter of the neocortex and thalamus, widespread damage to white matter and diffuse axonal injury, or bilateral thalamic lesions, particularly affecting the paramedian thalamic nuclei. Thalamic injury may arise as a secondary effect of diffuse cortical damage through retrograde degeneration. However, iso-

lated damage to the paramedian thalamus can lead to persistent unconsciousness [133].

MCS

MCS patients show intermittent but reproducible signs of consciousness. Similar to UWS, the MCS can be temporary or permanent [236]. The function of the cerebral cortex, diencephalon, and upper brainstem is variably impaired (Figure 1.11) [133].

Assessing level of consciousness: from bedside to neuroimaging

Extensive testing is required to evaluate levels of wakefulness and awareness before confirming a DoC. This assessment often includes cerebral exploration (brain scans and EEG), but the gold standard for diagnosing these states of consciousness is behavioral examination using a specialized scale.

Behavioral examination

Physicians utilize mainly the GCS and The Coma Recovery Scale-Revised (CRS-R) to assess an individual's level of consciousness.

- **The GCS** evaluates three aspects:
 - Eye opening: A score of 1 indicates no eye opening, while 4 signifies spontaneous eye opening.
 - Verbal response to commands: A score of 1 implies no response, while 5 indicates alertness and verbal interaction.
 - Voluntary movements in response to commands: A score of 1 denotes no response, while 6 signifies the ability to follow instructions.

A lower GCS score suggests more severe impairment of consciousness, potentially indicating a coma. This score is regularly monitored for any changes in the individual's condition [235].

- **The Coma Recovery Scale-Revised (CRS-R)**

More specialized scoring systems, such as the JFK CRS-R, provide detailed assessments of an individual's behavior [114]. This scale consists of 23 items divided into six subscales (auditory function, visual function, motor function, oromotor/verbal function, communication, arousal), each meticulously evaluated to gauge "perceptual awareness of the environment" [218].

These behavioral evaluations are deemed essential for diagnosing consciousness. Despite the availability of various electrophysiological and neuroimaging

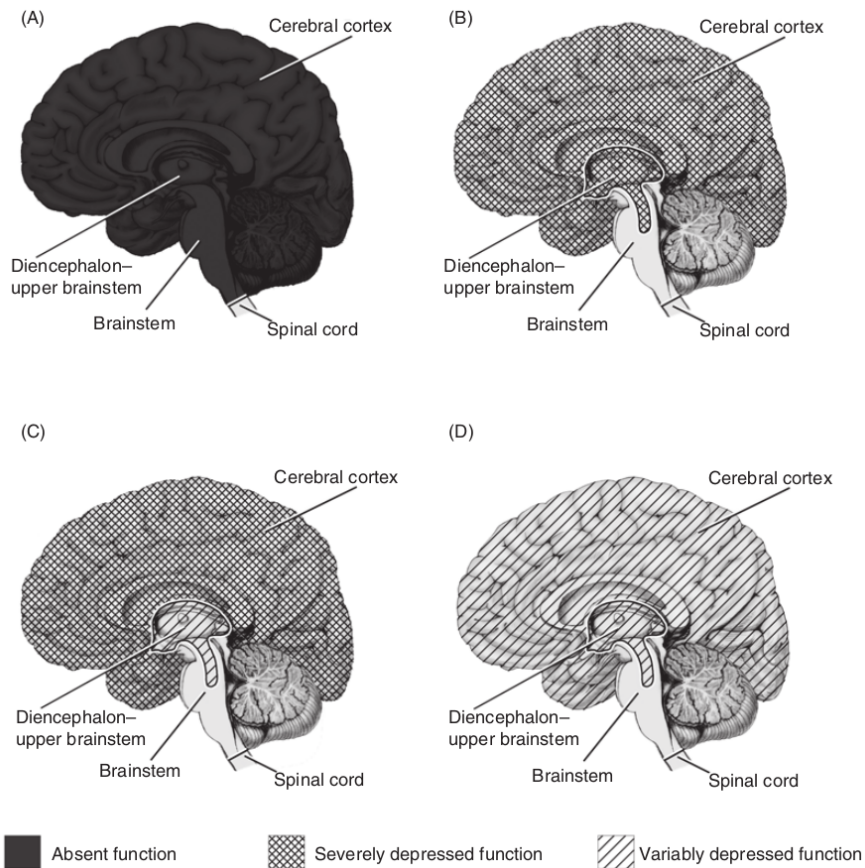


Figure 1.11: Schematic illustration of brain dysfunction in major states of impaired consciousness. (A) Brain Death: all cortical, subcortical, and brainstem functions are irreversibly lost. While spinal cord function may remain intact, eliciting responses from the patient is not possible except for spinal cord reflexes. (B) Coma: severe impairment of cortical function and dysfunction of the diencephalic/upper brainstem activating systems. Patients remain unarousable with closed eyes and lack purposeful responses, although brainstem reflex activity persists. (C) UWS: cortical function is significantly impaired, but some preserved function of the diencephalic/upper brainstem activating systems remains. Patients in this state are unconscious at all times, lacking purposeful responses, yet they may spontaneously open their eyes or do so in response to stimulation. They may also exhibit primitive orienting responses and sleep-wake cycles. (D) MCS or better: In this state, the degree of impairment in cerebral cortex and diencephalic/upper brainstem activating systems varies. Patients may display some purposeful responses along with deficits, with the severity of brain dysfunction determining the extent of impairment. From Chapter 2 of [133]

tools and the extensive research utilizing them, the behavioral examination involving direct interaction between the patient and the examiner remains the gold standard. Yet, fMRI research [181, 167, 175] and scalp EEG studies [41] have revealed that approximately 15–20% of patients diagnosed with DoC who exhibit no apparent behavioral responsiveness may, nonetheless, possess covert conscious-

ness [150]. This highlights the critical need for diagnostic tools that rely on brain activity. The use of brain scans, for example, have prompted revisions in certain categories of DoCs [218]. To clarify the prognosis of patients recovering slowly, several research teams are currently attempting to identify more effective prognostic markers, notably through functional explorations probing residual brain function. Is the dream of neuroimaging researchers to be able to decode levels of consciousness from recorded cortical activity, by studying how global changes in brain activity are linked to different levels of consciousness, becoming a reality?

Functional exploration

Until recently, the predominant neuroimaging techniques employed in DoC involved Positron emission tomography (PET) or Single Photon Emission Computed Tomography (SPECT) to evaluate resting cerebral blood flow and glucose metabolism. However, methodologies like fMRI offer the capability to associate particular cognitive processes with distinct physiological responses (changes in regional cerebral hemodynamics), even in the absence of any overt response (e.g., motor actions or verbal responses) from the patient [180].

Activation studies offer a means to evaluate cognitive functions in altered states of consciousness without requiring any overt response from the patient. For example, this approach has been utilized to identify residual brain functions in patients who meet all standard clinical criteria for UWS but retain cognitive abilities that elude detection using standard clinical methods. Even if they show no evidence of awareness at the bedside, the diagnostic label of MCS* has been suggested for them, as their neuroimaging data show atypical brain patterns using active paradigm (eg, brain activity in motor area during a motor imagery task) or metabolic resting state (eg, preservation of the fronto-parietal network) [180].

Similarly, in some patients diagnosed as MCS, functional neuroimaging has been used to demonstrate residual cognitive capabilities even when no clear and consistent external behavioral evidence supports this conclusion. These investigations have prompted several prominent research groups in this field to propose that integrating emerging functional neuroimaging techniques with established clinical and behavioral assessment methods will be crucial for enhancing our capacity to minimize diagnostic errors among these interconnected conditions [132, 211].

Studying alterations of consciousness: challenges

Unfortunately, studies in patients with DoCs are difficult to perform for several reasons and proxy should be used to study altered consciousness.

Ethical and technical challenges for studying DoCs

Several logistic challenges associated with scanning critically ill patients in the high magnetic field remained unresolved [180]. First, the patients encountered are extremely fragile. Subjecting them to repeated experiments is impractical due to

their rapid fatigue. Moreover, the clinical setting is constrained by time and budget limitations, and transporting patients to various facilities is not feasible. Ethical considerations also pose a significant debate. Disagreements among the patient's family members may arise, complicating further research endeavors that require unanimous consent [232].

Clinical relevance of anesthesia research to DoCs

As previously mentioned, there is a growing body of evidence indicating that misdiagnosis is common in DoCs, with up to 43% of patients considered as UWS showing at least minimal awareness [155]. However, inferring consciousness solely from paradigms involving stimulus presentation to non-responsive patients, followed by observation of brain response using neuroimaging, can be problematic.

In this context, studies examining the effects of anesthetic sedation in healthy human subjects have been particularly informative. Based on the findings discussed here, these studies suggest that intact fMRI responses to simple sensory stimuli, such as speech perception, in patients diagnosed as UWS cannot necessarily be interpreted as evidence of preserved awareness, especially when passive cognitive tasks are employed. Conversely, complex sensory processing, as evidenced by cortical reactivity in association cortices during active cognitive tasks, may indicate undetected consciousness in non-responsive patients [155].

Relevance of anesthesia in preclinical research to DoCs

However, studies involving anesthesia in healthy human subjects in France are not always easy to justify. Anesthesia, although generally considered safe, presents certain risks, as do all medical procedures. This is why some teams are opting for alternatives, such as studying loss of consciousness in NHP. The literature shows that this is a sensible choice, since very similar fMRI signatures of loss of consciousness have been found between comatose patients and NHPs under anesthesia in terms of brain activity patterns. In particular, it has been shown that the brain activity of anesthetized NHP and coma patients resided most frequently in a pattern of low connectivity resembling the anatomy, which was sustained for longer periods of time in comparison to more complex patterns [60, 14]. To study problems of consciousness, anesthesia-induced loss of consciousness is therefore a relevant proxy.

Animal models of traumatic brain injury

Another concept that could have emerged involves the development of an animal model for coma, similar to those established for other conditions such as Alzheimer's disease, Parkinson's disease, and multiple sclerosis. The creation of animal coma models could serve as valuable experimental frameworks for investigating the neural mechanisms underlying significant changes in brain states. This approach holds promise for elucidating the neural foundations contributing to the

reemergence of consciousness [182].

The exploration of the neuroanatomical underpinnings of consciousness has traditionally involved a lesional approach, wherein animal brains are investigated under various surgical conditions. Interestingly, in rats, cats, and dogs, the removal of both cerebral hemispheres does not result in coma; instead, these animals are capable of self-righting, feeding, and grooming. These findings underscore the difficulty in developing an animal model for coma induction and highlighting the crucial role of subcortical brain lesions in this process [182]. Presently, the NHPs, particularly rhesus macaques, serve as robust animal models for consciousness research. Their psychophysical performances bear many resemblances to those of humans, and both the anatomical and functional organization of their brains share significant similarities with humans brains [22]. However, the lesional approach is now being phased out due to evident ethical concerns, with preference given to studies on altered states of consciousness, such as those induced during general anesthesia or sleep.

1.2.2 . Anesthesia-induced loss of consciousness

General anesthesia stands out as the most prevalent method among external interventions to alter the level of consciousness. General anesthesia is a valuable tool for understanding the LOC, as it is a relatively safe practice widely used for surgical procedures. Sedation can be seen as a reversible state of drug-induced LOC, and pharmacological modulation offers the advantage of providing adjustable and reversible sedation [232].

Principal anesthetics

The level of consciousness induced by anesthesia depends on the dosage and the specific agent used, resulting in effects that vary from complete to partial unconsciousness [232]. Anesthetics are categorized into two primary classes: intravenous agents, like propofol and ketamine, typically employed for induction and often administered alongside sedatives such as midazolam and dexmedetomidine; and inhaled agents such as isoflurane, sevoflurane, and desflurane, as well as gases like xenon and nitrous oxide [133].

Effect at the cellular level of anesthetics

At the cellular level, numerous anesthetics have a combination of effects, ultimately leading to a reduction in neuronal excitability through either the augmentation of inhibition or the reduction of excitation. Primarily, anesthetics function by bolstering Gamma-Amino-Butyric-Acid (GABA) inhibition or inducing cell hyperpolarization via the augmentation of potassium leak currents. Additionally, they can impede glutamatergic transmission and counteract acetylcholine at nicotinic receptors [133].

What are the key neural pathways responsible for the loss of consciousness induced by anesthetics?

Effect of anesthetics on circuits

Sites of action: cortex, thalamus, other specific areas ?

A common target for several anesthetics is the posterior cingulate cortex, medial parietal cortical areas, and lateral parietal areas. However, it remains unclear whether anesthetics induce unconsciousness by affecting specific areas or by causing widespread deactivation of corticothalamic circuits. One of the most consistent effects of most anesthetics at LOC is a decrease in thalamic metabolism and blood flow, indicating that the thalamus may play a role of "consciousness switch" [133].

Regardless of the primary target of anesthetics, achieving LOC may not always involve the outright inactivation of neurons in these regions. Rather, subtle alterations in dynamic neural activity may suffice. Similar to the effects observed during sleep, there is evidence suggesting that anesthetic agents disrupt cortical integration, potentially contributing to impaired consciousness. Alternatively, anesthetics might induce a state of bistable, stereotyped cortical responses characterized by reduced information processing, leading to a loss of consciousness [133].

Thalamic consciousness switch hypothesis: the suppression of thalamocortical system activity by anesthesia may result from numerous anesthesia-induced interactions at different brain sites. These interactions collectively lead to the hyperpolarization of neurons within the thalamocortical system.

Disruption of large-scale integration

Anesthetics are recognized for their ability to decelerate neural responses, which can disrupt synchronization among different brain regions. This disruption is evidenced in the reduction of coherence in the gamma frequency range (typically 20 to 80 Hz) between various cortical areas, such as the right and left frontal cortices, and between frontal and occipital regions, as consciousness wanes [112]. Animal studies further support this observation, showing that anesthetics suppress gamma coherence between the frontal and occipital regions, both during visual stimulation and in the resting state. This effect is gradual and more pronounced for long-range coherence compared to local coherence [133]. A study by Uhrig et al. (2018) [245] investigated the corticocortical effects of ketamine, sevoflurane, and propofol anesthesia, revealing that all three anesthetics maintained long-range stationary connections but led to a reduction in both positive and negative correlations compared to the awake state. These findings support the notion that a disruption of long-distance corticocortical networks may explain the anesthesia-induced loss of consciousness.

The disruption of interactions between anterior and posterior regions of the

cortex may be of particular importance. At anesthetic concentrations inducing unresponsiveness in rats, there is a decrease in transfer entropy, a measure of directional information flow, in the front-to-back direction – from frontal to parietal and from frontal or parietal to occipital cortex – even when feedforward transfer entropy remains high [133]. This suggests that consciousness relies on feedback mechanisms.

Anesthetic agents may be especially effective at disrupting integration because the corticothalamic system seems organized like a small-world network with mostly local connectivity augmented by comparatively few long-range connections. Thus, anesthetics need only disrupt a few long-range connections to produce a set of disconnected components [133].

The pronounced effect on disrupting integration induced by anesthetic agents might be due to the organization of the corticothalamic system, which resembles a small-world network characterized by predominantly local connections with a few long-range connections. Consequently, anesthetics may only need to disrupt a limited number of long-range connections to induce a state of disconnected components [133].

Reduced repertoire of activity patterns

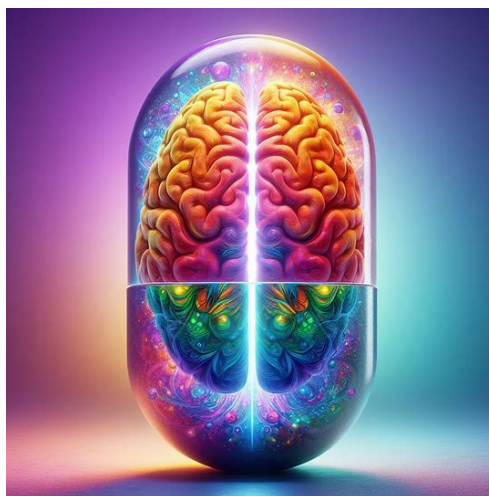
Another explanation for the loss of consciousness is linked to the decline in the quality of cortical responses, hence the notion of information. As the diversity of activity patterns generated by the corticothalamic system diminishes, neural activity becomes less informative, despite potential global integration [133]. Several general anesthetics induce a characteristic burst-suppression pattern characterized by a nearly flat EEG interrupted intermittently by brief, quasi-periodic bursts of global activation that are remarkably stereotypical. This stereotypical burst-suppression pattern, whether evoked or spontaneous, indicates that during deep anesthesia-induced unconsciousness, the corticothalamic system can still be active – indeed hyperexcitable – and can generate global, integrated responses. However, the range of responses has contracted to a stereotypical burst-suppression pattern, leading to a corresponding loss of information. The study by Uhrig et al. (2018) [245] supports the concept of diminished information. Under ketamine, sevoflurane, and propofol anesthesia, they demonstrate that the repertoire of brain states traversed by the anesthetized brain is reduced compared to the awake state.

1.3 . Restoring consciousness: from science fiction to science

Let's not sing a requiem too soon: lost consciousness could soon be restored, thanks to the advent of the latest neuromodulation techniques. Electrical stimulation of the brain to restore consciousness! A science fiction story? Some kind of gloomy dystopia? It's not a story about the future, though. Neuromodulation is well and truly in use in our hospitals. It's not a tool of torture, but on the contrary, a clinical technique used since 1987 to combat certain symptoms in patients suffering from Parkinson's disease. Results have shown a marked improvement in quality of life and motor skills in these patients. Electroceuticals, this novel class of therapeutic agents which target the neural circuits through electromagnetic stimulations, could also lead to improvements in consciousness disorders, with the ultimate aim of restoring consciousness to comatose patients.

1.3.1 . Pharmacological treatments

Obviously, implanting electrodes in the brain of a comatose individual was not the first idea of clinicians. But to date, few therapeutic options exist, and few studies have investigated the treatment of patients with DoCs. Based on the review by Thibaut et al (2019) [236], we first discuss pharmacological options for these patients with prolonged DoCs (i.e., more than 28 days).



Amantadine and zolpidem have been mostly used to improve consciousness and functional recovery in patients with DoCs [236].

Amantadine (*dopamine* agonist and *NMDA receptor* antagonist) enabled a group of patient with Traumatic Brain Injury (TBI) to recover faster than the placebo group as measured by the Disability Rating Scale [200] (a scale developed and tested with individuals with moderate and severe TBI derived from the Glasgow Coma Scale and reflecting impairment ratings). For patients experiencing DoCs due to causes other than TBI, the evidence is less clear, as there have been relatively few studies conducted on this population.

NMDA receptor

The NMDA (*N* – *methyl* – *D* – *aspartate*) receptor, is a glutamate receptor and ion channel found in neurons, essential to memory and synaptic plasticity.

Dopamine

Dopamine (contraction of 3,4 – *dihydroxyphenethylamine*) serves as a crucial neuromodulatory agent with multifaceted functions within cells. Within the brain, dopamine acts as a neurotransmitter, facilitating communication between neurons by transmitting signals from one nerve cell to another. While neurotransmitters are synthesized in distinct brain regions, their effects extend systemically across numerous brain areas.

Zolpidem (non-benzodiazepine GABA agonist) is classified as a hypnotic medication. It has been observed to enhance consciousness and facilitate functional recovery in approximately 5% of patients with DoCs. It is essential to characterize the behavioral and physiological attributes of individuals who respond positively to zolpidem to enhance the identification of patients who may benefit from this treatment [236].

GABA

GABA (γ – *Aminobutyric acid*) is the primary inhibitory neurotransmitter within the mature mammalian central nervous system. Its primary function involves diminishing neuronal excitability across various nervous system regions.

Studies on other molecules seem very anecdotal, given the few uncontrolled studies and case reports. More randomized controlled trials are needed [236]. Strangely enough, despite the absence of more studies on the subject, the use of pharmaceutical means is not really up for debate. Yet these molecules can have undesirable effects, and cannot target precisely the areas of the brain specifically linked to the disorders they treat.

1.3.2 . Neurostimulation to restore consciousness

In order to improve consciousness and functional recovery in patients with DoCs, non-pharmacological interventions have also been explored. These include non-invasive brain stimulations (e.g., transcranial Direct Current Stimulation (tDCS), repeated TMS, transcutaneous auricular vagal nerve stimulation, and low intensity focused ultrasound pulse), invasive brain stimulation (e.g., DBS or vagal nerve stimulation), and sensory stimulation programs (Fig. 1.12) [236].

In this thesis, we focus solely on DBS, but other neuromodulation techniques are not neglected in our team, which is also studying the effects of tDCS (G. Hoffner) and focused ultrasound (A. Bongioanni).



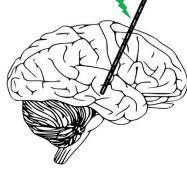
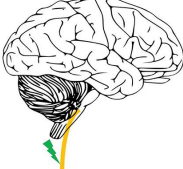
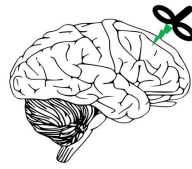
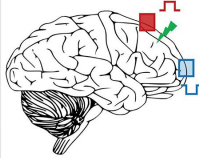
	Deep Brain Stimulation (DBS)	Vagus Nerve Stimulation (VNS)	Rhythmic Transcranial Magnetic Stimulation (rTMS)	Transcranial Direct Current Stimulation (tDCS)
TYPE				
TARGET	Midbrain Thalamus Pallidum Striatum	Vagus Nerve	Right or left dorsolateral prefrontal cortex or Right or left primary motor cortex	Left dorsolateral prefrontal cortex or Posterior parietal cortex
CURRENT	Low (8-30 Hz) or high frequencies (50-250 Hz) 1-20 V voltages	30 Hz 1.5 mA intensity	Single or repeated sessions 5-20 Hz	20 minutes sessions (single or repeated) 1-2 mA intensities
INVASIVE	Yes	Moderately	No	No

Figure 1.12: Illustration of various forms of stimulation utilized in DoC patients, including both invasive and non-invasive methods. The main targets and stimulation parameters are listed (intensities, voltages, frequencies, and number of sessions) used in clinical studies. From [24].

What is DBS ?

DBS involves the surgical implantation of electrodes into deep brain regions to administer electrical currents aimed at modulating neural activity. These electrodes

are implanted using stereotactic techniques, a neurosurgical method utilized to target specific brain areas. Stereotaxy relies on a 3D coordinate system to precisely locate structures within the brain using medical imaging tools. The electrodes are connected to an Implantable Pulse Generator (IPG), typically positioned subdermally beneath the clavicle (Fig. 1.13 (A)). A recent IPG comprises a battery and electronic components responsible for delivering electrical stimulation, and it can be externally controlled by patients or clinicians. Adjustments to stimulation parameters such as frequency, pulse width, and voltage are necessary to optimize efficacy [197].

The electrode serves as the intermediary component between the neuromodulation hardware and the specific nervous tissue being targeted. Electrical stimulation is achieved by establishing a connection between two poles of a stimulus source and the tissue. Typically, conventional current flows from the positive pole of the stimulus source to the negative pole, while electrons (negative charges) move in the opposite direction [126]. Its primary role is to deliver adequate current to selectively activate or deactivate the target neural tissue with which it interfaces.

Various metals, such as gold, stainless steel, platinum, platinum–iridium, among others, have been utilized for neurostimulation electrodes. The selection of the electrode's metal composition depends on factors like biocompatibility, the amount of charge injection required, and constraints related to surface area. The physical design of the electrode is tailored to fit the target anatomy appropriately and achieve the desired spatial activation and selectivity. This may involve shaping the electrode or configuring it with multiple contacts to enable multipolar stimulation. In the case of DBS, the electrode typically takes the form of a cylindrical shaft, as depicted in Figure 1.13 (B) [126].

Why using DBS ?

The narrative review of *Bourdillon et al.* (2019) [24] gives us more insights into the history of DBS for consciousness disorders. At first, performing lesions on deep, small brain structures with extensive projections to large cortical areas presented promising prospects in both psychiatric and neurological fields and significantly reduced the morbidity associated with surgical procedures. These lesional techniques were recommended for pathologies characterized by positive signs, such as tremors or dystonia, but were ineffective for conditions where negative signs predominated, such as DoCs. In this context, the application of electrical stimulation to human patients using stereotactically positioned intracranial deep electrodes emerged. DoC, which was viewed as a deficiency in cortical activation resulting from a disruption of projections from the Ascending Reticular Activating System (ARAS) through the diencephalon to the cortex, was among the first conditions for which DBS was employed [160, 95]. Even if promising effects were observed in the initial reports of brain stimulation on the arousal of vegetative patients, no further studies were conducted until the late 1980s when DBS gained popularity, particularly through

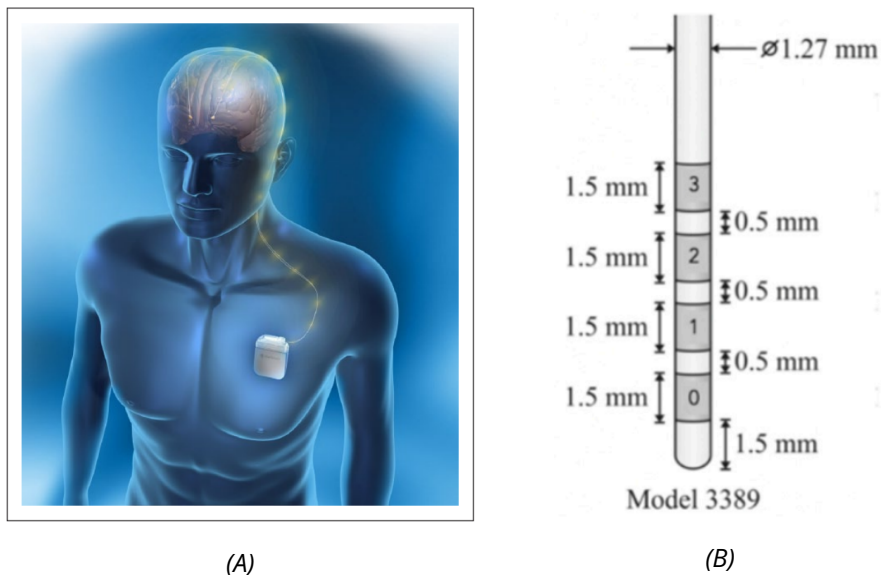


Figure 1.13: (A) Illustration of an implanted DBS system. From [197]. (B) Medtronic 3389 lead DBS model. From [148].

its application in Parkinson's disease [17]. In parallel, as seen previously, several theories of consciousness have been developed. While some authors postulate that consciousness stems from a brain-scale cortico-cortical communication (global workspace theory [58]), others claim that consciousness arises from the coordinated activity within thalamo-cortical as well as non-thalamic ARAS pathways [65, 110], or from fronto-pallido-thalamo-cortical loops (meso-circuit hypothesis, [213]). According to all of these theories, the common feature in DoC pathophysiology would be the disruption of a complex and organized high-order activity among large-scale neural networks [24]. Regardless of its cause, DoC is characterized by a widespread cessation of excitatory synaptic activity throughout the entire cerebral cortex [32]. Recovery from coma relies on cellular and circuit mechanisms that restore excitatory neurotransmission through connections between the cortex, thalamocortical system, and thalamus. DBS can be used as a surgical tool to provide adjustable stimulation to restore dysfunctional circuits while compensating for lost arousal regulation typically controlled by intact frontal lobes.

How DBS acts ?

Central thalamus: a target for consciousness restoration The underlying mechanisms of DBS are not yet fully understood. Despite the apparent diversity in the targets of DBS, all published studies have consistently observed modulation of the same pathway, simplifying the interpretation of overall results [24]. The primary target is the Central Thalamus (CT), to elicit excitation of the thalamo-cortical projection. The CT encompasses various intra- and paralaminar

thalamic nuclei situated between the brainstem/basal forebrain arousal systems and the cortex (Fig. 1.14). Neurons within the central thalamus play a crucial role in regulating arousal through their anatomical connections with large-scale cortical networks [212]. Typically, electrodes are implanted in the intralaminar nuclei due to their apparent association with recovery in patients with DoCs and the pathophysiological mechanisms linked to brain injury and cellular loss in the central thalamus [214, 236, 140, 32]. In some studies, stimulation is applied unilaterally, while in others, it is bilateral [236, 24, 32].

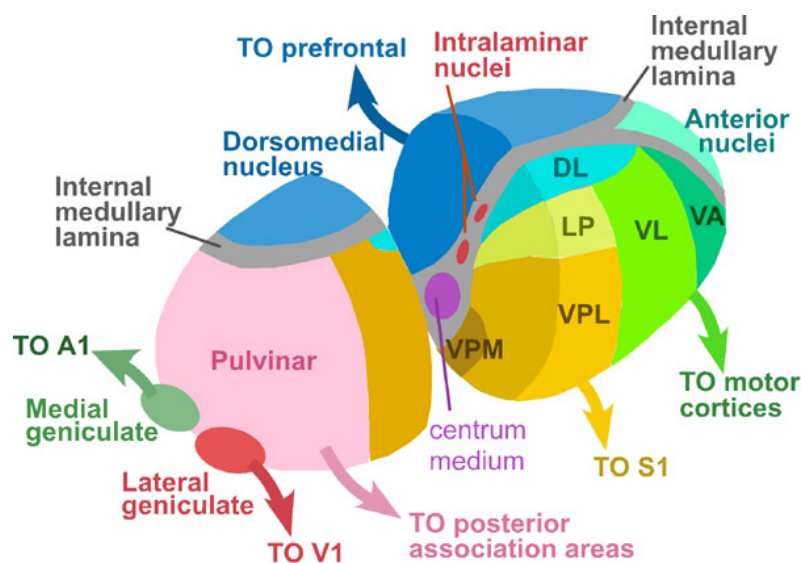


Figure 1.14: Overview of the primate thalamus. From [88].

Parameters of stimulation Most studies predominantly employed low-frequency stimulation (up to 50 Hz), although some investigations also explored high-frequency stimulations (up to 100 Hz). However, the influence of stimulation parameters on clinical outcomes remains uncertain [24].

Encouraging but variable responses across the DoC population Based on the comprehensive review by Bourdillon *et al.* (2019) [24], a positive response was observed in 30 out of 67 patients diagnosed with UWS and in 6 out of 11 patients diagnosed with MCS. However, the definition of "response" varies significantly across studies, reflecting the evolution of outcome measures since the 1970s. Nonetheless, the clinical descriptions provided in older studies consistently align with improvements on the Coma Recovery Scale-Revised (CRS-R) [24]. Detailed results from the studies cited in the review are summarized in Table 1.15. The etiologies of DoC included TBI (27 patients), anoxic causes (12 patients), and vascular causes (13 patients). However, it remains unclear from the literature whether etiology serves as a predictive factor for outcomes.

In addition to the existing review, recent findings from *Chudy et al.* (2023) [40] contribute valuable insights. In their study, 32 patients were implanted (27 with UWS and 5 with MCS). The stimulation target was the centromedian-parafascicular complex in the left hemisphere for patients with hypoxic brain lesions or the better-preserved hemisphere for those with TBI. Consciousness levels improved in 7 patients, with 3 out of 5 MCS patients transitioning to full awareness, demonstrating the ability to interact and communicate. Two of them can live largely independently. Among the 27 UWS patients, 4 showed improvement in consciousness, with 2 emerging to full awareness and the other 2 progressing to MCS. Notably, spontaneous recovery is rare in patients with DoC lasting longer than 12 months following TBI or 6 months following anoxic-ischemic brain lesions.

Promising responses in animals *Alkire et al.* [6, 5, 143], using rodent models, initially showcased that anesthesia-induced loss of consciousness could be reversed through direct manipulation of the central medial thalamus. In NHP, electrical stimulation of CT neurons was observed to heighten arousal and improve cognitive responses to visuomotor tasks [10], ultimately reversing the effects of anesthesia [15]. *Redinbaugh et al.* [202] conducted electrical stimulation of the central lateral thalamus in anesthetized macaques, successfully modulating vigilance by regulating cortical interactions in specific layers. They showed that thalamic stimulation reinstated coherence between frontoparietal regions in both the feedforward and feedback pathways. Recently, *Tasserie et al.* [234] demonstrated that DBS targeting the CT restored both arousal and conscious access, following loss of consciousness, thereby laying the groundwork for its therapeutic translation in patients with DoC.

1.3.3 . Challenges

Implementation complexity Conducting studies involving comatose patients is inherently complex, as evidenced by a case of patient selection in a well-designed prospective open-label study spanning seven years. Out of 40 patients considered, only five (13%) met the inclusion criteria, which included parameters such as EEG desynchronized activity being less than 5% of the recorded time, as well as the presence of somatosensory and auditory evoked potentials on at least one side. However, out of the five eligible patients, two were unable to undergo surgery due to issues with legal representation [156, 236].

Side effects Severe adverse effects have been documented in DBS, as outlined in Table 1.15 [24]. An inherent drawback of current DBS methods is the requirement to insert an electrode through the scalp, skull, and brain, which carries various risks [197].

Study	Design/Control	Population	Target/Stimulation parameters	Behavioral effects	Electrophysiological/metabolic effects	Side effects
McLardy et al., 1968	Case report/ None	1 (considered as) VS/UWS	Left thalamus; midbrain (intralaminar nuclei/reticular formation) / 250Hz, 1ms	No modifications of consciousness, left hand spontaneous movement	No post procedure electrophysiological nor metabolic evaluation available	None
Hassler et al., 1969	Case report/ None	1 (considered as) VS/UWS	Left ventral anterior thalamus; right pallidum / Left, 25-30Hz, 20V, 1-3ms; Right 8Hz, 30V, 1-3ms	"Improvement" of consciousness, vocalizations, left limbs spontaneous movement	EEG recordings showed a disappearance of a unilateral delta focus which is replaced by an alpha activity	None
Tsubokawa et al., 1990	Open-label/ None	8 patients (VS/UWS)	Central thalamic nuclei; nucleus cuneiformis (reticular formation)/50 Hz, 0-10 V	4 recoveries (PCS 2-4 = > 8-9) 1 responder (PCS 2-4 = > 7) 3 failures (PCS 2-4 = > 3-5)	Increase of spectral power and desynchronization on EEG in the 4 patients who recovered/Increase on the brain perfusion on MRI in these patients	None
Cohadon and Richer, 1993	Open-label/ None	25 patients (VS/UWS)	Central nucleus of the thalamus/50 Hz, 5-10 V, 5 ms	1 moderate disabilities (GOS) 10 severe disabilities (GOS) 12 no effect (2 patients died before the endpoint)	No post procedure electrophysiological nor metabolic evaluation available	2 died (unrelated to surgical procedure)
Schiff et al., 2007	Case report, Cross-over RCT/ Sham	1 MCS	Anterior intralaminar thalamic nuclei / 100Hz, 4V	Fluctuant increase in CRS-R subscales, better feeding and motor behaviors, restoration of communication	No post procedure electrophysiological nor metabolic evaluation available	None
Yamamoto et al., 2010 (includes publications since 2002)	Open-label/ None	21 patients (VS/UWS)	Centro-median nucleus of the thalamus; midbrain (reticular formation) / 25Hz, various intensities	8 became MCS or EMCS 13 remain VS/UWS	The 8 patients who recovered from VS showed desynchronization on continuous EEG frequency analysis/Increase on the brain perfusion on MRI in these patients	None
Wojtecki et al., 2014	Case report/ None	1 MCS	Internal medullary lamina; nuclei reticularis thalami/70-250 Hz, various intensities	No modifications of consciousness	Modulation of oscillatory activity in the beta and theta band within the central thalamus accompanied by an increase in thalamocortical coherence in the theta band	None
Magrassi et al., 2016	Open-label/ None	3 patients (1 MCS, 2 VS/UWS)	Anterior intralaminar nuclei; paralaminar Areas/80-110 Hz, various intensities	Increase of CRS-R in all of the 3 patients: 14 = > c15 8 = > 11 6 = > 9	Increase of theta and gamma power spectrum in EEG after 1 month of stimulation. No modifications of the evoked potentials.	1 postoperative intraparenchymal hematoma
Adams et al., 2016	Case report/ None	1 MCS	Anterior intralaminar thalamic nuclei/100 Hz, 4 V	Variable increase of CRS-R (11-14)	Long term re-emergence of sleep patterns	None
Chudy et al., 2018	Open-label/ None	14 patients (4 MCS, 10 VS/UWS)	Central thalamic nuclei / 25 Hz, 2.5-3.5 V, 90 μ s	3 MCS became EMCS; 1 VS became MCS; 7 had no improvement of consciousness (3 patients died before the endpoint)	No post procedure electrophysiological nor metabolic evaluation available	3 died (unrelated to surgical procedure)
Lemaire et al., 2018	Cross-over RCT/ Sham	5 patients (4 MCS, 1 VS/UWS)	Dual pallido-thalamic / 30-Hz, 6V, 60 μ s	1 VS/UWS and 1 MCS had a significant improvement of the CRS-R.	The metabolism of the medial cortices increased specifically in the two responders	1 postoperative bronchopulmonary infection

CRS-R, Coma Recovery Scale – Revised; DoC, disorders of consciousness; EEG, electroencephalogram; EMCS, Emergence from Minimally Conscious State; GOS, Glasgow Outcome Scale; MCS, Minimally Conscious State; PCS, Prolonged Coma Scale; RCT, randomized controlled trial; VS/UWS, vegetative state/unresponsive wakefulness syndrome.

Figure 1.15: DBS studies in DoC patients. From [24].

No sham-controlled trial No sham-controlled trial investigating DBS in patients with DoCs has been published. There remains a necessity to establish a treatment protocol that evaluates the generalizable effects of DBS using standardized criteria. Moreover, numerous clinical and ethical concerns, such as the risk of infection and resulting clinical deterioration, still require attention [236]. Animal studies can offer a promising alternative for addressing this challenge, as demonstrated by the recent study in NHP [234], which allows for control over both the stimulated region and the stimulation parameters.

Could DBS be truly attributed as the catalyst for recovery? One of the most significant critiques of the published studies concerns the timeframe. Spontaneous recovery from non-anoxic UWS lasting longer than 1 month occurs in 30% of patients at 6 months and in 43% at 12 months. This observation extends beyond UWS, as 83% of patients emerged from MCS after 6 months. However, most studies report DBS performed within the year following the brain injury, so in the 29 out of the 41 patients who improved after DBS, spontaneous recovery cannot be excluded [24].

In the review by *Thibaut et al.* (2019)[236], the same observation is made: untangling the impacts of DBS from spontaneous recovery is a difficulty due to the enrollment of patients 2–11 months after injury.

For this challenge, the proxy through LOC via anesthesia ensures that it is not a spontaneous recovery of consciousness.

Physio-pathological heterogeneity Another constraint lies in patient selection based on clinical criteria [24]. Diverse lesions in the central nervous system can yield identical clinical manifestations. For instance, UWS can arise from diffuse cortical lesions or from a highly focal lesion in the brainstem of the ARAS. In the former scenario, DBS will modulate a damaged cortex with compromised long-distance synchronization capacity, whereas in the latter, thalamic modulation will affect an intact cortex [24]. Recent studies tend to address this issue by excluding anoxic causes or attempting to identify the potential connectivity that DBS may restore. However, most studies amalgamate patients with similar clinical presentations but with potentially significant physio-pathological heterogeneity. Studying connectivity following consciousness restoration is an intriguing avenue to identify the exact connectivity that DBS can restore, thereby customizing treatments [24]. Additionally, the proxy of LOC induced by anesthesia is also valuable in studying cases of consciousness restoration as it helps overcoming lesion heterogeneity.

Consciousness, like happiness according to Balzac, is a soap bubble that changes color as the iris and that breaks when touched. Consciousness can be had, lost or covered up. Anesthesia serves as a potent method for simulating altered consciousness in NHP. Electrical neuromodulation through DBS represents a robust approach for inducing large-scale cortical reconfiguration. Thalamic DBS not only reinstates vigilance but also enhances cortical responses in the fronto-parieto-cingulate network for higher-order processing. DBS holds promise for restoring wakefulness and awareness in patients with DoCs. The optimal target for DBS may vary among individuals, and in some cases, stimulating multiple targets could be considered for improved outcomes. Despite its therapeutic potential, the cortical effects of DBS remain largely unknown, and target selection is currently based on empirical evidence. Future research should strive to unravel the underlying mechanisms and functional connectivity associated with each target. Moreover, progress in neuroimaging techniques like fMRI and EEG holds potential for identifying patient-specific biomarkers, that could help in treatment planning and target selection [236, 32].

2 - fMRI: the brain's cloak of visibility

Contents

2.1	Theory	45
2.1.1	Magnetic Resonance Imaging : from magnet to image	46
2.1.2	Functional MRI (fMRI)	47
2.1.3	fMRI parameters	48
2.1.4	Contrast agents in fMRI	50
2.1.5	Resting-state fMRI	51
2.1.6	Relation with neuronal activity	52
2.1.7	Benefits, drawbacks and comparison	52
2.2	Datasets	54
2.2.1	Anesthesia dataset	54
2.2.2	DBS dataset	57
2.2.3	External resources: ROI template & reference anatomical connectivity	62
2.2.4	Functional connectivity (FC)	63
2.3	Limits and associated challenges	67
2.3.1	The limits of acquisition	68
2.3.2	The limits of preprocessing	69
2.3.3	The limits of connectivity computation	71

As we have seen, brain exploration using neuroimaging tools has enabled us to highlight the neural correlates of consciousness. But how exactly do we make the brain talk? How does neuroimaging work, and why does it allow us to read through the cranium as if through a crystal ball?

2.1 . Theory

Functional Magnetic Resonance Imaging (fMRI) has a relatively recent history. It was developed barely 30 years ago, but has grown rapidly. According to [195], by 1996 it was possible to have read all the fMRI literature in a week, whereas at the time the authors were writing, it was barely possible to read all the papers published the week before. A little over 10 years later, fMRI has not been shelved, as one might have thought would happen with a trendy new object that is only

considered for a short while. It continues to be the talk of the town, at a steady rate of 6,200 papers a year for the past 5 years (Fig. 2.1).

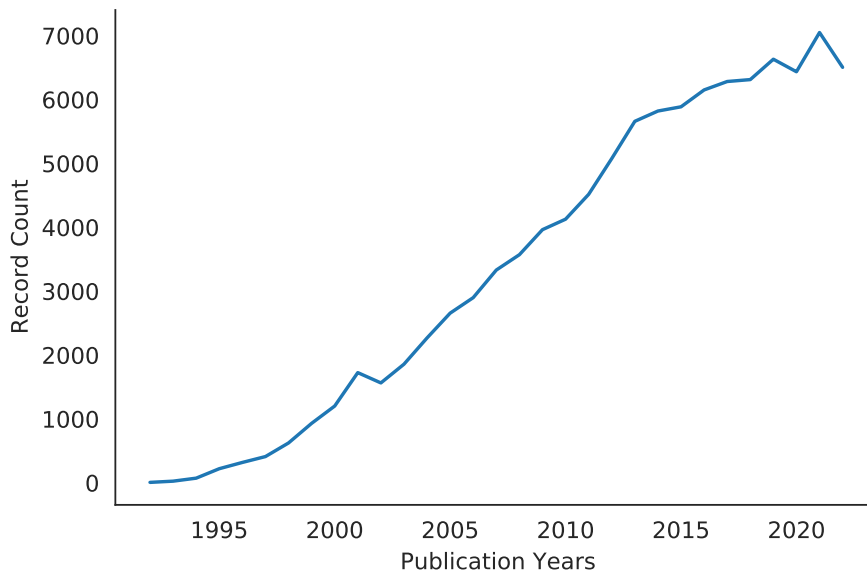


Figure 2.1: Number of citations in the WebOfScience database matching the query [“fMRI” OR “functional MRI” OR “functional magnetic resonance imaging”] for every year since 1992.

2.1.1 . Magnetic Resonance Imaging : from magnet to image

The concept of fMRI relies on the technology of MRI scanning and the understanding of the properties of oxygen-rich blood.

MRI brain scans utilize a robust, permanent, static magnetic field to align nuclei in the specific brain region under examination. The primary magnet generates the static magnetic field responsible for the observable macroscopic magnetization. Predominantly utilized, superconducting electromagnets constitute the most common type of magnets. These electromagnets involve a coil rendered superconducting through liquid helium cooling. Gradient fields are additional magnetic fields applied along the main magnetic field’s x, y, and z axes. These gradient fields help to spatially encode the signals emitted by the protons in the patient’s body. These gradient fields, in combination with the main magnetic field and RF pulses, allow MRI scanners to precisely localize the signals emitted by the protons within the patient’s body. This enables the creation of detailed 3D images [104]. Different tissues in the body have varying relaxation times, and these differences contribute to the contrast observed in the final image.

Technically, during an MRI scan, complex data is acquired in the frequency

domain (and stored in the so-called k-space matrix). The Fourier transform allows this frequency-encoded spatial information to be converted into an interpretable image [229]. Other reconstruction algorithms can be applied to the k-space matrix to generate the final image [229]. Finally, MRI is a non-invasive technique that provides an in-vivo structural view of the brain with excellent spatial resolution and soft tissue contrast [104].

2.1.2 . Functional MRI (fMRI)

The primary motivation behind fMRI was to expand MRI capabilities to capture functional changes in the brain triggered by neuronal activity. It is thus a specialized MRI technique. fMRI enables the visualization of brain areas activated during the execution of a task (involving motor, sensory, or cognitive functions) or during rest. Although it does not directly measure neuronal activity, it highlights changes in blood flow associated with this activity [116]. Its principle is based on the Blood Oxygenation Level Dependent (BOLD) method, which represents variations in Cerebral Blood Volume (CBV). When cerebral activation occurs, there is a significant local increase in cerebral blood flow and an increased oxygen consumption, although proportionally less (*neurovascular coupling*). This results in an excess of oxyhemoglobin (HbO₂) (oxygenated form of hemoglobin) in the venous capillaries of the activated area, leading to a relative decrease in deoxyhemoglobin (Hb) concentration (cf. Fig. 2.2). Deoxyhemoglobin exhibits *paramagnetic* properties, possessing a higher magnetic susceptibility compared to the *diamagnetic* oxyhemoglobin. This paramagnetic nature typically leads to a signal decrease due to magnetic susceptibility effects.

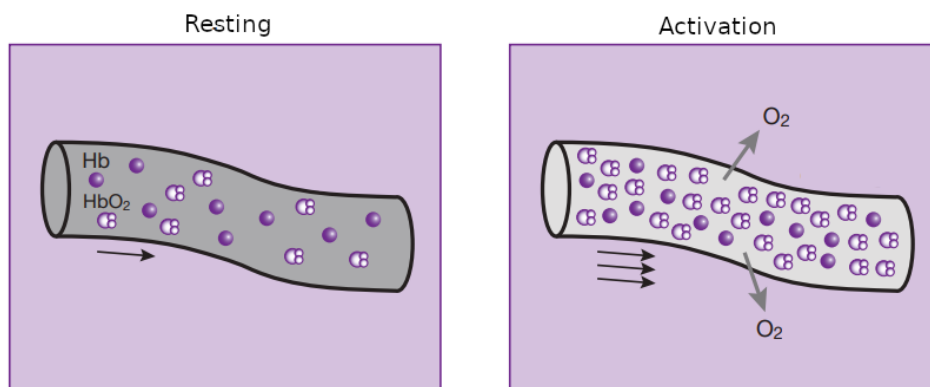


Figure 2.2: Principle of the BOLD effect. Adapted from [116]

Consequently, the reduction in deoxyhemoglobin concentration induces a modest signal increase in the activated region in T₂*-weighted sequences, attributable to the prolonged T₂* in blood vessels. As illustrated in Fig. 2.3, the BOLD signal (depicted in blue) doesn't exhibit an instantaneous increase and doesn't promptly

return to the baseline once the stimulus concludes (depicted in red). Due to the relatively gradual changes in blood flow evolving over several seconds, the BOLD signal provides a blurred and delayed representation of the original neural signal. The Hemodynamic Response Function (HRF) can be conceptualized as the ideal, noiseless reaction to an infinitesimally brief stimulus. The HRF is a critical component of fMRI data analysis because it represents the relationship between neural activity and the resulting hemodynamic response. The shape and timing of the HRF can vary between brain regions and individuals and can be influenced by factors such as age, health and cognitive state. However, for a given analysis, the HRF is often fixed. It extends over a period of 10 to 15 seconds, gradually rising, peaking at 4 to 6 seconds, and subsequently declining.

Neurovascular coupling (NVC)

Process by which the supply of oxygen and nutrients is adjusted to neuronal activity through the regulation of blood flow.

Paramagnetism

Type of magnetism of a material medium which is not magnetized in the absence of a magnetic field, but which acquires, under the effect of such a field, a magnetization oriented in the same direction as the field.

Diamagnetism

Behavior of materials which, when subjected to a magnetic field, creates a very weak magnetization opposite to the external field, and therefore generates a magnetic field opposite to the external field.

The signal enhancement observed in activated areas is minimal (around 2 to 5%), necessitating many measurements. The "BOLD contrast" amplifies with the B0 magnetic field's intensity, as well as the SNR, spatial resolution, scanning efficiency and susceptibility artifacts. The most widely used sequence is Echo Planar Imaging (EPI). EPI allows for whole-brain coverage with relatively short acquisition times, making it well-suited for studying dynamic brain processes such as task-evoked activations and resting-state networks.

2.1.3 . fMRI parameters

An MRI sequence is a set of excitation pulses whose parameters (TE, TR) are adjusted to obtain images with a given contrast (T1, T2, T2*). For fMRI, the slight transient increase in BOLD signal can be detected in T2* weighting. A description of all fMRI parameters is beyond the scope of this thesis, but a quick overview of the main parameters used is given here.

The fundamental parameter for time resolution, known as the sampling time, is denoted as Repetition Time (TR). TR determines how frequently a specific brain slice is stimulated and permitted to lose its magnetization. TRs can range from

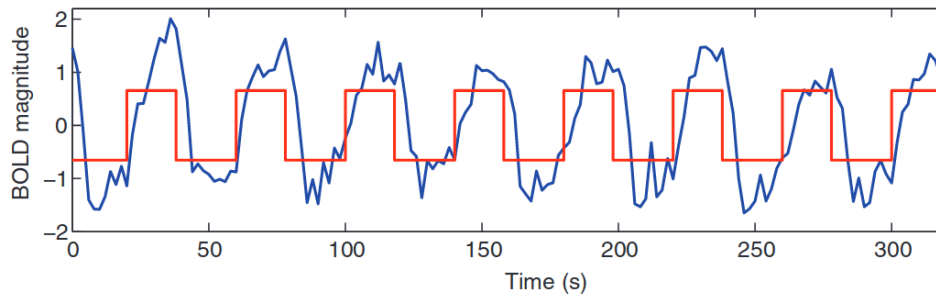


Figure 2.3: Visualization of time series data in BOLD fMRI for an active voxel and representation of a signal used for modeling the response. The BOLD signal from an active voxel is depicted in blue, alongside the stimulus time series in red. From [195]

very short intervals (e.g., 500 ms) to much longer durations (up to 3 s). Changes in the blood-flow and vascular systems integrate responses to neuronal activity over time. Since this response represents a smooth continuous function, employing ever-faster TRs doesn't provide additional benefits; it merely yields more points on the response curve, which can be obtained through simple linear interpolation anyway.

The Echo Time (TE) is the time interval between excitation and the occurrence of the MRI signal.

T2 relaxation is associated with the presence of "molecular" origin field inhomogeneities (small local magnetic fields that overlap with B0, causing the dephasing of protons or spins), responsible for the "irreversible" decay of transverse magnetization of the Nuclear Magnetic Resonance (NMR) signal (*Free Induction Decay (FID)*). If the magnetic field B0 of the magnet were perfectly homogeneous, we would observe a signal decay according to a decreasing exponential in T2. In reality, on a macroscopic scale, the magnetic field B0 of the magnet can be considered fairly homogeneous, but on a microscopic scale, it is not: these inhomogeneities in the B0 field of "instrumental" or "inherent" origin are constant and lead to further dephasing of spins. Thus, the observed FID signal is linked both to the inhomogeneities of the B0 field of "molecular" origin (T2), to which the inherent (constant) inhomogeneities of the external magnetic field B0 are added; the symbol T2* is used to represent the combination of these two effects (cf. Fig 2.4). Therefore, the FID signal decreases more rapidly than expected according to an exponential in T2* (and not in T2) [116].

Free Induction Decay (FID)

Observable NMR signal generated by non-equilibrium nuclear spin magnetization precessing about the magnetic field (conventionally along z).

2.1.4 . Contrast agents in fMRI

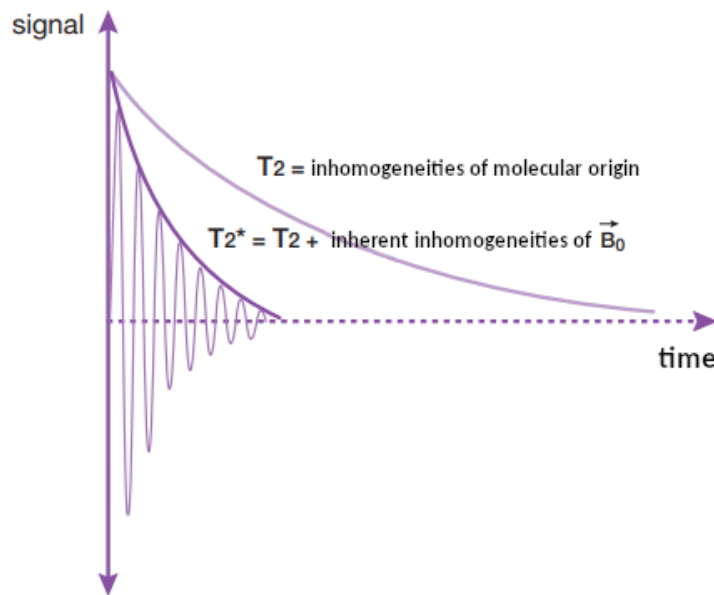


Figure 2.4: T_2^* principle. Adapted from [116]

Exogenous contrast agents are artificial substances administered, usually intravenously, to modify the contrast of vessels and organs. MRI contrast agents are not directly visible. The shortening of their T1 and/or T2 relaxation times on nearby hydrogen nuclei is responsible for the contrast modification [100]. We can distinguish two main classes of contrast agents: if the contrast agent shortens the T1 time (paramagnetic contrast agents), a T1-weighted hypersignal is observed. If it shortens T2 (super-paramagnetic contrast agents), on the other hand, we'll see a reduction in T2 and T2*-weighted signal. For the fMRI signal, if contrast agents are used then these are the ones. These are superparamagnetic ferrite particles (Superparamagnetic Iron Oxide (SPIO) and Ultrasmall Superparamagnetic Iron Oxide (USPIO)) [100].

These nanoparticles are accepted for punctual diagnostic use, the toxicity of the iron they contain being considered negligible as it is eliminated via the normal endogenous iron cycle. They are mainly used in humans as contrast agents to image vascular lesions, tumors and lymph nodes [241].

Monocrystalline Iron Oxide Nanoparticles (MION) are a subset of USPIO (10–30 nm diameter). In addition to their use in humans for specific images, MION have been utilized in anesthetized rodents to amplify fMRI sensitivity and investigate the CBV physiology in connection with the BOLD signal following neuronal activation. *Leite et al.* [137] showed the benefits of employing an exogenous agent for repetitive neuroimaging in awake, nonhuman primates using a clinical 3 Tesla scanner. A MION solution was administered in two macaque monkeys.

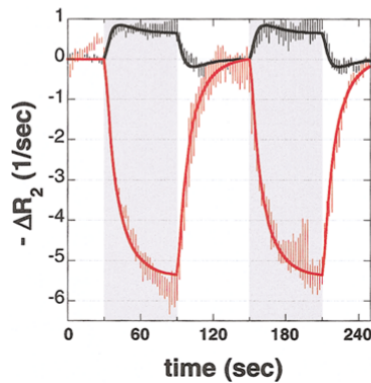


Figure 2.5: Alterations in relaxation rates for BOLD (depicted in black) and MION (depicted in red) contrasts, along with their respective linear model fits, are illustrated over two cycles of 60 seconds of stimulus (indicated by gray shaded intervals) followed by 60 seconds of baseline. From [137]

No adverse behavioral effects attributable to the contrast agent were observed in either monkey. In comparison to BOLD imaging at 3 Tesla, MION enhanced functional sensitivity by an average factor of 3 across the entire brain for a stimulus of prolonged duration (cf. Fig. 2.5). Overall, the contrast agent yielded a significant enhancement in functional brain imaging outcomes in awake, behaving primates at this field strength.

2.1.5 . Resting-state fMRI

The inception of Resting-State fMRI (RS-fMRI) can be traced back to the research conducted by *Biswal et al.* [20], where they showcased highly correlated low-frequency (<0.1 Hz) variations in BOLD signal between sensorimotor and supplementary motor cortices on both sides in individuals at rest. They also observed synchronous fluctuations in the auditory and visual systems, identifying these patterns as manifestations of the brain's functional connectivity. RS-fMRI resembles conventional task-fMRI but doesn't necessitate subjects to engage in a task or respond to stimuli. Subjects simply recline in the scanner for 5-10 minutes, either with their eyes closed or fixed on a point, while comprehensive whole-brain BOLD data is collected. The task-free nature of RS-fMRI allows it to be applied to diverse subjects, including infants, children, individuals with neurological disorders, patients under anesthesia, and even animals.

RS-fMRI has led to the identification of at least 20 distinct patterns of brain connections known as Resting State Network (RSN)s. Among the most notable are the default mode network (most active at rest, associated with introspection and mind wandering), networks for visual and auditory processing, executive control, dorsal attention, and salience (identification of unusual/remarkable events) (cf. Fig. 2.6). These networks have yielded valuable insights into the cognitive organization of the brain in both health and disease.

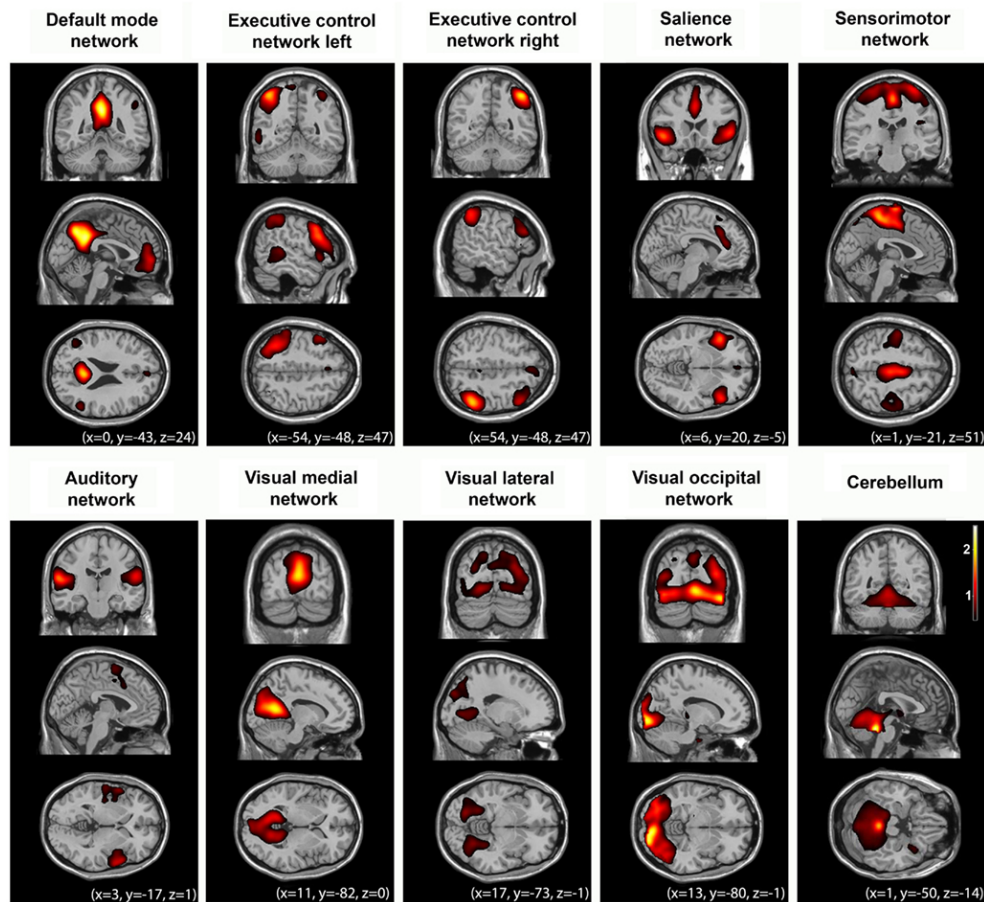


Figure 2.6: Example of resting-state networks in healthy controls ($N = 10$) during the typical wakeful resting state, utilizing Independent Component Analysis. To illustrate, group-level spatial maps (z values) are superimposed on a structural T1 magnetic resonance template, where x , y , and z values denote the Montreal Neurological Institute coordinates of the depicted sections. From [96]

2.1.6 . Relation with neuronal activity

The BOLD signal often corresponds relatively closely to the Local Field Potential (LFP), the electrical field potential surrounding a group of cells. In many cases, neuronal discharges, the local potential and the BOLD signal are closely correlated [147, 178].

2.1.7 . Benefits, drawbacks and comparison

MRI has the advantage of being non-invasive and painless. It has a relatively high spatial resolution (the typical fMRI voxel size is 1.5-4 mm or even less with higher field magnets) [83]. On the other hand, the acquisition time is somewhat long (between 20 and 40 minutes) and unpleasant due to the repetitive noise inside the device and the cramped conditions. Any ferromagnetic object in the body

is also potentially dangerous (prosthetic devices, pacemakers, metal splinters, especially intraocular ones, projectiles (bullets, shell fragments), so there are a few contraindications. The "missile" effect of extracorporeal metal objects attracted at high speed into the magnet is also potentially very dangerous. A significant limitation of fMRI lies in its modest temporal resolution. The temporal resolution of fMRI is constrained by the hemodynamic response time, which is considerably slower than the underlying neural processes, resulting in substantial blurring of temporal information [83]. Furthermore, fMRI is dependent on neurovascular coupling, and therefore provides indirect information on neuronal activity, which can sometimes complicate the interpretation of results obtained in the case of pathologies [87]. Finally, its primary drawbacks both for preclinical and clinical imaging are its portability, machine costs, maintenance and accessibility [57].

Is the grass greener elsewhere?

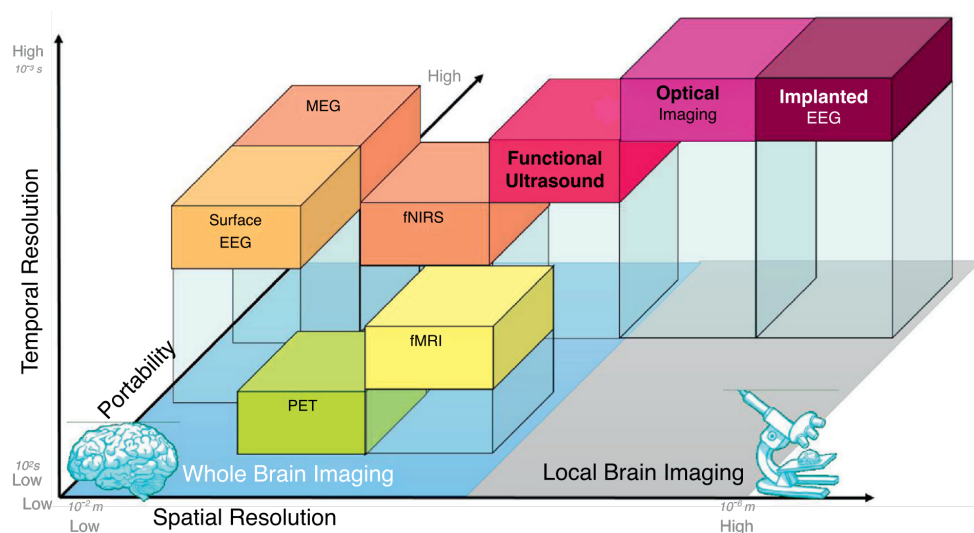


Figure 2.7: Main brain functional imaging technique resolutions. From [57]

Positron emission tomography (PET) stands as an additional functional imaging method employing injected radioactive and biologically active tracers. This approach allows the visualization of brain molecular processes, providing insight into the consumption of glucose associated with brain metabolism. PET calculates a 3D reconstruction of the concentration of positron-emitting radionuclides from the pairs of indirectly emitted gamma rays [57]. EEG and Magnetoencephalography (MEG) are two other non-invasive electromagnetic techniques that measure electric potential and magnetic field respectively [260] with unrivalled temporal resolution.

PET scans provide relatively similar information to fMRI in terms of brain mapping, but at the cost of more pronounced irradiation and invasiveness. Spatial and

temporal resolution are also reduced (cf. Fig. 2.7). MEG and EEG have difficulty highlighting sources of activity located deep in the brain. Spatial resolution is also poor (cf. Fig. 2.7).

Combining these hemodynamic and electromagnetic measurements therefore represents a solution for studying brain activations with high spatial and temporal resolution.

2.2 . Datasets

This study uses two previously acquired RS-fMRI datasets: the anesthesia dataset [14, 245] and the DBS dataset [234]. The data were acquired for a previous project with the goal of discovering a new signature of anesthesia-induced loss of consciousness and consciousness restoration. In the following work, we propose a retrospective analysis of these data. In this study, we strive to use the same data without additional experiments, to maximize their use, and to shed new light on them, while being aware of the ethical issues associated with the data acquisition in animals, and even more so in NHP.

All procedures were conducted in accordance with the European convention for animal care (86-406) and the National Institutes of Health's Guide for the Care and Use of Laboratory Animals. Animal studies were approved by the Institutional Ethical Committee (CETEA protocols #10-003 #12-086 #12-086 and #16-040) (for details of the anesthesia protocol, see [245, 234]).

2.2.1 . Anesthesia dataset

RS-fMRI data were collected from rhesus macaques either in the awake state or under anesthesia (deep ketamine, moderate/deep propofol, or moderate/deep sevoflurane). Two different levels of anesthesia were considered, either moderate sedation or deep sedation equivalent to general anesthesia. Three monkeys were scanned for each arousal state. 156 RS-fMRI runs of 500 volumes each were acquired on a 3T scanner with a repetition time of 2400 ms. Study design is summarized in Figure 2.8.

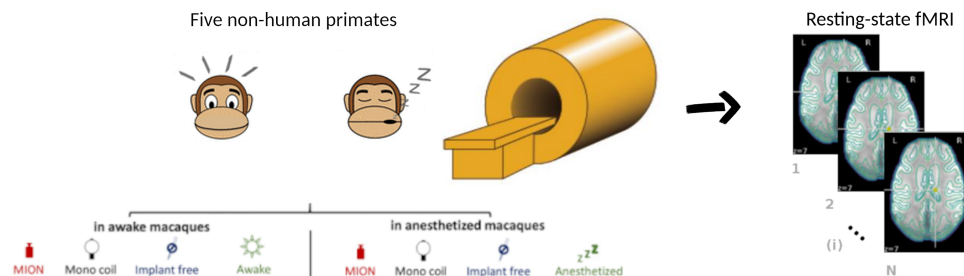


Figure 2.8: *Experimental design anesthesia dataset.*

Animals

Acquisitions involved five rhesus macaques (*Macaca mulatta*), comprising one male (designated as monkey J) and four females (monkeys A, K, Ki, and R). The monkeys, weighing between 5 to 8 kg and aged 8 to 12 years, were distributed across three arousal states: awake (monkeys A, K, and J), propofol anesthesia (monkeys K, R, and J), ketamine anesthesia (monkeys K, R, and Ki), and sevoflurane anesthesia (monkeys Ki, R, and J) [245]. It is important to note that NHP studies often involve a limited number of animals.

Awake protocol

For the awake condition, monkeys were implanted with a magnetic resonance-compatible headpost and trained to sit in the sphinx position in a primate chair [243]. Monkeys sat in the dark inside the MRI scanner without any task [14, 245]. The eye position was monitored.

Anesthesia protocol

Monkeys underwent scans in an awake resting state and under various levels and types of sedation, including ketamine [244], propofol [14, 244], or sevoflurane anesthesia [245]. Anesthesia levels were determined using the monkey sedation scale, considering spontaneous movements, responses to external stimuli, and continuous EEG monitoring [244] at the beginning and the end of the scanning session and continuous EEG monitoring [244] (cf. Appendix 1). Monkeys in an awake state responded positively to all stimuli. Under ketamine, deep propofol anesthesia, and deep sevoflurane anesthesia, monkeys ceased responding to stimuli, reaching a general anesthesia state.

Monkeys (K, R, and Ki) undergoing deep ketamine anesthesia [244] were administered an intramuscular ketamine injection (20 mg/kg) for anesthesia induction. This was followed by a continuous intravenous infusion of ketamine (15 to 16 mg · kg⁻¹ · h⁻¹) to maintain anesthesia. Atropine (0.02 mg/kg intramuscularly) was administered 10 minutes before induction to reduce salivary and bronchial secretions. For propofol anesthesia, monkeys (K, R, and J) were scanned during both moderate propofol sedation and deep propofol anesthesia (equivalent to general anesthesia) [14]. Monkeys were trained to receive an intravenous propofol bolus (5 to 7.5 mg/kg) for anesthesia induction. Induction was followed by target-controlled infusion of propofol (moderate propofol sedation, 3.7 to 4.0 μg/ml; deep propofol anesthesia, 5.6 to 7.2 μg/ml). For sevoflurane anesthesia, monkeys (Ki, R, and J) underwent scans during both moderate and deep sevoflurane anesthesia. Monkeys received an intramuscular ketamine injection (20 mg/kg) for anesthesia induction, followed by sevoflurane anesthesia (moderate sevoflurane anesthesia, deep sevoflurane anesthesia, sevoflurane inspiratory/expiratory, 4.4/4.0 volume percent). At

least 80 minutes were waited before initiating sevoflurane anesthesia scanning sessions to allow for a washout of the initial ketamine injection [245].

To mitigate potential motion-related artifacts during magnetic resonance imaging, a muscle-blocking agent (cisatracurium, 0.15 mg/kg bolus intravenously, followed by continuous intravenous infusion at a rate of $0.18 \text{ mg} \cdot \text{kg}^{-1} \cdot \text{h}^{-1}$) was coadministered during the ketamine and moderate propofol sedation sessions. In all anesthesia experiments (ketamine, moderate and deep sevoflurane, moderate and deep propofol anesthesia), monkeys were intubated and ventilated with specified parameters [14]. Physiological monitoring encompassed heart rate, noninvasive blood pressure, oxygen saturation (SpO₂), respiratory rate, end-tidal CO₂ (EtCO₂), and cutaneous temperature. Intravenous hydration comprised a mixture of normal saline (0.9%) and 5% glucose (250 mL of normal saline with 100 mL of 5% glucose) at a rate of 10 mL/kg/h. After each fMRI session, anesthesia was stopped, and the animal was carefully monitored during recovery, eventually placed in individual housing and observed until fully recovered from anesthesia. Animals were positioned in a sphinx position, mechanically ventilated, and their physiologic parameters were monitored.

fMRI data acquisition

The data were gathered between July 2011 and August 2016 [14, 245]. Monkeys were scanned using a 3-Tesla horizontal scanner (Siemens Tim Trio, Germany) with a single transmit-receiver surface coil customized for monkeys. Each functional scan comprised gradient-echo planar whole-brain images (repetition time = 2,400 ms; echo time = 20 ms; 1.5-mm³ voxel size; 500 brain volumes per run). Before each scanning session, a contrast agent, MION (Feraheme; 10 mg/kg, intravenous), was injected into the monkey's saphenous vein [14].

In total, 156 runs were acquired: 31 awake runs (monkey A, 4 runs; monkey J, 18 runs; monkey K, 9 runs), 25 ketamine anesthesia runs (monkey K, 8 runs; monkey Ki, 7 runs; monkey R, 10 runs), 25 moderate sevoflurane sedation runs (monkey J, 5 runs; monkey Ki, 10 runs; monkey R, 10 runs), 20 deep sevoflurane anesthesia runs (monkey J, 2 runs; monkey Ki, 8 runs; monkey R, 11 runs), 25 moderate propofol sedation runs (monkey J, 2 runs; monkey K, 10 runs; monkey R, 12 runs), and 30 deep propofol anesthesia runs (monkey J, 8 runs; monkey K, 10 runs; monkey R, 12 runs) [245] (for summary cf. Table 2.1).

monkey	awake	moderate propofol	deep propofol	moderate sevoflurane	deep sevoflurane	ketamine	
J	18	2	8	5	2	-	35
A	4	-	-	-	-	-	4
K	9	11	10	-	-	8	38
Ki	-	-	-	10	8	7	25
R	-	12	12	10	10	10	54
	31	25	30	25	20	25	156

Table 2.1: Description of the acquisition conditions across NHP.

fMRI preprocessing

Functional images underwent reorientation, realignment, and rigid coregistration to the anatomical template of the monkey Montreal Neurologic Institute (MNI) space using the NeuroSpin Monkey (NSM) preprocessing [233, 243, 245]. Regression was applied to eliminate the *global signal* from the images -addressing potential confounding effects related to physiological changes (e.g., respiratory or cardiac changes)- and movement parameters -resulting from rigid body correction for head motion [73]. Voxel time series underwent filtering with lowpass (0.05-Hz cutoff) and high-pass (0.0025-Hz cutoff) filters, along with a zero-phase fast-Fourier notch filter (0.03 Hz) to eliminate an artifactual pure frequency present in all the data [14, 245].

Global signal

Mean time course computed over all voxels within the brain [255]. Global signal regression (GSR) is often used in RS-fMRI. The utilization of GSR has been identified as significantly enhancing the functional specificity of resting-state correlation maps, mitigating the impact of motion on functional connectivity estimates and being effective in eliminating global artifacts arising from motion, cardiac activity, and respiratory activity [145].

2.2.2 . DBS dataset

RS-fMRI data were collected from rhesus macaques either in the awake state or under anesthesia with a DBS electrode implanted. Three subjects were included for the awake (non-DBS) experiments and two for the DBS experiments. Two different levels of DBS (low 3V or high 5V) were applied during the entire run on either the Central Thalamus (CT) or VentroLateral (VL) thalamic nucleus. We aimed at stimulating the CT as a main DBS target and the VL as a control target (cf. Figure 2.9). 199 resting-state runs of 500 brain volumes each were acquired on a 3T scanner, with a TR of 1250 ms [234]. Study design is summarized in Figure 2.9.

Animals

Five male rhesus macaques (*Macaca mulatta*), aged 9 to 17 years and weighing 7.5 to 9.1 kg, were involved in the study. Three monkeys (B, J, and Y) were included for the awake (non-DBS) experiments, while two monkeys (N and T) were included for the DBS experiments.

Awake protocol

In the awake experiments, monkeys B, J, and Y were implanted with an MR-compatible headpost [14, 243]. Animals were conditioned to maintain a sphinx position in a primate chair, securing their heads without engaging in any specific task. Eye movements were tracked. [14, 243, 245, 234]. To protect against noise, protective ear caps were inserted into the monkeys' external auditory canals, and

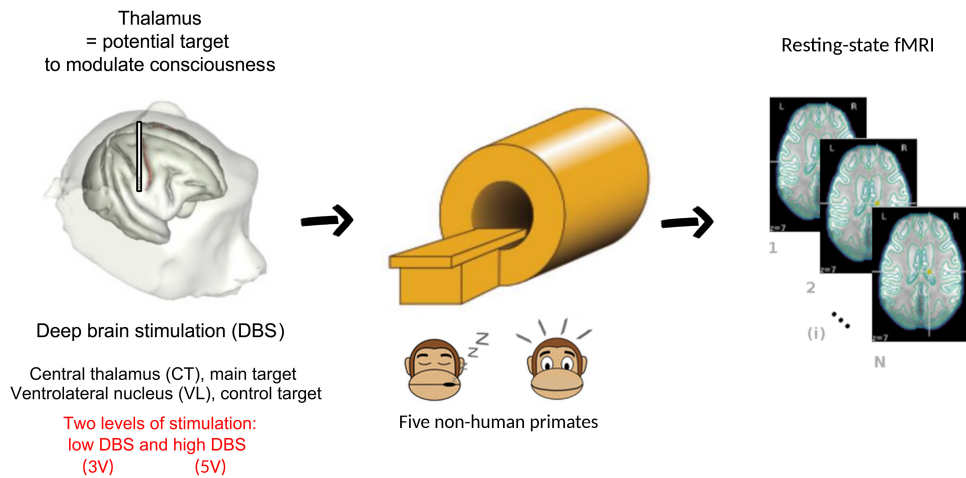


Figure 2.9: *Experimental design DBS dataset.*

a reward pump was introduced into the mouth for positive reinforcement. The experimental setup involved alternating between periods with no fixation task and a fixation task, where the monkeys were required to focus continuously on a red square ($0.35 \times 0.35^\circ$) within a $2 \times 2^\circ$ window on a black screen. Fluid reward was linked to fixation performance and regularly delivered during passive acquisitions. This strategy maintained the monkeys' motivation at a high level, but only runs without a task were included in the analysis to avoid potential visual activations. Experiments were concluded at the first sign of frustration or a decrease in performance, and the monkeys were returned to their housing cages with additional positive reinforcement in the form of fruits and vegetables.

Anesthesia protocol

Anesthesia initiation involved an intramuscular injection of ketamine (10 mg/kg) and dexmedetomidine (20 $\mu\text{g}/\text{kg}$). Maintenance was achieved through a target-controlled infusion (TCI) of propofol (monkey T: TCI, 4.6 to 4.8 $\mu\text{g}/\text{ml}$; monkey N: TCI, 4.0 to 4.2 $\mu\text{g}/\text{ml}$) (80). A waiting period of at least 80 minutes after the initial ketamine induction was observed to allow for the washout period before acquiring images under pure propofol sedation.

Monkeys underwent intubation and mechanical ventilation. Physiological parameters, including heart rate, noninvasive blood pressure, oxygen saturation, respiratory rate, end-tidal carbon dioxide, and cutaneous temperature, were continuously monitored. To prevent artifacts, a muscle-blocking agent (cisatracurium, 0.15 mg/kg, bolus i.v., followed by continuous intravenous infusion at a rate of 0.18 mg/kg per hour) was administered during all anesthesia fMRI sessions.

The level of sedation was determined through a combination of clinical scoring and continuous EEG, utilizing an MR-compatible EEG system that included a

custom-built 13-channel EEG cap (EasyCap), an amplifier, and the Vision Recorder software. Our target was a deep sedation level (general anesthesia), characterized by the complete absence of movements and responses to various stimuli, diffuse delta waves, waves of low amplitude, and anterior alpha waves [234].

Surgical procedures and electrode location

DBS electrode implantation Monkeys N and T underwent implantation with a clinical MRI-compatible DBS electrode [234]. The DBS lead featured four active contacts for electrical stimulation (1.5-mm contact length, 0.5-mm spacing, and 1.27-mm diameter). Stereotaxic surgery targeting the right Centro-Median (CM) thalamus was conducted, guided by rhesus macaque atlases [185, 207] and preoperative and intraoperative anatomical MRI (Magnetization Prepared-Rapid Gradient Echo (MPRAGE), T1-weighted, TR = 2200 ms, Inversion Time (TI) = 900 ms, 0.80-mm isotropic voxel size, sagittal orientation). The trajectory was virtually simulated using the neuronavigation module (BrainSight, Rogue, Canada) to anticipate implantation and visualize blood vessels or sensitive cerebral structures to avoid. In macaque monkeys, the CM thalamus is a diamond-shaped structure of around 9 × 6 × 5 millimeters (about half the size of a human). The extracranial part of the DBS lead was accommodated using a homemade three-dimensional (3D) printed MRI-compatible chamber. A waiting period of at least 20 days after implantation was observed before commencing the DBS-fMRI experiments [234].

Anatomical localization of the DBS lead Two methods were employed to ensure the anatomical localization of the DBS lead and the DBS contacts [234]. Firstly, a reconstruction method based on in vivo brain imaging was utilized. Secondly, a postmortem brain histology study was conducted in one of the implanted monkeys. The volume of activated tissue for all DBS conditions was simulated using the modeling module available in the Lead-DBS toolbox, and the activated thalamic nuclei (minimum 40% of the whole size) for each DBS condition were reported. All approaches led to a consistent localization of the four distinct stimulating DBS lead contacts (referenced as contacts 0, 1, 2, and 3). For monkey N, contact 0 was in the subthalamic nucleus, contact 1 in the zona incerta, contact 2 in the CM nucleus, and contact 3 in the VL. For monkey T, contact 0 was in the ventral posterior nucleus of the thalamus, contact 1 in the CM, contact 2 in the centrolateral nucleus of the thalamus, and contact 3 in the VL. During the DBS sessions, we stimulated the CT (contact 2 in monkey N and contact 1 in monkey T) or the VL (contact 3 in both monkeys) (cf. Figure 2.10).

Electrical stimulation protocol for the DBS experiments

For the DBS experiments, monkeys were anesthetized as detailed previously. Electrical stimulation was independently administered to each of the four leads (0, 1, 2, and 3) of the clinical DBS electrode. The DBS electrode was connected to an

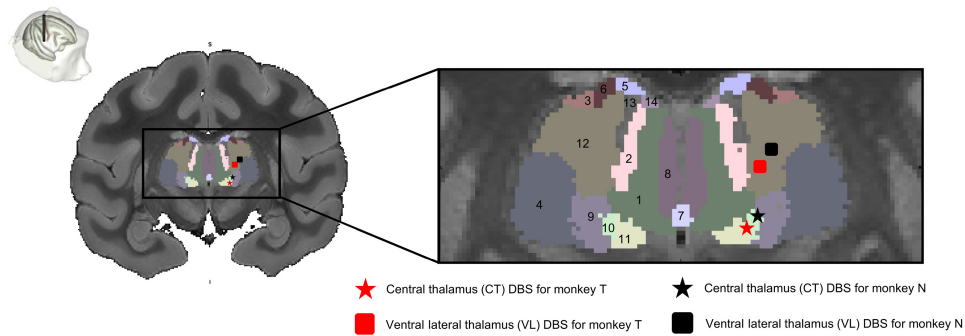


Figure 2.10: Anatomical localization of the DBS lead and active contacts. Coronal section of an anatomical MRI. Zoomed-in view of the thalamic nuclei as segmented with the CIVM atlas [30]. DBS was delivered to either the CT (star label) or the ventral lateral thalamus (VL) (square label). 1, mediodorsal nucleus, central part; 2, mediodorsal nucleus, lateral part; 3, ventral lateral nucleus, lateral part; 4, ventral posterolateral nucleus; 5, lateral dorsal nucleus, superficial part; 6, ventral anterior nucleus, lateral part; 7, intermediodorsal nucleus; 8, mediodorsal nucleus, medial part; 9, ventral posteromedial nucleus; 10, CM nucleus, lateral part; 11, CM nucleus, medial part; 12, ventral lateral nucleus, medial part; 13, centrolateral nucleus; 14, mediodorsal nucleus, dorsal part. Adapted from [234].

external stimulator, and all parameters were adjusted to fixed values of frequency ($f = 130.208$ Hz, $T = 7.68$ ms), waveform (monopolar signal), and pulse width (monkey N, $w = 320$ μ s; monkey T, $w = 140$ μ s). The absolute voltage amplitude was set to 3 V ("low" DBS) or 5 V ("high" DBS).

Behavioral assessment

A clinical arousal scale, adapted from [244], was employed to characterize monkey behavior in the awake state, under anesthesia, and during DBS (low CT-DBS, high CT-DBS, low VL-DBS, and high VL-DBS). This scale is based on the exploration of the surrounding world (0, absence; 1, small search for external clues; 2, total investigation of the environment, like head orientation to a sound), spontaneous movements (0, absence; 1, small torso and/or limb movement; 2, large torso and/or limb movement), shaking/prodding (0, nothing; 1, small body movement; 2, large body movement), toe pinch (0, nothing; 1, body movement or eye blinking or cardiac rate change; 2, body movement and eye blinking or eye opening and cardiac rate change), eye opening (0, nothing; 1, small blinks or eye movements; 2, full eye opening), and corneal reflex (0, absent; 1, present). Behavioral assessment was conducted outside the scanner in the awake state, under anesthesia, and under each DBS condition when the animal was not paralyzed [234]. Notably, only the stimulation of the CT could modulate arousal in the two anesthetized monkeys (Figure 2.11).

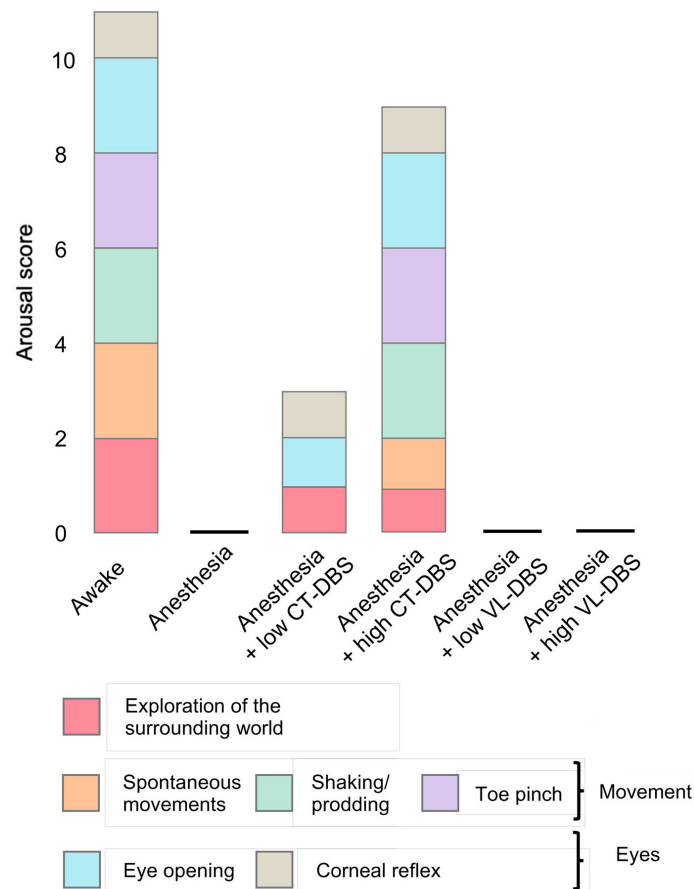


Figure 2.11: Clinical arousal scale in anesthetized monkeys, as a function of the electrode location and the level of stimulation (low-voltage versus high-voltage DBS). Adapted from [234].

fMRI data acquisition

The rhesus macaques underwent imaging on a 3-Tesla horizontal scanner (Siemens Prisma Fit, Erlanger, Germany) using a specialized eight-channel phased-array surface coil (KU Leuven, Belgium). The MRI sequence parameters for resting-state experiments were as follows: EPI, TR = 1250 ms, echo time (TE) = 14.20 ms, and a voxel size of 1.25 mm isotropic, with 500 brain volumes per run. No contrast agent was used.

The RS-fMRI experiments were conducted in either the awake condition or under anesthesia, with or without low or high DBS throughout the entire run targeting either the CT or VL thalamic nuclei. DBS initiation occurred a few seconds prior to the commencement of the fMRI sequence and ceased immediately after the conclusion of the MRI sequence. In total, 199 runs were acquired: 47 awake runs (monkey B, 18 runs; monkey J, 13 runs; monkey K, 16 runs), 20 DBS-3V-control runs (monkey T only), 20 DBS-5V-control runs (monkey T only), 38

anesthesia (monkey N, 16 runs; monkey T, 22 runs), 36 DBS-3V-CT (monkey N, 18 runs; monkey T, 18 runs), 38 DBS-5V-CT (monkey N, 17 runs; monkey T, 21 runs) [234] (for summary cf. Table 2.2).

monkey	awake	DBS-3V-control	DBS-5V-control	anesthesia	DBS-3V-CT	DBS-5V-CT	
J	13	-	-	-	-	-	13
B	18	-	-	-	-	-	18
K	16	-	-	-	-	-	16
T	-	20	20	22	18	21	101
N	-	-	-	16	18	17	51
	47	20	20	38	36	38	199

Table 2.2: Description of the acquisition conditions across NHP.

fMRI preprocessing

The preprocessing of images was carried out using Pypreclin [233]. Functional images were corrected for slice timing and B0 inhomogeneities, reoriented, realigned to T1 of the session, resampled (1.0 mm isotropic), masked, and smoothed (3.0-mm Gaussian kernel). Anatomical images were corrected for B1 inhomogeneities, normalized to the anatomical MNI macaque brain template, and masked. As previous dataset, the voxel time series underwent filtering with lowpass (0.05-Hz cutoff) and high-pass (0.0025-Hz cutoff) filters, along with a zero-phase fast-Fourier notch filter (0.03 Hz) to eliminate an artifactual pure frequency present in all the data.

2.2.3 . External resources: ROI template & reference anatomical connectivity

ROI template were sourced from the CoCoMac2.0 [11] database, derived from the F99 macaque standard cortical surface template co-registered to the MNI space [14]. This parcellation consists of 82 cortical ROIs (41 per hemisphere; Appendix 2), including interhemispheric connections, with mirror-symmetrical representation across hemispheres.

The reference anatomical connectivity matrix is a matrix displaying these 82 cortical regions of interest on the x-axis and y-axis. Each matrix cell signifies the strength of the anatomical connection between any pair of cortical areas. If information about the connectivity between two regions was unavailable in CoCoMac, the connection strength was set to 0 [224]. The reference CoCoMac connectivity matrix categorizes the strength of anatomical connections as weak, moderate, or strong, denoted as 1, 2, and 3, respectively.

2.2.4 . Functional connectivity (FC)

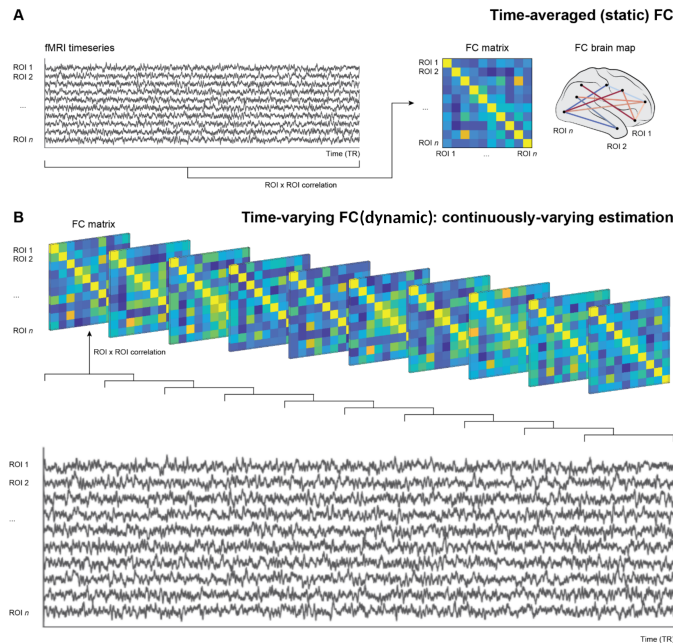


Figure 2.12: Static and dynamic FC. A) By correlating all pairs of ROIs over the entire timeseries we get time-averaged (static) FC. The resulting FC can be illustrated as (ROI \times ROI) FC matrix or as FC map in the brain. B) With sliding window approaches, we can estimate FC separately on portions (windows) of the timeseries. Adapted from [4].

Static FC

Let p be the fixed number of ROIs defined by the atlas (in our case $p = 82$ as defined in the CoCoMac atlas). The feature matrix C_r , where r represents the run, contains the averaged time series computed across all voxels within each ROI. From each feature matrix C_r , a static Functional Connectivity (FC) matrix reflecting the data empirical covariance structure is derived and will be analyzed afterwards (Fig. 2.12-A) [11].

Dynamic FC (dFC)

Empirical estimations of Dynamic Functional Connectivity (dFC) enable researchers to investigate how the degree of interregional coupling changes over time. These estimations serve as the foundation for empirical investigations into dFC. In their simplest manifestation, such as time-resolved correlations, they offer insights into the pathways through which static ("time-averaged") functional connectivity (FC) manifests. Time-resolved estimations also facilitate detailed assessments of the connection between FC and ongoing cognitive processes, as well as how aggregate measures (e.g., FC variability) may correlate with phenotypic characteristics in both health and disease [153].

Detailed methodology of the analysis of the dynamic RS-fMRI was previously

described in the literature [7, 14, 245]. dFCs were derived from the sliding window covariance matrix $C_{r,w}$, where r represents the run, and w the time window ranging from 1 to W [7, 14]. Covariance matrices were computed from segmented time series using a Hamming window with a width of 35 TR, sliding with 1 TR steps, resulting in $W = 464$ windows per session [14]. The variance of each time series was normalized, resulting in covariance matrices that corresponded to correlation matrices [7]. To address potential information insufficiency in short time segments, regularization was applied. A penalty on the L1 norm of the regularized matrix, promoting sparsity (with a regularization parameter λ set to 0.1), was employed following the graphical LASSO method by Friedman et al. (2008) [77].

This process yielded a 3D matrix $C_{r,w}$ of size $82 \times 82 \times 464$ for each run r , which was Fisher transformed ($Z_{r,w}$) before further analysis (Fig. 2.12-B).

Quality control (QC)

Anesthesia dataset Only a visual control quality was previously applied: no spiking, ghosting or artifacts... For reasons of reproducibility, we have retained all the data analyzed previously.

DBS dataset

Cleaning procedure

To ensure the data quality, we apply a manual cleaning procedure adapted from [226]. For each run, this procedure consists of a visual inspection of the averaged time series of each ROI, the z-scored FC matrix, and the dynamical FC matrix. We keep runs with no row signal artifacts and FC matrices coherent with the average FCs across the population (Figure 2.13). Finally, on the 199 available runs, 186 runs pass the checks. They form 6 conditions: 41 awake, 32 DBS-off (stim-off), 36 DBS-5V-CT (stim-on-5V), 35 DBS-3V-CT (stim-on-3V), 20 DBS-5V-VL (stim-cont-on-5V), and 20 DBS-3V-VL (stim-cont-on-3V).

Detection electrode artifact

We also perform a semi-automatic detection of the cortical voxels affected by the electrode artifact. We proceed with the anatomical MRI and the functional MRI volumes. Since the voxels affected by the artifact are of relatively low intensity, we apply intensity thresholding followed by a connected component analysis. We also use the position of the electrode as prior information to generate a rough bounding box composed of the superior-right part of the image. Finally, we intersect the found artifactual voxels with the CoCoMac atlas [11].

The electrode artifact impacts few voxels. The electrode detection with the proposed semi-automatic method (Fig. 2.14-A) highlights that 9 of the 41 cortical ROIs of the right hemisphere are corrupted by electrode-related artifacts. Concerning the ratio of voxels involved in each corrupted ROI across all runs, only one ROI has a ratio greater than 10% (the ROI corresponding to the primary somatosensory

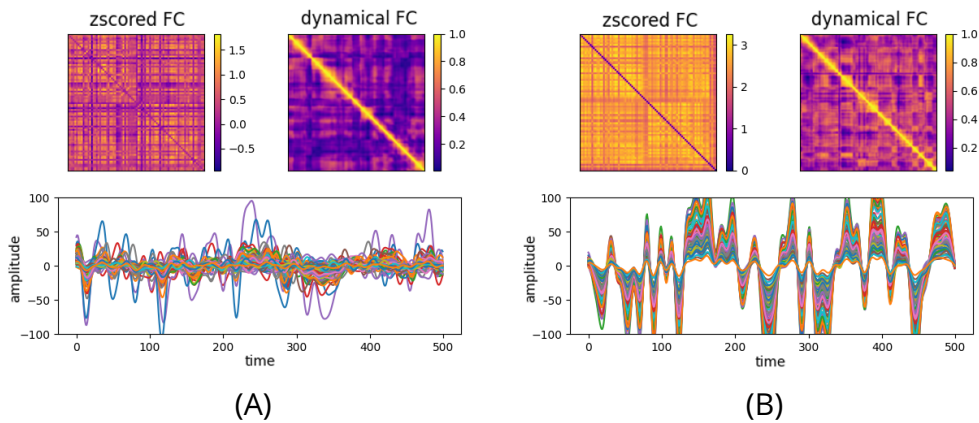


Figure 2.13: The control board of two runs is shown: run (A) shows no indices of abnormality so is kept after the post-processing step; run (B) shows abnormal time course, z-scored FC matrix, and dynamical FC matrix and is eliminated.

cortex S1), the others having a ratio under 5% (Fig. 2.14-B). As proposed in [234], we neglect this artifact in the rest of this study.

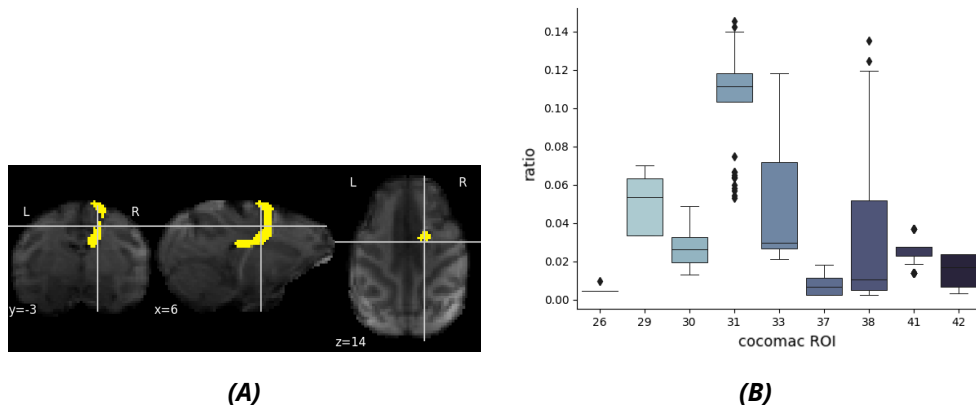


Figure 2.14: A) Visualization of the detected electrode on the T1w image of one run, and B) ratio of voxels across runs affected by the artifact for each ROI concerned by this artifact.

Coupling fMRI with EEG

Anesthesia dataset To assess the depth of anesthesia during ketamine, propofol, and sevoflurane sessions, we employed scalp EEG with a magnetic resonance imaging-compatible system and custom-built caps (EasyCap, 13 channels) [14]. The parameters were set as follows: sampling rate at 5,000 per channel, impedance below 20 M Ω , and a band-pass filter of 0.01 Hz < f < 500 Hz during data collection. EEG gel (One Step EEG gel; Germany) was applied to ensure low impedances.

EEG scalp recordings to verify the anesthesia level were obtained 10 minutes before commencing MRI acquisition, performed outside the scanner room.

An online analysis through visual assessment of EEG traces was performed [245]. The EEG traces were visually interpreted to establish anesthesia levels for clinical sedation during ketamine, propofol, and sevoflurane anesthesia. For ketamine, sedation levels were defined as follows [244, 245]: level 1, awake state, characterized by posterior α waves (eyes closed) and anterior β waves; level 2, light ketamine sedation, marked by the loss of α rhythm with a decrease in amplitude; level 3, moderate ketamine sedation, featuring persistent rhythmic θ activity (4 to 6 Hz) with increasing amplitude and fast β activity (14 to 20 Hz) of low amplitude; level 4, deep ketamine anesthesia, displaying intermittent polymorphic δ activity (0.5 to 2 Hz) of large amplitude, superimposed by a β activity (14 to 20 Hz) of low amplitude, an increase in γ power (30 to 100 Hz). For propofol, sedation levels were defined as follows [14, 244, 245]: level 1, awake state, characterized by posterior α waves (eyes closed) and anterior β waves; level 2, light propofol sedation, marked by an increase in the amplitude of α waves and anterior diffusion of α waves; level 3, moderate propofol sedation, featuring diffuse and wide α waves, and anterior θ waves; level 4, deep propofol anesthesia (general anesthesia), displaying diffuse delta waves, waves of low amplitude, and anterior α waves (10 Hz); level 5, very deep sedation (deeper than the level of general anesthesia), characterized by burst suppression. For sevoflurane, sedation levels were defined as follows [245]: level 1, awake state, characterized by posterior α waves (eyes closed) and anterior β waves; level 2, light sevoflurane sedation, marked by an increase in frontal and central β waves (no fMRI data collected at this sedation level); level 3, moderate sevoflurane sedation, featuring increased frontal delta, α , and β waves; level 4, deep sevoflurane anesthesia (general anesthesia), displaying diffuse delta waves and anterior α waves; level 5, very deep sedation (deeper than the level of general anesthesia), characterized by burst suppression (no fMRI data collected at this sedation level).

DBS dataset Scalp EEG data were obtained using an MR-compatible system and custom-built caps (EasyCap, 13 channels). This EEG acquisition not only facilitated the assessment of the sedation level but also served as an additional means to control the delivery of electrical stimulation. EEG recordings spanned the entire duration of the experiments, from the loss of animal consciousness to the initiation of vigilance recovery, both outside and inside the MRI environment.

Key parameters were set as follows: a sampling rate of 5000 per channel, a common reference electrode, impedance maintained at <20 megohms, and band-pass filtering in the range of $0.01 \text{ Hz} < f < 500 \text{ Hz}$ during data collection. To achieve low impedances, EEG gel (One Step EEG gel, Germany) was applied [14, 244, 245].

The correction of scanner artifacts was executed, employing a two-step process.

Initially, EEG epochs affected by the MR scanner were averaged, and this average template was then subtracted. For artifact detection, the gradient method and automatic detection of scanner episodes were utilized. All available channels with a signal were considered in the detection process (Fp1, Fp2, F3, F4, T3, T4, P3, P4, O1, Oz, and O2). The artifact type was specified as continuous, indicating detection on artifacts occurring consecutively without interruption, with a TR value of 1250 ms (offset set to 10 ms). The gradient trigger was set to 200 $\mu\text{V}/\text{ms}$. Baseline correction was computed over the entire artifact duration (0 to 200 ms). Sliding average calculation was activated with 21 intervals used for the correction template calculation. The correction was applied to the same channels mentioned above.

Subsequent to the MR artifact cleaning, the signal underwent filtering between 1 Hz (IIR Butterworth high-pass zero-phase two-pass forward and reverse noncausal filter, order 12—effective, after forward-backward) and 25 Hz (IIR Butterworth lowpass zero-phase noncausal filter, order 16—effective, after forward-backward) and was downsampled to a 250-Hz sampling rate. The data were then trimmed from 15 s after the start of the stimulation until 15 s before the end of the scanning. The remaining time series was divided into epochs of 0.8 s, with a random jitter ranging from 0.55 to 0.85 s. For the cleaning of artifacted epochs or channels, the Python package Autoreject [111] was employed with the number of channel interpolations set to 1, 2, 4, or 8. An average EEG reference projection was applied. All post-MR artifact removal preprocessing was conducted in Python using the MNE-Python [111] and Autoreject packages.

To examine distinctions among the five conditions in this study (anesthesia and 3 or 5 V of DBS in both CT and VL), we computed EEG markers that have been previously identified for distinguishing between individuals with and without disorders of consciousness [228, 66]. We focused on markers falling under the category of spectral measures. These markers encompass the normalized spectral power of delta (1 to 4 Hz), θ (4 to 8 Hz), and alpha (8 to 13 Hz) oscillatory bands; the Signal Entropy (SE), a measure of signal predictability; and the Median Power Frequency (MSF), representing the frequency dividing the power spectrum into two equal areas. The computation of these markers was performed using the NICE-tools package [66].

2.3 . Limits and associated challenges

The table 2.3 summarizes the different parameters of the data we use in the rest of the manuscript. Due to their major differences in acquisition, pooling these two datasets together is almost impossible. For this reason, the studies we have carried out have focused on either dataset, taken separately.

2.3.1 . The limits of acquisition

Name	Data type	Dataset	Signal	N	TR (s)	Preprocessing	QC
static anesthesia	static FC	anesthesia	MION	156 runs	2.4	NSM	No
dynamic anesthesia	dynamic FC	anesthesia	MION	72384 win	2.4	NSM	No
static DBS	static FC	DBS	BOLD	186 runs	1.25	pypreclin	Yes
dynamic DBS	dynamic FC	DBS	BOLD	86304 win	1.25	pypreclin	Yes

Table 2.3: Summary of the datasets used in this work.

Contrast agent

One of the fundamental differences is whether or not the MION contrast agent is used. In the first study, the signal enhancement provided by the contrast agent was favored to reduce the number of sessions. Indeed, the use of MION improves SNR. Nevertheless, this approach comes with several drawbacks [253]:

- the signal reflects alterations in blood volume [137] instead of the usual BOLD response, making its translation to conventional human studies more complex,
- the HRF is slightly slower [137, 189]
- it can accumulate in the brain leading to potential long-term health risks [173]

In the second dataset (DBS), the choice was made not to use a contrast agent, in order to approximate as closely as possible the experimental conditions in humans. A larger dataset was acquired, yet this approach is costly and impractical, and it poses unavoidable risks to subjects if general anesthetics are employed [253].

Small number of subjects

Both datasets were collected from five rhesus macaques, which is a small sample. Despite the repeated number of sessions and the even greater number of data sets obtained using the sliding windows method, the small number of individuals remains a major limitation. This is a common limitation of NHP studies. We shall see that these datasets remain highly sensitive to overfitting despite the precautions taken for the analysis methods we propose. This raises concerns about the generalizability of the results, particularly regarding the risk of overfitting.

Generative methods modeling brain dynamics can simulate new data to address the data scarcity issue. Computational models provide a solution by generating realistic data for augmentation during training. The approach we employed for investigating dFC directly focuses on the observed BOLD signal without explicitly modeling the underlying neural activity. It is a data-driven method. Another approach exists that aims to model the underlying neural fluctuations and interactions giving rise to BOLD dFC. This perspective suggests that observed BOLD

time series are generated by underlying nonlinear brain dynamics, which are then corrupted by measurement noise. According to this view, activity in large-scale neural systems is inherently dynamic and exhibits complex phenomena such as partial synchronization, multistable attractor landscapes, and edge-of-chaos behavior indicative of criticality [44, 97, 55, 205, 254] (for review, see [153]). These dynamics produce physiological time series with a highly nonlinear structure and can be formally modeled by biophysically derived differential equations. By integrating these equations with models of the observation process (e.g., neurovascular coupling), it is feasible to simulate how these underlying dynamics would manifest in the BOLD signal (i.e., after adding measurement noise). Exploratory computational efforts involve refining the model structure and adjusting parameters to generate synthetic BOLD data that mimics the dependency structure and dynamics seen in empirical observations [54, 115, 61, 249]. Modern physical models, like the Hopf model, can simulate fMRI activity across the entire brain by reproducing dynamics resulting from mutual interactions among brain regions when coupled by structural connectivity [128, 56]. These models effectively capture aspects of brain dynamics observed in electrophysiology [74, 75], MEG [52], and fMRI [128, 56]. However, model-based approaches must rely on strong assumptions about the processes underlying observed BOLD data [55].

Sparsity of conditions

Moreover, the data are heterogeneous: not all conditions are represented for all individuals (cf. tables 2.1 & 2.2). For instance, two subjects did not have an fMRI scan in the awake state, and one had scans only in the awake state, not under anesthesia, for the Anesthesia dataset. For the DBS dataset, due to the implant constraint, monkeys scanned awake are not those scanned anesthetized with or without DBS. Only one monkey has both control DBS and effective DBS conditions.

To mitigate the risk of overfitting, we take great care when selecting data: we use a train set and an independent test set with leave-one-subject-out for the Anesthesia dataset. This forces us to work with one subject only. For the DBS dataset, the heterogeneity of the conditions does not allow us to make this choice. Each awake subject is then used once as a test while the $k - 1$ remaining awake subjects form the training set.

2.3.2 . The limits of preprocessing

Confounding effects

What does the change in fMRI reflect? Rs-fMRI likely mirrors a spectrum of conscious and unconscious cognitive processes, alongside inherent noncognitive processes [153, 129]. However, distinguishing between the neural and physiological factors associated with arousal and their resultant effects, such as alterations

in head motion, heart rate, and respiration, is not straightforward. Hence, it is important to exercise caution in considering them as artifacts.

Neuromodulators Neuromodulators can also affect neurovascular coupling [153, 27, 134]. Consequently, it is crucial to ensure that the effects observed in fMRI studies involving pharmacological manipulation are genuinely associated with alterations in neural activity, rather than solely resulting from hemodynamic changes [153].

Temperature Body temperature, particularly cortical temperature, influences neurovascular coupling and consequently the BOLD signal in fMRI [23]. Induction of anesthesia typically does not result in significant drops in body temperature in older children and adults. However babies and young children experience significant decreases in body temperature during this brief period. The size of the subjects influences this; similarly, macaques also experience a drop in temperature during anesthesia. To prevent heat loss in these subjects, it is essential to establish a thermoneutral environment by increasing the temperature of the operating room, positioning a heating lamp above the operating table, activating a heating mattress (40°C) before bringing the subject into the operating room and before inducing anesthesia [39]. Despite these precautions, temperature differences between the awake and anesthetized states may persist.

Head motion Head motion is considered one of the most influential confounding factors affecting the estimation of BOLD functional connectivity [196, 246]. Several studies have shown that even minor head movements can lead to biased estimates of static FC. Further research in this area is required to develop a more comprehensive understanding of how various aspects of BOLD dFC are affected by head motion and to devise effective preprocessing techniques [153].

Global signal

In terms of preprocessing, certain choices had been made prior to our study, which we haven't gone back on. However, these may have an influence on the results. In particular, the global signal is regressed, which remains controversial in the literature [145, 172, 248].

Electrode artefact

A semi-automatic artifact detection method was proposed, as described in section 2.2.4. We would certainly need to refine this electrode segmentation. Furthermore, although we found that relatively few voxels were affected and decided to continue with the original data, it would seem advisable not to average over the voxels concerned when using this data in the future.

Non-standardized NHP preprocessing

There isn't a standardized approach for analyzing monkey fMRI data, particularly in terms of preprocessing. The preprocessing of monkey fMRI data faces various technical and experimental challenges specific to primate research, such as artifacts caused by body movements or intracranial leads [233]. Preprocessing methods have differed between the Anesthesia and DBS datasets due to issues with the previous standard preprocessing method NSM, which failed to register BOLD images from a monkey implanted with a DBS lead. The pypreclin pipeline [233] successfully addressed this challenging task.

In particular, there are several differences between the two preprocessing methods (see [233] for a complete list). Pypreclin can reorient images regardless of the monkey's positioning during fMRI acquisition. Regarding normalization and B1 inhomogeneities correction, NSM preprocessing doesn't utilize the monkey's own anatomical image but rather directly registers a selected template image (e.g., macaque MNI template) with the functional images. Automation and QC reporting also differ: NSM preprocessing is fully automated without the option for manual initialization and lacks a quality check report, whereas pypreclin offers these features [233].

To overcome both the lack of data and the heterogeneity of pre-processing, initiatives have been launched to share primate MRI data. One example is the Primate neuroimaging Data-Exchange (Prime-DE) open access platform, which has been recently introduced to develop open resources for non-human primate imaging [162].

2.3.3 . The limits of connectivity computation

Computing dynamic functional connectivity

Metric Functional brain connectivity, as detected through distant correlations in fMRI signals, holds promise as a source of biomarkers for brain pathologies. However, there are several methods for calculating this connectivity, and the resulting outcomes are not always consistent. The study by Dadi et al (2019) [49] compares different types of functional connectivity between regions of interest: correlation, partial correlation, and tangent space embedding. These connectivity coefficients can distinguish between children and adults. Generally, tangent space embedding outperforms standard correlations [49], although it was not the previous choice made by the team, and this decision has not been revisited.

Sliding windows limitations Many variations in the approach are used to estimate the pairwise correlations within a sliding window. For the different options, the review by [153] lists most of the options available, including the type of window used (square, tapered, or exponentially decaying), the flexibility of the window (fixed or adaptive), as well as the length of the window.

If a small window size is selected, the correlation coefficient will be calculated based on a limited number of data points, leading to increased sampling variability. Consequently, shorter window lengths may produce signals indicating dynamic changes in correlation across time, even if the FC remains static. This issue becomes less pronounced as the window length increases, but longer windows sacrifice sensitivity to transient correlation changes. Moreover, overlapping windows can induce autocorrelation in the estimated dFC values, potentially resulting in artificially smoothed changes in FC. However, recent research suggests that the optimal window length to mitigate these concerns may be shorter than the recommended minimum of approximately 60 seconds. Hence, the choice of window size can be viewed as a tunable filter that can be optimized based on the specific research question of interest [153].

Another problem associated with sliding windows is overlapping. By construction, dFCs are highly correlated with each other, which can lead to a bias in the analysis. Alternative methods for avoiding the overlapping induced by the sliding windows method are proposed as the phase synchronization [82] to which we will return in Chapter 4.1.2.

Computing structural connectivity

The structural connectivity matrix used is not derived from a tractography directly performed on our data, but uses a previously defined parcellation, an average parcellation. Moreover, it has weaknesses in the discrete way it quantifies the strength of anatomical connections. It may not be perfectly reliable. An existing continuous matrix could be used for more detailed work on the relationship between brain structure and function. *Luppi et al.* [151], for example, use the recent macaque connectome of [225], which combines diffusion MRI tractography with axonal tract-tracing studies.

In addition, the CoCoMac atlas comprises 82 uniquely cortical regions. They are also functionally defined, which can be inconvenient for interpretation from an anatomical point of view [232]. Including more regions and in particular deep structures such as core areas of the ARAS (for instance, reticular formation, basal forebrain, thalamus, striatum), seems important for the study of vigilance. Atlas, such as the CIVM atlas [30], that was revised [234, 232] could be interesting.

Motivations and aims of the thesis

We have described the brain as a mystical crystal ball, whose interior, from which consciousness emerges, is obscure, wisps of smoke that escape us, slip through our fingers. We have explored this consciousness, which we gain, lose and recover. Once unreachable because of the jewel box that guarded it so well, we have seen that neuroimaging makes the box less black, almost transparent. We can now touch the object of our desire with our fingertips and see the brain in action. Neuroimaging holds up a magic mirror, where we can observe the brain at work, both at rest and in action.

Because it's not always easy to read a crystal ball, we propose a number of reading keys in the rest of this thesis. We seek to model states of consciousness and their modulation. Simulation tools and the rise of machine learning (including deep learning) are enabling us to vitrify consciousness, to model it in order to understand it better. We aim to identify signatures of anesthesia-induced loss of consciousness, as well as those of restoration of consciousness induced by DBS, from fMRI signals. To do this, we are starting from the experimental data at our disposal and working towards computational models of consciousness. In this way, we hope to make sense of the signals sent back to us by this mysterious object, and that this modeling will be the key to interpretation. The contributions of this thesis can be seen as tools of interpretability.



Illustration adapted from Magritte's La trahison des images. From PhD Days 2018 : de l'autre côté du miroir, au delà de notre réalité ?

Understanding a story from the observation of a crystal ball is no easy task. Before you start believing that I think I'm a witch, a crazy doctoral student turned

Madame Irma, this idea is not confined to sorcery. Images don't tell us everything; they tell only partially, revealing sparingly. In 1929, Magritte painted a famous picture of a pipe, accompanied by the caption "Ceci n'est pas une pipe" ("This is not a pipe"). The title of this work, "La Trahison des images" (The betrayal of images), leaves no room for doubt as to the painter's intentions: an image, however realistic, is only a representation and will never be identical to the object itself.

However, unlike a pipe, which is easier to understand how it works by holding it in our hands than by seeing it painted, when we study the brain, and more specifically the brain in action, using neuroimaging techniques gives us many clues as to how the brain works. Although the signals we acquire only partially represent what the brain is and what it does, they remain informative but high-dimensional, noisy and tangled, making them complex to interpret.



Illustration of Plato's Allegory of the Cave. Author unknown.

Magritte, who made no secret of his interest in mystery, said: "Each thing we see conceals something else. There is an interest in what is hidden and what the visible does not show us". It's also an idea described by Plato, years before Magritte, in the cave allegory: men are chained and imprisoned in a cave. Blocked in front of a wall, they can't see the real sources of the shadows dancing on it. Since these shadows are their only knowledge of the world, they believe they are seeing reality when in fact, they are only seeing a distorted and partial projection of it. To understand the real image of the world, we need access to the latent variables that are inaccessible to the observer's eyes. Seeking out this hidden information is essential to understanding how much of the data is informative about brain function, and how much is not.

Can we identify and interpret latent variables in fMRI signals? To what extent does identifying these variables inform us about the modulation of states of

consciousness? These questions lead to new ones: *what characterizes a state of consciousness? Is it a substrate that defines it, a set of regions or networks? Is it a sequence of events, a spatio-temporal dynamic? Is it both?*

We made several hypotheses:

- (i) Latent variables models can inform in an unsupervised way about cortical networks specifically related to conscious information processing.
- (ii) Latent variables encode information about dynamics and transitions between states.
- (iii) The traditional categorical approach does not account for the continuum of the dynamics of states of consciousness.
- (iv) The reduction of the observation space to a very low-dimension is sufficient to separate the levels of consciousness.

This work has led to several contributions, which we present in the following three chapters.

In Chapter 3, on the basis of hypothesis (i), we propose the identification of an interpretable spatial signature of consciousness in the awake state or under anesthesia, using a method that highlights particular networks involved in conscious access. In a translational approach, in order to study the restoration of consciousness, we have replicated this analysis in NHPs awake or awakened by CT-DBS. Our model puts forward, in an unsupervised manner, that both the anterior and posterior cortex contribute to consciousness, which is still under debate, today, in the community. It also highlights key regions of the global neuronal workspace, a major theory of conscious access.

Because we believe that integrating the temporal aspect into these analyses is crucial (hypothesis (ii)), in Chapter 4, we next propose to challenge the dFC methods traditionally employed. We apply a contrastive deep learning model to predict brain patterns characteristic of states of consciousness. A preliminary experiment shows that network predictions based on dFC enable the analysis of different transient brain states, with potential clinical implications for anesthesia monitoring and diagnosing disorders of consciousness.

Finally, in Chapter 5, in order to test hypothesis (iii) and gain a better understanding of the characteristics of the dynamics of states of consciousness, we depart from the traditional framework of subgroup classification of dynamic brain states and propose a dimension-reduction method to capture the whole continuum. We choose to reduce them to a small number of variables that can be interpreted and explained. A virtual connection ablation experiment is used to test hypothesis (iv).

Part II

Seeing patterns, deciphering messages: the spatial signatures of consciousness



3 - Functional connectivity at rest: studying conscious networks

Contents

3.1	Introduction	80
3.2	Related works to identify stationary networks	81
3.2.1	Multivariate decomposition	82
3.2.2	Graph	86
3.3	Material and Method	88
3.3.1	Datasets and atlas choice	88
3.3.2	The Modular Hierarchical Analysis (MHA)	89
3.3.3	Decoding Brain Network Activities (BNAs)	91
3.4	Results on Anesthesia Dataset	92
3.4.1	Consciousness connectivity can be decomposed into few consistent brain networks	92
3.4.2	Brain network 1 intersects the Global Neuronal Workspace (GNW)	93
3.4.3	Which network best predicts the depth of anesthesia from BNAs?	94
3.4.4	Influence of time window size on predictions: sensitivity study	96
3.4.5	Conclusion	96
3.5	Results on DBS dataset	97
3.5.1	Consciousness connectivity can be decomposed into few consistent brain networks	97
3.5.2	Identified networks	97
3.5.3	Reconciling the front and back of the brain in the processing of conscious information	98
3.5.4	Networks capture differences induced by the effective DBS condition	98
3.5.5	Conclusion	98
3.6	Discussion	99

3.1 . Introduction

"Tell me who your friends are, and I'll tell you who you are". So, to define consciousness at last, do we need to find its friends? Or rather, can we define consciousness by identifying the network that characterizes it? Isn't that what this proverb is all about? We're just a knot in the middle of the fabric formed by our acquaintances, our family, our colleagues, our friends, hooked up to them by a whole host of connections. And what if consciousness, too, could be defined in this way, by the brain tissue that makes it up, by the links that are created between different regions of interest, sometimes far apart? Theories clash and, as we saw earlier, we can't decide which is the network of consciousness.

RS-fMRI is a neuroimaging technique with undeniable advantages for non-invasively studying fluctuations in brain activity. Although we are working here in a pre-clinical setting, we feel it is important to emphasize the transposability for clinical studies. Indeed, the resting state paradigm holds significant appeal as it eliminates the necessity for a sophisticated experimental setup to administer external stimuli and obviates the requirement for active patient participation. Consequently, resting state protocols prove to be a suitable approach for studying clinical populations where direct communication at the bedside is challenging, such as individuals with disorders of consciousness [96]. Numerous studies utilizing RS-fMRI have uncovered consistent patterns of long-distance interactions, termed Resting State Network (RSN)s, across participants during rest. While the specific origin and function of RSNs continue to be debated, their spatial patterns disclose functionally relevant brain subsystems [67].

These cerebral networks have been compared in humans in normal and pathological conditions but also in different species of NHP [80]. Homologies of several networks have been found as the executive or the Default Mode Network (DMN) [80]. The majority of studies have focused on network differences during anesthesia-induced loss of consciousness [107, 105, 80, 101] or in patients with disorders of consciousness [96]. A few studies have also looked at the effect of Subthalamic Nucleus DBS (STN-DBS) neuromodulation in terms of networks in Parkinson's disease [102, 67] or the intrinsic reconfiguration of brain activity networks after active tDCS. With regard to the effects of CT-DBS, some studies have suggested that cortical and subcortical networks are affected, but their analysis is based on seed-based correlations only [140, 234]. To the best of our knowledge, no study to date has focused on how CT-DBS affects brain activity at the level of RSNs in the context of restoring consciousness. This could help identify possible mechanisms of action and permit novel therapeutic strategies targeted at specific network patterns [67]. To help bridge this gap, we ask if we can identify a dimension of consciousness that is reflected in different RSN patterns. Is this reproducible on a dataset with neuromodulation? Can we identify potential targets for DBS and sites where DBS is not effective? We make the following assumptions:

1. Networks activities can predict the state of consciousness
2. Neuromodulation reconfigures networks between the unconscious and awake states.
3. Some networks inform us about the cortical impact of neuromodulation

To answer these questions, in this part II, we'll first look at the literature on methods for identifying stationary brain networks. Then, we will present the method and results of our two conference papers "Interpretable Signature of Consciousness in Resting-State Functional Network Brain Activity" (MICCAI 2022) [89] an extension of which is available as a preprint ("Revisiting the standard for modeling functional brain network activity: application to consciousness" [90]) and "Exploration of the Neural Correlates of Consciousness Using Linear Latent Model" (ISBI 2023) [84]. In these studies, we utilize the previously collected RS-fMRI data in NHP to reveal new insights on interpretable, spatial signatures of consciousness. We propose a computational framework to find resting-state functional brain networks from FC, using the Modular Hierarchical Analysis (MHA) method, a linear latent variable model. In particular, we focus on how these networks break down according to the conditions of acquisition and how they can be interpreted, specifically in the context of the theoretical framework of consciousness, GNW. The method we propose here approaches networks in a static way, i.e. we don't consider the dynamic variations that occur during an acquisition. The aim is to get a global view of the underlying networks in our data and to study whether this organization informs us about the modulation of consciousness. The interest of comparing the results of the two retrospective datasets considered is to reveal a global signature of the modulation of consciousness, from loss to recovery, and in particular, to compare the networks that stand out and are involved in separating the levels of consciousness.

3.2 . Related works to identify stationary networks

Analyzing RS-fMRI data can give insights into the function of specific brain regions or the functional connectivity between different regions. Analytic approaches in this domain generally fall into two categories: functional segregation and functional integration. Functional segregation emphasizes the local function of specific brain regions and is commonly utilized for brain mapping purposes. On the other hand, functional integration focuses on the functional relationships or connectivity between different brain areas, treating the brain as an integrated network. While functional segregation techniques analyze RS-fMRI activity, functional integration methods primarily explore RS-fMRI connectivity [154].

However, the trend in neuroscience has shifted toward considering the brain as an integrated network rather than isolated regions. Consequently, enthusiasm

for standalone functional segregation methods has gradually diminished in favor of functional integration methods [154].

Functional integration, specifically, delves into the functional connectivity between distinct brain regions, measuring correlations in activity between spatially remote areas. It assesses the degree of synchrony in the BOLD time series across different brain regions. It's important to note that functional connectivity does not always indicate a direct causal influence between regions; it could result from direct anatomical connections, indirect paths via mediating regions, or may lack a known anatomical correlation. The interpretation of functional connectivity analyses should be cautiously approached due to the potential influence of common sources of input signals [195].

Common computational methods employed for assessing functional integration features include ROI-based functional connectivity analysis, Independent Component Analysis (ICA), Dictionary Learning (DL), and graph analysis [154]. The first method, a seed correlation analysis, aims to identify the correlates of specific regions at rest. However, as our focus is on gaining a global perspective of the detectable networks involved in consciousness, the subsequent sections will concentrate on the latter three methods.

3.2.1 . Multivariate decomposition

There are a number of ways to decompose a matrix into separate components, which can be used to identify coherently active networks from fMRI data. In the matrix algebra language, these are known as matrix factorization methods. Each of these methods assumes that the data are composed of some underlying components mixed together to form the observed data. The main differences between methods center on how the underlying components are related to one another and how they are estimated from the data [195].

Independent component analysis (ICA)

Method ICA was developed to address the challenge of detecting unknown signals in a dataset, commonly referred to as the blind source separation problem. This problem is often illustrated using the analogy of a cocktail party [108]. Consider a scenario with microphones positioned throughout a room where numerous people are engaged in conversation. The blind source separation problem involves isolating the speech stream of each individual solely based on the recordings from these microphones. The concept assumes that the recording from each microphone reflects a mixture of all speakers (weighted by factors like their distance from the microphone and head direction). The goal is to separate the sources from the recordings, assuming that both the speakers and microphones remain stationary during the recording, making the mixing process consistent throughout.

Formally, the ICA model is defined as $x = As$, where x is the signal under decomposition, s represents a set of unknown sources or components, and A is

the unknown mixing matrix combining the components to produce the observed signal. Since both A and s are unknown, certain assumptions must be made to find a unique solution. Generally, assumptions about the relationship between different components in s are necessary. If we assume these components are orthogonal and Gaussian, Principal Component Analysis (PCA) can be used to solve the problem. However, in cases where the signals from different sources are neither orthogonal nor Gaussian, PCA may fail to identify them. Formally, ICA constitutes a linear latent variable model; however, unlike PCA, it does not assume that the latent variables follow a Gaussian distribution [169]. ICA relies on the assumption that the components in s are statistically independent. In situations where the signal components are generated by independent processes (such as independent speakers at a cocktail party or independent neural processes in fMRI) ICA may outperform PCA in correctly identifying the component sources due to their likely non-Gaussian nature [195].

Independence is a concept related to, but distinct from, orthogonality (or uncorrelatedness). It is possible for two variables to be statistically dependent even if they are orthogonal, a situation that arises when the data deviate from a Gaussian distribution. In the context of ICA, the independent components are estimated by identifying non-Gaussian signals in the data. Given that ICA seeks non-Gaussian signals, many ICA algorithms initially whiten the data using PCA to eliminate any Gaussian signals present [195].

When applying ICA to fMRI timecourses data, a crucial decision is whether the algorithm should search for components that are spatially independent or temporally independent. Most methods assume spatially independent components, yielding a set of spatial components along with a mixing matrix that indicates the contribution of each spatial pattern to the observed signal at each time point. The assumption of spatial independence is justified by the notion that the brain harbors numerous potentially independent networks, each with similar timecourses during task performance. This approach facilitates the detection of spatially distinct effects, such as differentiating task activation from task-correlated head motion, even if their timecourses are correlated [195].

In summary, ICA employs multivariate decomposition to segregate the BOLD signal into distinct independent functional networks represented by spatial maps that exhibit temporal correlation. Each functional network, or component, encapsulates an independent set of ROIs with synchronized BOLD activity [154].

Results Various resting-state networks commonly arise from ICA analyses in RS-fMRI studies. These networks include, but are not limited to, the default mode network, auditory network, salience network, executive control network, medial visual network, lateral visual network, sensorimotor cortex, dorsal visual stream (frontoparietal attention network), basal ganglia network, limbic network, and precuneus network (cf. section 2.1.5). These networks exhibit resting-state connectivities,

with some showing up- or down-regulation during specific cognitive tasks [154].

The number of networks identified by ICA and group-ICA ranges from 8 to 30 networks in macaque [199, 105, 253]. The number of networks depends on the number of components chosen, the atlas and, above all, the initial image resolution.

Challenges ICA can be conducted without incorporating any a priori assumptions, except for the need to specify the number of independent components to identify. The process of dimensionality reduction and model selection are somewhat arbitrary, as one must determine the number of components to estimate, and there is no consensus on the ideal dimensionality for understanding the neurophysiology of multiple distributed systems [253]. Depending on the specified number of independent components, a single network may be fragmented into subnetworks [154].

Furthermore, the perceived synchrony within functional networks, as identified by ICA, might stem from non-neural factors such as breathing, pulsation, motion, scanner artifacts, and noise. This inherent characteristic of ICA complicates result interpretation [154]. In most studies, components derived from ICA are manually examined and labeled based on anatomical and functional criteria, with some components being discarded due to displaying noisy, nonspecific, low correlation activation patterns, or corresponding to large veins [253].

Several considerations need to be taken into account when utilizing ICA. ICA presents brain networks individually and does not illustrate connections between modules or communications among different brain networks. The solutions produced by ICA are not unique, requiring additional constraints to identify an optimal solution. Due to its sensitivity to non-Gaussian structure, ICA can be influenced by outliers in the data, which can be advantageous for identifying potential artifacts but should be carefully addressed to ensure that results are not driven by outliers [195]. The presence of negative values in the loading matrix adds complexity to the interpretation of such matrices [169]. ICA has faced criticism for assuming independence among underlying sources, as adjacent neurons are typically correlated in their responses, potentially impacting the assumption of independence [50].

Dictionary Learning

Recent studies have demonstrated that favorable outcomes could be obtained when imposing sparsity, rather than independence, in spatial decomposition through the use of a dictionary learning formulation [161]. Techniques based on dictionary learning outperform ICA in terms of stability. In neuroimaging, dictionary learning aims to extract a small set of representative temporal elements, accompanied by their sparse spatial loadings, resulting in well-defined extracted maps [161].

Method Dictionary learning serves as a sparsity-based decomposition technique designed for extracting spatial maps. The maps it extracts are inherently sparse

and typically exhibit greater clarity compared to those obtained through ICA [161]. This method generates probability maps by concatenating individual records from groups of subjects [80].

Results In the study by *Garin et al. (2021)* [80], sparse components-based DL was applied to mouse lemurs and humans, aiming to emphasize large networks and facilitate their comparison. In lemurs, six sparse components were utilized, while humans employed eight, determined by the number of modules identified through graph theory analysis. The resulting maps, revealing voxels belonging to distinct networks, were generated without relying on any atlas [80]. However, to identify brain regions associated with large-scale networks, 3D atlases were employed, assigning names to regions within each network based on arbitrary criteria [80]. We briefly present the networks obtained by this group as it represents one of the most recent pieces of literature on the topic, offering high-resolution insights into NHP (mouse lemur) brain networks using high field 11.7T MRI.

In mouse lemurs, the following prominent networks were identified (Figure 3.1) [80]:

1) The **somato-motor network** encompassed the frontal anterior lateral area, all parietal regions, anterior cingulate cortex/supplementary motor area/frontal superior region, and medium cingulate/paracentral lobule.

2) The **occipito-parietal network** involved all occipital regions, along with areas in the parietal posterior, temporal middle/inferior, and cingulum posterior/precuneus cortices. The presence of occipital and parietal regions suggests similarity to the visual network observed in humans.

3) The **fronto-parietal network** included the frontal anterior lateral cortex and the dorsal part of the frontal superior medial (dlFC regions), parietal posterior cortex, medial and posterior cingulate cortices, as well as retrosplenial regions. It also incorporated the temporal middle/inferior cortex, hippocampus, and occipital regions.

4) The **fronto-temporal network** comprised the frontal anterior medial and lateral regions, precentral cortex, all temporal regions, parietal posterior cortex, anterior and medial cingulum cortices, and the insular cortex.

5) The **sensory-limbic network** involved limbic structures (basal forebrain, septal nuclei, midbrain, hippocampus, hypothalamus) and numerous regions associated with vision (occipital cortex, superior colliculi) or audition (inferior colliculi). Additionally, it encompassed the cingulum posterior/precuneus and subcortical regions (thalamus, caudate nucleus, and the globus pallidus).

6) The **evaluative-limbic network** embedded limbic structures (basal forebrain, septal nuclei, amygdala, hippocampus), the insula, and subcortical structures (striatum including the caudate nucleus, putamen, and the accumbens nucleus of the ventral striatum, as well as the globus pallidus).

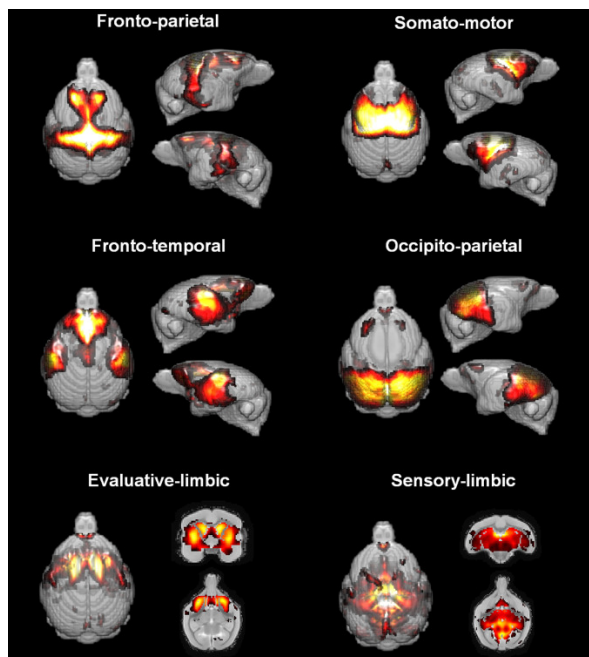


Figure 3.1: *Cerebral networks in mouse lemurs. From [80]*

Challenges Maps obtained through Dictionary Learning are often more straightforward to interpret due to their increased contrast compared to ICA maps, with more clearly defined blobs. However, similar to ICA, the dictionary method employed to characterize networks necessitates a priori selection of the number of components the observer intends to extract. Given the gradual nature of interactions among different brain regions, there are no distinct boundaries between large-scale networks, and various methods have been proposed to partition a network into communities [80]. Therefore, the selection of the number of networks in a resting-state study can be perceived as "arbitrary." In the study presented above, the number of networks identified in lemurs was determined using the "stability of a network partition" method derived from graph theory. Increasing the number of networks could have resulted in the subdivision of some networks into subnetworks [80].

3.2.2 . Graph

An alternative method for modeling connectivity in fMRI data originates from an unexpected source: the examination of social networks. Sociologists have historically explored the "six degrees of separation" concept, suggesting that almost everyone can establish a path of friends to anyone else in the world within six or fewer steps. In the 1990s, a cohort of physicists delved into the analysis of complex networks, including the World Wide Web and the brain, leading to the formulation of new models for comprehending the structure of diverse complex networks [195].

Method Graph theory offers a theoretical framework for analyzing the topology of brain networks, examining both local and global organization. In this framework, functional brain networks are defined as a graph (G), characterized by nodes (V) and functional connections (E), represented as $G = f(V,E)$. Nodes (V) typically represent voxels or ROIs [154]. To optimize computational efficiency, signal extraction often focuses on a limited number of regions of interest, as analyses involving more than a few thousand nodes become computationally intensive [195].

Functional connectivity (E) is quantified using measures that assess the strength of the relationship between signals at each node. The Pearson correlation coefficient (r) is a common measure of adjacency. The resulting adjacency matrix is usually thresholded at a relatively liberal value (e.g., $r > 0.1$) to exclude noise-related edges. Post-thresholding, the adjacency matrix becomes a binary $V \times V$ matrix indicating the presence or absence of links between nodes [195].

Following network estimation from adjacency measures, various aspects of the network can be characterized using key graph analysis parameters [154]:

- "Clustering coefficient" signifies the degree of local neighborhood clustering, reflecting local connectedness.
- "Characteristic path length" represents the average number of connections between all pairs of nodes, indicating global connectivity and network efficiency.
- "Node degree" indicates the number of connections for each node, identifying highly connected nodes.
- "Centrality" represents the number of short-range connections for each node, with nodes of higher centrality contributing more to overall network efficiency.
- "Modularity" gauges the extent to which groups of nodes connect with members of their own group, revealing subnetworks within the overall network.

Results In functional brain connectivity networks, the organization of nodes has been shown to be crucial, for example, in distinguishing different states of consciousness [33, 2]. Graph analyses can help answer questions about critical elements of brain network organization, for brain functions such as consciousness [2].

Graph theoretical methods were employed to investigate the topology of brain networks using RS-fMRI data collected from 17 patients with severe consciousness impairment and 20 healthy individuals [2]. The study revealed that many global network characteristics remained unchanged in comatose patients. Specifically, there were no significant abnormalities in global efficiency, clustering, small-worldness, modularity, or degree distribution within the patient cohort. However,

each patient exhibited evidence of a significant reorganization of nodes with high degree or high efficiency, known as "hub" nodes. Regions of the cortex that served as hubs in healthy brain networks tended to lose their hub status in comatose brain networks, and vice versa. These findings suggest that while the overall topological properties of complex brain networks in comatose patients did not differ quantitatively from those of the normal control group, consciousness likely relies on the specific anatomical localization of hub nodes in human brain networks [2].

A recent study delves into the Connectome Harmonic Decomposition (CHD), a method within the domain of Graph Signal Processing, which examines how a property of nodes in a graph (in this case, brain activation) treated as a signal correlates with the organization of the graph itself (in this case, structural connectivity). Due to its nonlinear nature, this graph-based approach offers a more comprehensive characterization compared to linear methods such as ICA and PCA [152]. The central premise of this research is that examining brain activity through the lens of connectome harmonics will yield insights about consciousness that complement the spatially-localized perspective prevalent in current neuroimaging studies. The findings reveal heightened structure-function coupling across various scales during states of unconsciousness induced by anesthesia or brain injury. This coupling can discern between behaviorally similar subcategories of brain-injured patients and track the presence of covert consciousness [152].

Challenges Graph analysis of RS-fMRI reveals a highly efficient organization of the brain network optimized toward a high level of local and global efficiency. Graph analysis can be automatically performed, with little a priori assumptions and with minimal bias. However, the results are often not intuitive and may be difficult to interpret [154].

They still rely on a choice of an atlas and are also prone to noise, which may result due to head motion, cardiac and respiratory effects as well as signals from white matter and cerebro-spinal fluid, and hence require many temporal data points for better estimation [51].

3.3 . Material and Method

3.3.1 . Datasets and atlas choice

Here, we work with "Static Anesthesia" and "Static DBS" data (see Table 2.3). In order to unify the results of both studies, the results presented here focus solely on those obtained from the CoCoMac atlas [11]. Comparison work with other atlases has been carried out on the "Static Anesthesia" dataset, but will not be presented here [89, 90].

3.3.2 . The Modular Hierarchical Analysis (MHA)

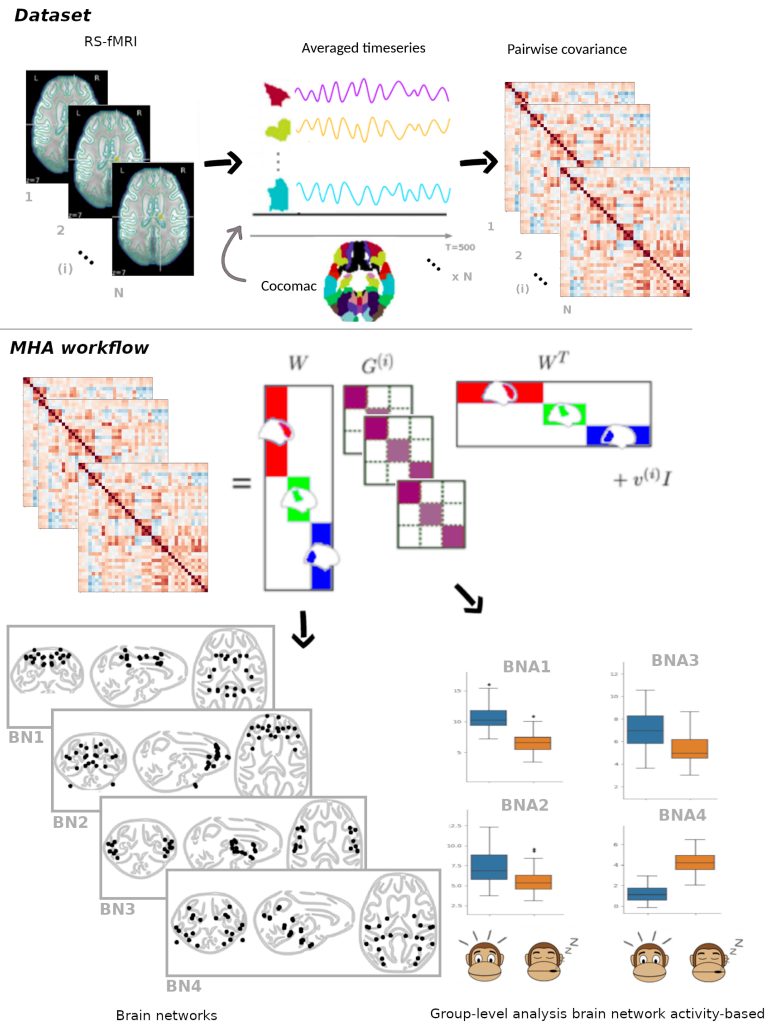


Figure 3.2: Graphical abstract

A latent variable model constitutes a statistical framework linking high-dimensional observed variables to low-dimensional latent variables, enabling the representation of intricate brain properties challenging to directly quantify. Utilizing empirical ROI-based covariance structures or FC matrices, the model deduces a distinct set of brain networks common to all experimental runs and their associated Brain Network Activities (BNA). This reinforces the identification of recurring patterns of brain activity during diverse anesthesia states [245].

In this model, the latent variables represent the BNAs specific to each run i . Within a regime of a small number of networks, the MHA linear latent variable model demonstrates enhanced reproducibility and interpretability compared to other methods like PCA or ICA [169]. The MHA approach aligns with the probabilistic PCA formulation [237]. Briefly, rs-fMRI observations $X^{(i)}$ stem from a linear projection of low-dimensional latent variables $Z^{(i)} \in \mathbb{R}^k$. Both observations

and latent variables are taken to follow a multivariate Gaussian distribution. This yields the generative model for observed data [169] :

$$\begin{aligned} Z^{(i)} &\sim \mathcal{N}(0, G^{(i)}) \\ X^{(i)} | Z^{(i)} = z^{(i)} &\sim \mathcal{N}(Wz^{(i)}, v^{(i)}\mathbb{I}) \end{aligned} \quad (3.1)$$

where $G^{(i)} \in \mathbb{R}^{k \times k}$ is the covariance of latent variables, k denotes the number of disjoint brain networks, and $v^{(i)} \in \mathbb{R}_+$ is the measurement noise. By capturing the low-rank covariance structure via the shared across-runs loading matrix W , the MHA model can reconstruct the covariance matrix in $\Sigma^{(i)}$:

$$\Sigma^{(i)} = WG^{(i)}W^T + v^{(i)}\mathbb{I} \quad (3.2)$$

$W \in \mathbb{R}^{p \times k}$ describes brain networks that are reproducible across the entire population. Each column j of W encodes the j^{th} brain network. For each run i , the matrix $G^{(i)}$ contains the latent variables of run i . More specifically, the j^{th} diagonal element of $G^{(i)}$ estimates the so-called BNA associated with the j^{th} brain network for run i . To compute the model parameters, the optimization maximizes the model log-likelihood \mathcal{L} between $\Sigma^{(i)}$ and the empirical covariance structure $K^{(i)} = X^{(i)}X^{(i)'} \in \mathbb{R}^{p \times p}$ across all runs as follows:

$$\begin{aligned} \mathcal{L} &= \sum_{i=1}^N p \log(2\pi) + \log \det \Sigma^{(i)} + \text{tr}(\Sigma^{(i)-1} K^{(i)}) \\ \hat{W} &= \underset{W: W^T W = \mathbb{I}; W \geq 0}{\text{argmax}} \mathcal{L} \end{aligned} \quad (3.3)$$

In comparison to PCA, the MHA model introduces a non-negativity constraint in addition to the orthonormal constraint. This incorporation empowers MHA to unveil distinct brain networks in matrix W and their corresponding run-specific BNA in $G^{(i)}$. Matrix W exhibits a block structure and is uniquely defined and identifiable. It can be conceptualized as a shared basis representing k non-overlapping brain networks across all runs.

As demonstrated in the work of Monti and colleagues [169], determining the optimal number of disjoint networks k in the model is treated as a hyperparameter tuning process. A leave-one-subject-out split is executed to create training and test sets. The MHA model is trained on the former, maximizing the log-likelihood \mathcal{L} over the unseen test set. The use of unseen data is crucial for evaluating model performance and generalization, guarding against overfitting. However, caution is exercised in maximizing log likelihood over unseen data for hyperparameter selection, given the potential for overfitting in our dataset where multiple anesthetic states were administered to the same monkey. To mitigate this risk, we include all of a monkey's data in the test set.

3.3.3 . Decoding Brain Network Activities (BNAs)

The proposed analysis considers only the k -diagonal elements in $G^{(i)}$. Let $S^{(i)} = (s_{G^{(i)}}^1, \dots, s_{G^{(i)}}^k)$ denote the associated individual activities across the k discovered BNs. For all runs $i \in [1, N]$, let $G \in \mathbb{R}^{N \times k}$ be the concatenation of these individual activities $S^{(i)}$. Our focus is on comparing the different BNAs contained in each column of G : $BNA^j = (s_{G^{(1)}}^j, \dots, s_{G^{(N)}}^j)$, where $j \in [1, k]$. Therefore, we interpret the BNAs as a metric of the activity within the corresponding brain networks. In essence, the BNAs reflect the amount of variability carried by each brain network. Our hypothesis posits that the off-diagonal entries of the latent variable covariances may not be the most discriminative features for the clinical question at hand. Non-zero off-diagonal values suggest redundancy in the data or some level of correlation between variables.

BNA-based statistical inference

Group-level analysis is conducted on BNAs using an atlas basis to emphasize the primary distinctions among anesthetic conditions. Examination with the Shapiro-Wilk test indicates that the BNAs deviate from normal assumptions. Consequently, pairwise nonparametric Wilcoxon signed-rank tests are employed to compare paired grouped BNAs. The null hypothesis (H_0) posits the absence of a significant difference between two awake/anesthetized conditions. To account for multiple comparisons, p-values undergo adjustment using the Benjamini/Yekutieli False Discovery Rate (FDR) correction.

BNA-based multivariate analysis

Consider the previously mentioned matrix $G \in \mathbb{R}^{N \times k}$ representing decomposed BNAs, and let $y \in \mathbb{Z}_+^N$ denote labels encoding anesthetic conditions (awake state or moderate/deep ketamine, propofol, or sevoflurane anesthesia). Supervised machine learning is applied to predict outcomes y based on input features G . The proposed classification relies on Support Vector Machines (SVM) with a Radial Basis Function (RBF) kernel, implemented using scikit-learn [186]. The gamma hyperparameter is automatically determined, while the C hyperparameter is set to 1. To mitigate overfitting in our small-sized dataset during training, bagging is introduced. This technique aggregates multiple models trained from the base SVM-RBF estimator by randomly selecting training subsets, contributing to the creation of a more robust predictor. As described earlier, the model is trained using a leave-one-subject-out splitting to generate both a training set and a test set. Model fitting incorporates five-fold cross-validation on the training set. It's important to note that the aforementioned classifier treats each class as non-ordinal, potentially miss the intrinsic relationships among the categories. To address this, we also evaluate the advantages of using another base classification estimator employing an Ordinal Logistic model with l2 regularization [203], as implemented in

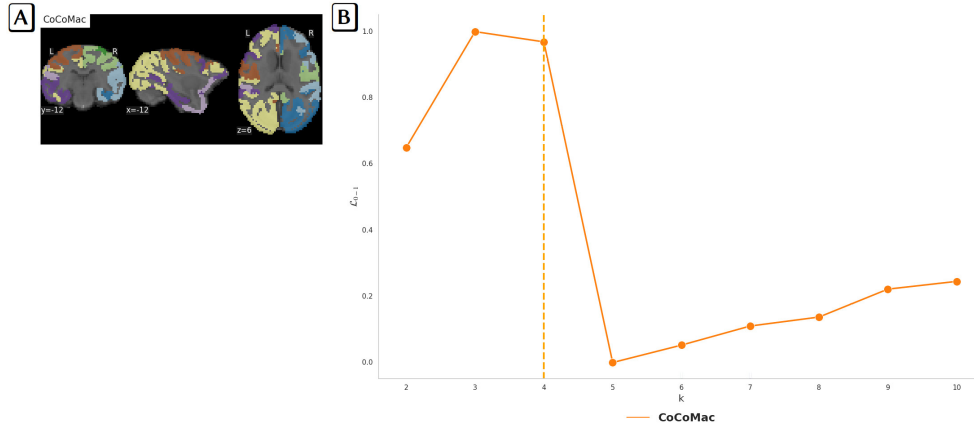


Figure 3.3: Illustrating how to select the best number of brain networks: A) the CoCoMac atlas, and B) associated 0-1 normalized log-likelihood \mathcal{L}_{0-1} on the unseen validation set for $k \in [2, 10]$. The optimal number of networks is represented by a vertical dashed line: $k=4$ brain networks for the CoCoMac atlas.

mord [188]. The regularization parameter is set to 1. In both cases, the specified hyperparameters are not assessed in an internal cross-validation process.

Brain network importance

Ultimately, a pertinent question arises: which brain networks contribute to the different predictions? Employing a model-agnostic permutation importance technique, as implemented in scikit-learn [186], enables the assessment of feature importance. This method involves randomly permuting the values of a feature and assessing its impact on the model's performance. By comparing the model's performance with permuted features to its original performance, one can discern which features influence most the predictions. Features exhibiting a significant performance drop post-permutation are deemed more critical. This technique aids in pinpointing the features or BNAs and their associated brain networks that exert the most substantial influence on the model's performance. It provides valuable insights into the relationships between BNAs and anesthetic conditions.

3.4 . Results on Anesthesia Dataset

3.4.1 . Consciousness connectivity can be decomposed into few consistent brain networks

The selection of an atlas determines the quantity of input regions, denoted as p supplied to the model. Maximizing \mathcal{L} results in the identification of four optimal brain networks ($k = 4$) for the CoCoMac atlas (Figure 3.3). While the likelihoods for $k = 3$ and $k = 4$ are close for the CoCoMac atlas, we opt for $k = 4$ to maximize the coverage of the ROIs (see Appendix 4 for a detailed listing

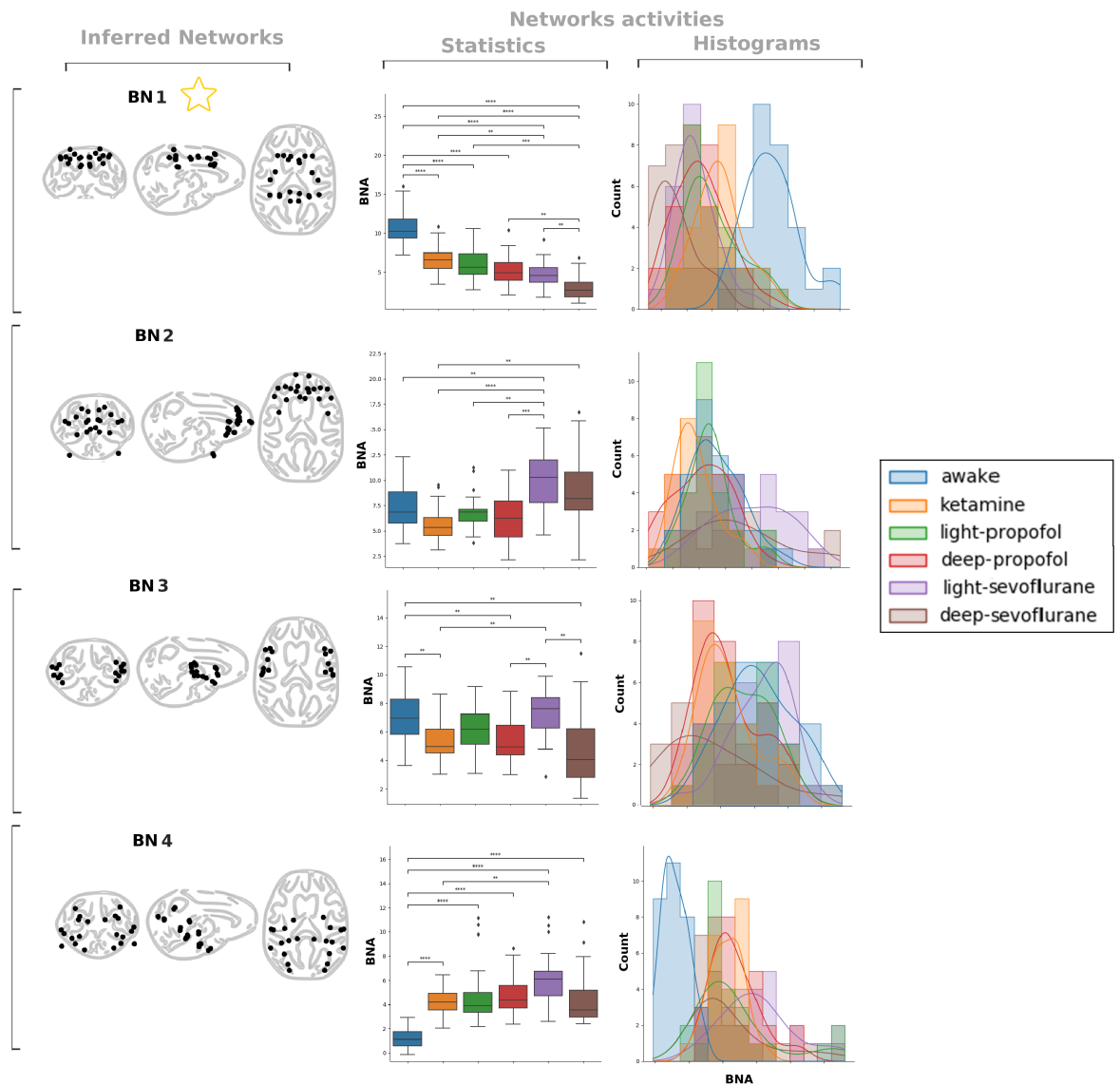


Figure 3.4: Inferred Brain Networks (BNs) from the CoCoMac ($k=4$) atlas, and associated pairwise statistical analysis of BNA. p -value annotation legend: **: $1.00e - 03 < p \leq 1.00e - 02$, ***: $1.00e - 04 < p \leq 1.00e - 03$, ****: $p \leq 1.00e - 04$

of coverage for $k = 3$). Recall that the MHA constraints drive this coverage by conditioning the loading matrix W to have at most one non-zero entry per row, imposing sparsity with non-negativity.

3.4.2 . Brain network 1 intersects the Global Neuronal Workspace (GNW)

Pairwise statistics (Figure 3.4), bring attention to the most substantial differences in terms of BNAs between conditions, i.e., states of consciousness. In

	name	hemi	location
CCp	posterior cingulate cortex	left, right	cingulate cortex
CCa	anterior cingulate cortex	left, right	cingulate cortex
S1	primary somatosensory cortex	left, right	parietal cortex
PC1	inferior parietal cortex	left, right	parietal cortex
PCm	medial parietal cortex	left, right	parietal cortex
PCip	intraparietal cortex	left, right	parietal cortex
PCs	superior parietal cortex	left, right	parietal cortex
M1	primary motor cortex	left, right	frontal cortex
FEF	frontal eye field	left, right	frontal cortex
PMCm	medial premotor cortex	left, right	frontal cortex
PMcdl	dorsolateral premotor cortex	left, right	frontal cortex

(A)

	name	hemi	location
TCpol	temporal polar	left, right	temporal cortex
PFCoi	orbitoinferior prefrontal cortex	left, right	frontal cortex
PFCom	orbitomedial prefrontal cortex	left, right	frontal cortex
PFCol	orbitolateral prefrontal cortex	left, right	frontal cortex
PFCpol	prefrontal polar cortex	left, right	frontal cortex
PFCvl	ventrolateral prefrontal cortex	left, right	frontal cortex
PFCm	medial prefrontal cortex	left, right	frontal cortex
PFCcl	centrolateral prefrontal cortex	left, right	frontal cortex
PFCdm	dorsomedial prefrontal cortex	left, right	frontal cortex
PFCdl	dorsolateral prefrontal cortex	left, right	frontal cortex
CCs	subgenual cingulate cortex	left, right	cingulate cortex

(B)

	name	hemi	location
TCs	superior temporal cortex	left, right	temporal cortex
A1	primary auditory cortex	left, right	temporal cortex
A2	secondary auditory cortex	left, right	temporal cortex
G	gustatory cortex	left, right	gustatory cortex
PMCvl	ventrolateral premotor cortex	left, right	frontal cortex
Ip	posterior insula	left, right	insular cortex
Ia	anterior insula	left, right	insular cortex
S2	secondary somatosensory cortex	left, right	parietal cortex

(C)

	name	hemi	location
Amyg	amygdala	left, right	temporal cortex
TCc	central temporal cortex	left, right	temporal cortex
TCi	inferior temporal	left, right	temporal cortex
PHC	parahippocampal cortex	left, right	temporal cortex
HC	hippocampus	left, right	temporal cortex
TCv	ventral temporal cortex	left, right	temporal cortex
VACv	anterior visual area (ventral)	left, right	occipital cortex
V2	visual area 2	left, right	occipital cortex
VACd	anterior visual area (dorsal)	left, right	occipital cortex
V1	visual area 1	left, right	occipital cortex
CCR	retrosplenial cingulate cortex	left, right	cingulate cortex

(D)

Table 3.1: Listing of Brain Networks (BNs) inferred from the CoCoMac atlas: A) the BN1 highlighting the difference between the awake state and anesthesia (the BN1 indicated by a star in Figure 3.4-A), B) BN2, mainly prefrontal and B) BN3 mainly temporal, including auditory cortex, D) the inferred BN4 that is driven by the visual pathway. The detected GNW areas are depicted in blue, and the associated sensory areas in green.

contrast to the sliding window synchronization patterns [92], our statistical approach highlights a larger number of significant differences, notably emphasizing distinctions between the awake state and all anesthetic conditions. This outcome underscores the significance of Brain Network 1 (BN1), marked with a star in Figure 3.4. Intriguingly, a focused examination of this network (BN1 in Figure 3.4) reveals that the ROIs underlying this difference exhibit perfect symmetry and closely align with the macaque GNW nodes [243]. These are the posterior cingulate cortex (CCp), anterior cingulate cortex (CCa), intraparietal cortex (PCip), Frontal Eye Field (FEF), dorsolateral prefrontal cortex (PFCdl), prefrontal polar cortex (PFCpol), and dorsolateral premotor cortex (PMcdl), encompassing sensory regions like the primary motor cortex (M1), primary somatosensory cortex (S1), primary visual cortex (V1), and primary auditory cortex (A1) (Table 3.1-A). Although obtained through unsupervised constraints, BN1 notably aligns closely with the GNW theory, covering 7 out of 11 nodes.

3.4.3 . Which network best predicts the depth of anesthesia from BNAs?

Utilizing the BNA distributions, we can distinctly differentiate the wakefulness state from the anesthesia state, irrespective of the administered anesthetics suppressing consciousness (Figure 3.4). Noteworthy variations persist among different anesthetics. Consequently, we conduct a multivariate analysis of the BNAs, incor-

CoCoMac			
SVM RBF	All	DeepModerate	Anesthesia
train	0.76 ± 0.06	0.84 ± 0.02	1.0 ± 0.0
validation	0.58 ± 0.09	0.78 ± 0.14	0.99 ± 0.012
test	0.24 ± 0.05	0.84 ± 0.05	0.97 ± 0.03
Ordinal Logistic	All	DeepModerate	Anesthesia
train		0.7 ± 0.02	1.0 ± 0.0
validation		0.73 ± 0.08	0.99 ± 0.02
test		0.79 ± 0.09	0.98 ± 0.01

Table 3.2: Brain activity based prediction of acquisition conditions using the CoCoMac atlas. The Balanced Accuracy (BAcc) metric is used to evaluate model performances. Three settings are considered: the awake state and all anesthetics are considered separately (All), the anesthetics are grouped by dosage (DeepModerate), or all anesthetics are encoded in the same group (Anesthesia). Two models are evaluated: the SVM-RBF and Ordinal Logistic models.

porating the downstream task of anesthetic state classification. To delve deeper into BNAs, SVM-RBF and Ordinal Logistic models are employed to address three classification tasks, each corresponding to distinct sets of target labels: 1) treating the awake state and each anesthetic individually (label set: All), 2) categorizing anesthetics based on sedation level (label set: DeepModerate), or 3) grouping all anesthetics into a single category (label set: Anesthesia). BNA-driven predictions are listed in Table 3.2. The best performance, unsurprisingly, is the separation between levels of vigilance (unconscious/conscious - label set : Anesthesia). For the different levels of sedation (label set : DeepModerate), performance is good, well above chance (which is ~ 0.33), demonstrating its relevance in characterizing the depth of anesthesia. On the other hand, for the different conditions taken separately (label set : All), performance drops drastically.

In the subsequent experiments, the SVM-RBF has superior performance and is retained for the rest of the study. Our current objective is to offer a deeper understanding of the identified brain networks. An analysis of brain network importance highlights two significant networks (Figure 3.5-A). Firstly, BN4 (Table 3.1-B), primarily located in the temporo-occipital region, encompasses the visual pathway. This may correspond to the observation that awake monkeys typically have open eyes, potentially experiencing visual stimulation. Secondly, BN1, largely parieto-cingular and detailed in the preceding paragraph, contains the majority of GNW nodes. In contrast, networks 2 and 3 are of relative minor importance in distinguishing the different conditions. BN2 (cf. Table 3.1-C) is predominantly prefrontal, while BN3 (cf. Table 3.1-D) is predominantly temporal, encompassing in particular the auditory cortex.

3.4.4 . Influence of time window size on predictions: sensitivity

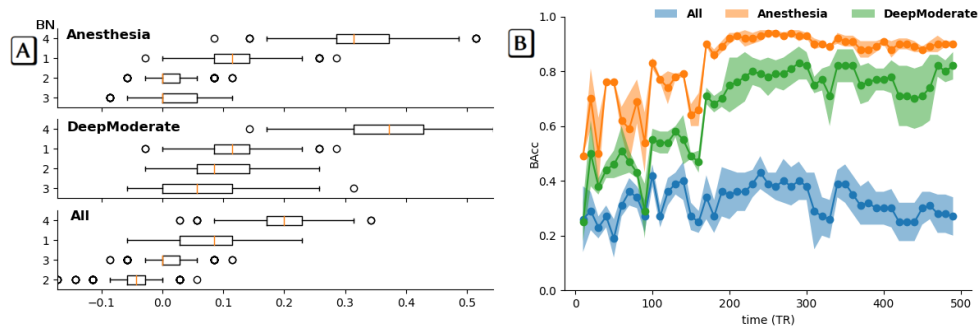


Figure 3.5: Illustration of the results of the study on network influence and the impact of time window size in the prediction of different acquisition conditions. A) Brain networks (BNs) exerting the most substantial influence on prediction, as determined by the feature importance analysis. Three scenarios are examined: individual consideration of the awake state and each anesthetic (All), grouping of anesthetics based on dosage (DeepModerate), and encoding all anesthetics in a single group (Anesthesia). The x-axis values represent the extent to which model performance decreases with random shuffling, using balanced accuracy as the performance metric. The level of randomness is gauged through repeated iterations of the process. Positive values indicate significant features, while negative values suggest predictions more accurate than actual data. This can occur when a feature is unimportant, but randomness coincidentally improves predictions—an observed behavior with small datasets more susceptible to random errors. B) Learning curves concerning acquisition time. The Balanced Accuracy (BAcc) metric is employed to assess the performance of the SVM-RBF model. To achieve satisfactory performance, a minimum of 200 TR is required.

study

Decoding the level of consciousness from neural activity holds significant potential for clinical applications, specifically in the development of innovative tools for objectively monitoring the depth of anesthesia. To assess the method’s adaptability, we examine its learning curve in relation to acquisition duration. This experiment employs the three labeling settings outlined in the previous paragraph: All, DeepModerate, and Anesthesia (Figure 3.5-B). The duration of the truncated time series varies from 10 to 500 TR in increments of 10 TR. Despite the limited dataset size, the plots suggest that a run length of 200 TR yields accurate performances (~ 0.8 balanced accuracy for the Anesthesia setting). Thus, the proposed solution requires a 200 TR duration to make reliable predictions. This learning curve analysis for a given time series duration represents the practical method for establishing a steady state for the model. While additional data would be necessary for confirmation, this result provides an estimate of the minimum buffer size for such an approach.

3.4.5 . Conclusion

Consistent with the GNW theory of consciousness, the brain network involving the frontal, parietal, and cingulate cortices emerges as crucial in discerning

consciousness levels. This pioneering approach enables the development of an interpretable brain decoding model, offering a distinctive signature of consciousness and anesthesia-induced loss of consciousness. The model does not rely on any biological assumption about anesthetics, and provides results that are relatively insensitive to different anesthetics. Thus, one might assume that we are getting a general signature of consciousness, disentangled from potential markers related to a particular anesthetic effect.

3.5 . Results on DBS dataset

3.5.1 . Consciousness connectivity can be decomposed into few consistent brain networks

Consistently with the MHA literature, we find that the optimal number of networks in the decomposition is low ($k = 2$ or $k = 3$). We perform a leave-one-subject-out cross-validation, repeated three times, with a different individual left out between trials. The highest log-likelihood is obtained for $k = 3$ on two over the three trials. We also clearly see a drop in the log-likelihood at $k = 4$. Thus, we choose to work with $k = 3$ in the rest of this study (Figure 3.6).

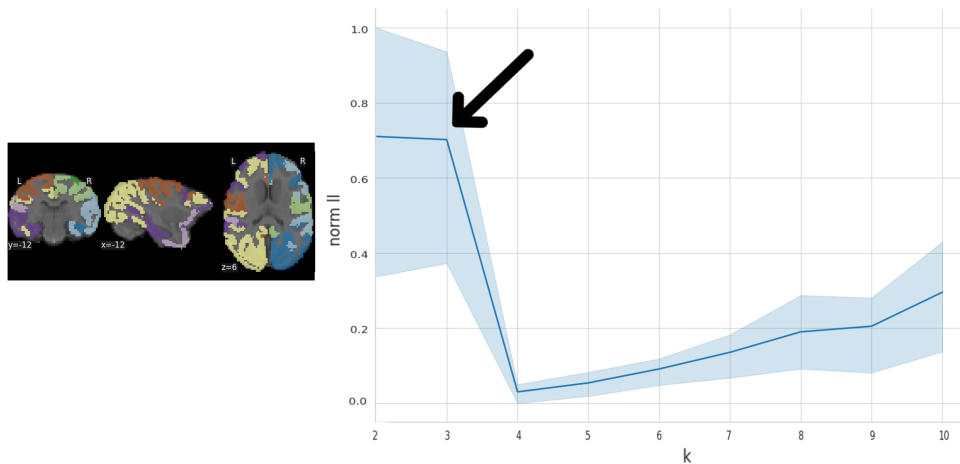


Figure 3.6: CoCoMac atlas and associated 0-1 normalized log-likelihood on three different unseen test sets for $k \in [2, 10]$.

3.5.2 . Identified networks

BN1 is almost symmetrical: the network consists of 36 ROIs, and only 4 do not appear together in the opposite hemisphere. This network emphasizes the back of the brain. It is mainly parieto-temporal (12 ROIs in the parietal cortex, 9 in the temporal cortex). The occipital cortex is also represented (8 ROIs), with the presence of all visual areas (V1, V2, VACd, and VACv) as well as the cingular cortex (Table 3.3). It is also interesting to note that it contains 7/8 sensory ROIs,

and 2 nodes (CCp and PCip) involved in the monkey's Global Neuronal Workspace (GNW) theory (respectively highlighted in green and yellow in Table 3.3-a).

BN2 has 20 ROIs distributed symmetrically in both hemispheres and 8 regions present only in the left hemisphere. Among the symmetrical regions, a large number of monkey GNW nodes appear (5/11) (Figure 3.3-b). This network is predominantly prefrontal (16 ROIs) and cingulate.

Finally, BN3 is almost entirely asymmetric with ROIs in the right hemisphere. Note that the electrode was implanted in the right hemisphere. It is mainly prefronto-temporal.

3.5.3 . Reconciling the front and back of the brain in the processing of conscious information

BN1 and BN2 are posterior and anterior, respectively. They highlight a significant difference in BNAs between the awake and the anesthetized states (stim-off). Such a trend is not captured in BN3 (Figure 3.7). It highlights the importance of the first two networks when processing conscious information. It also shows that the third network mainly captured signal variations related to the DBS direct stimulation in the right hemisphere without any link with the neuronal correlate of consciousness. In particular, it is interesting to note that BN3 does not contain any of the regions impacted by the electrode artifact (identified in section 2.2.4), which suggests an effect related to the repercussions of the stimulation on the whole cortex rather than a local effect.

3.5.4 . Networks capture differences induced by the effective DBS condition

We can notice a BNA increase in the stim-on-5v acquisition condition (effective DBS) compared to the stim-off in the three networks. Such a BNA increase is not observed in the control state at 5V (stim-cont-on-5V). Finally, no significant difference is observed between the BNA in the awake state and under stim-on-5v.

3.5.5 . Conclusion

To conclude, the MHA model highlights two relevant networks whose BNAs support that conscious information is processed in anterior (prefronto-cingular) - posterior (parieto-cingular) networks and confirms the positive impact of the stim-on-5v DBS stimulation on the consciousness signatures restoration. The model also highlights a network capturing the stimulation effects. The MHA model disentangles the different sources of signal variability. This model could thus be helpful to evaluate the cortical impact of a cerebral stimulation and thus identify the regions at risk of collateral damage. Moreover, in future work, we would like to address one issue. We would like to remove the artifactual voxels before averaging the time series. By doing so, we will eliminate one known source of variability, which will improve the identification of the other sources of variability.

label	hemi	location	name
TCs	right, left	temporal cortex	Superior temporal cortex
TCc	right, left	temporal cortex	Central temporal cortex
VACv	right, left	occipital cortex	Anterior visual area (ventral)
V2	right, left	occipital cortex	Visual area 2
VACd	right, left	occipital cortex	Anterior visual area (dorsal)
V1	right, left	occipital cortex	Visual area 1
PFCcl	right	frontal cortex	Centrolateral prefrontal cortex
A2	right, left	temporal cortex	Secondary auditory cortex
CCR	right, left	cingulate cortex	Retrosplenial cingulate cortex
CCp	right, left	cingulate cortex	Posterior cingulate cortex
S2	right, left	parietal cortex	Secondary somatosensory cortex
S1	right, left	parietal cortex	Primary somatosensory cortex
M1	right, left	frontal cortex	Primary motor cortex
PCI	right, left	parietal cortex	Inferior parietal cortex
PCm	right, left	parietal cortex	Medial parietal cortex
PCip	right, left	parietal cortex	Intraparietal cortex
PCs	right, left	parietal cortex	Superior parietal cortex
PHC	left	temporal cortex	Parahippocampal cortex
TCv	left	temporal cortex	Ventral temporal cortex
A1	left	temporal cortex	Primary auditory cortex

(A)

label	hemi	location	name
PFCol	right, left	frontal cortex	Orbitolateral prefrontal cortex
PFCpol	right, left	frontal cortex	Prefrontal polar cortex
CCs	right, left	cingulate cortex	Subgenual cingulate cortex
PFCm	right, left	frontal cortex	Medial prefrontal cortex
CCa	right, left	cingulate cortex	Anterior cingulate cortex
PFCdm	right, left	frontal cortex	Dorsomedial prefrontal cortex
FEF	right, left	frontal cortex	Frontal eye field
PFCdl	right, left	frontal cortex	Dorsolateral prefrontal cortex
PMCm	right, left	frontal cortex	Medial premotor cortex
PMCdl	right, left	frontal cortex	Dorsolateral premotor cortex
PFCoi	left	frontal cortex	Orbitoinferior prefrontal cortex
la	left	insular cortex	Anterior insula
PFCom	left	frontal cortex	Orbitomedial prefrontal cortex
G	left	gustatory cortex	Gustatory cortex
PMCVl	left	frontal cortex	Ventrolateral premotor cortex
Ip	left	insular cortex	Posterior insula
PFCvl	left	frontal cortex	Ventrolateral prefrontal cortex
PFCcl	left	frontal cortex	Centrolateral prefrontal cortex

(B)

label	hemi	location	name
TCpol	right, left	temporal cortex	Tempolar polar
PFCoi	right	frontal cortex	Orbitoinferior prefrontal cortex
la	right	insular cortex	Anterior insula
PFCom	right	frontal cortex	Orbitomedial prefrontal cortex
TCi	right, left	temporal cortex	Inferior temporal
PHC	right	temporal cortex	Parahippocampal cortex
G	right	gustatory cortex	Gustatory cortex
PMCVl	right	frontal cortex	Ventrolateral premotor cortex
Ip	right	insular cortex	Posterior insula
HC	right, left	temporal cortex	Hippocampus
PFCvl	right	frontal cortex	Ventrolateral prefrontal cortex
TCv	right	temporal cortex	Ventral temporal cortex
A1	right	temporal cortex	Primary auditory cortex

(C)

Table 3.3: Listing of Brain Networks (BNs) inferred from the CoCoMac atlas: A) the BN1 emphasizes the back of the brain, B) the inferred BN2 that is predominantly prefrontal and cingulate, C) BN3, asymmetric, right-hand side. The detected GNW areas are depicted in yellow, and the associated sensory areas in green.

3.6 . Discussion

In this study, we introduced an innovative approach for analyzing RS-fMRI data acquired during distinct states of consciousness in non-human primates. Our method successfully identified anatomically relevant cortical brain networks associated with different consciousness states. The uniqueness of our framework stems from the application of a constrained linear latent variable model, yielding BNA across identifiable and non-overlapping brain functional networks. Within the recorded brain activity, specific properties (the BNAs) are highlighted, showcasing predictable distinctions between conscious and unconscious states. These properties could serve as reliable indicators for accurately and objectively classifying individuals based on their state of consciousness. These findings reinforce the importance of quantitative biomarkers of RSN for assessing the widespread effects of brain stimulation techniques that transcend the stimulated network. Understanding how CT-DBS modulates brain activity at the level of RSNs can be essential to advance in the design of more effective and personalized therapeutic strategies.

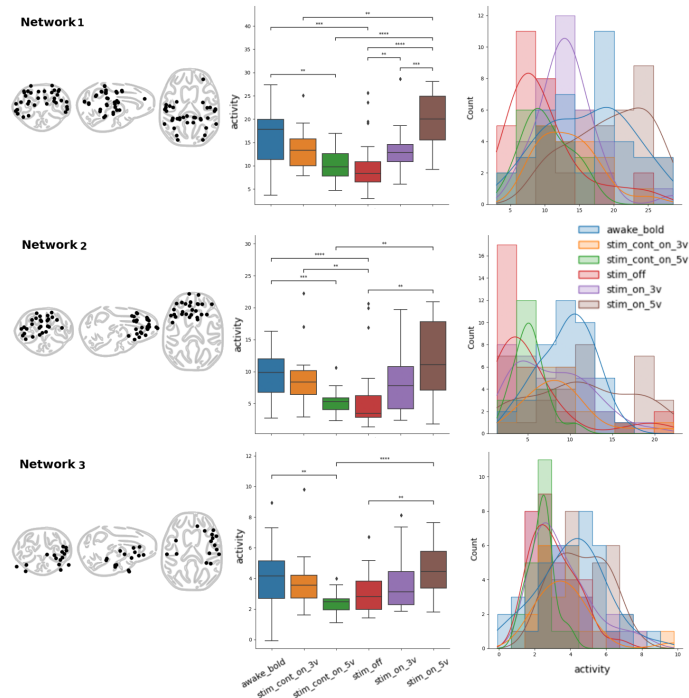


Figure 3.7: Networks inferred by the MHA model from the time series and associated statistical analysis on BNAs and sample distributions. *p*-value annotation legend: **: $1.00e - 03 < p \leq 1.00e - 02$, ***: $1.00e - 04 < p \leq 1.00e - 03$, ****: $p \leq 1.00e - 04$

Parcellation and number of networks

Parcellation Generally, in neuroimaging studies, specific templates that define anatomical regions of the brain are used. The atlas serves as a common spatial landmark for analysis and interpretation of the data. In our case, it allows us to map and compare brain activity across individuals or groups.

When studying functional connectivity, the choice of brain atlas is usually a trade-off between the characterization of brain structure and signal averaging for data and noise reduction. Fine-grained brain areas can capture brain activity descriptions with more functionally-specific regions at the expense of signal-to-noise ratio loss. Fine-grained brain areas also generate high-dimensional input features that are challenging to learn in generic predictive models, a problem known as the curse of dimensionality.

To test the robustness of the results obtained, a second functional atlas is used (CIVMR; containing 222 ROIs), on the anesthesia dataset. As this atlas contains more ROIs, we unsurprisingly also find more BNs (Appendix 3). Some networks strongly resemble those found on the CoCoMac cortical atlas, others are the decomposition of some of them, and finally, new, sub-cortical ones make their appearance. The networks are always symmetrical.

It could also be interesting to test a data-driven approach at the voxel level

without relying on an atlas-based framework.

Number of networks There's substantial research on resting-state networks in humans, often involving 7-17 networks. In macaque, the number of networks ranges from 8 to 30 networks [199, 105, 253]. In this study, the number of optimal brain networks varies across three atlases from 4 to 7. Does a $k=4$ adequately capture this organization?

When using classical ICA, there are presently no firmly established criteria to assist in determining the optimal number of ICs for a given dataset [105]. Methods using ICA select between 20 and 30 components, then decide to eliminate some of them after processing, as they may represent noise (correlation with things other than grey matter, for example). For methods using dictionary learning, a criterion based on network modularity is generally used. For example, in the mouse lemur primate, 6 RSNs have been identified, two of which include subcortical regions [80]. However, the CoCoMac is cortical, so finding only 4 networks, in our work, is not aberrant. The method proposed here dispenses with a priori selection and automatically selects a number of components. The optimal number of brain networks depends on the input atlas, but a clear decision can be made by monitoring the log-likelihood. Our model scales to any dataset and the associated hyperparameters are tuned numerically from maximum likelihood estimation. The discovered brain networks are tailored, spatially consistent, and symmetric.

Cerebral networks identified and associated activities

In the present work, we highlight differences directly related to interacting cortical regions by calculating statistics on BNAs grouped by level of consciousness. In the anesthesia dataset, we found two main brain networks that effectively differentiate between the awake and all anesthetic states, underscoring their significance as a robust finding. One of them, previously described in the literature, supports the above hypothesized GNW theory, and the other emphasizes the visual network. These networks 1 and 4 are actually quite similar to networks 1 and 4 found in the small monkey (see 3.2.1), highlighting the fronto-parietal and occipital cortices. These findings are also intriguing when juxtaposed with the global disconnection syndrome observed in UWS patients, characterized by a disconnect between higher-order association cortices and primary cortical areas. In contrast, MCS patients exhibit preserved large-scale cortical networks associated with language and visual processing [2, 210, 81].

Low-level cortical networks

The BN4 network is primarily associated with the visual pathway, and we wonder if it could be an artifact of our experimental setting, considering that awake non-human primates (NHPs) had their eyes open, potentially exposing them to visual stimulation. However, previous studies using RS-fMRI in both humans and NHPs have consistently identified visual networks even under anesthesia [16, 80].

Additionally, it is expected that functional connectivity remains intact in low-level sensory cortices across different sedation stages, including auditory and visual networks [25].

Network 3, which involves the auditory cortex, is associated with BNAs that do not significantly contribute to the differentiation between anesthetic and awake states. Conversely, in the case of the visual network, BNAs are lower in the awake state compared to the anesthetized state. This unexpected outcome may indicate reduced integration of a more complex network present during wakefulness under anesthesia. Consequently, this over-presence of the visual network in anesthetized NHPs might stem from fewer cross-network interactions in the unconscious state.

For additional clarity and as a potentially significant control measure, we could exclude saccade windows from the awake data and rerun the model. Saccades might also introduce slight motion artifacts specifically in the awake state, or artificially induce neural coherence in the awake condition, potentially affecting the modeling outcomes. In particular, the FEF, responsible for saccadic eye movements for the purpose of visual field perception and awareness, as well as for voluntary eye movement, appears as a hub of BN1.

High level cortical networks

Networks 1 and 2 involve regions associated with high-level cortical networks (including GNW regions). However, associating them specifically with known human high-level cortical networks, linked to high-level information processing, remains complex. In the BN1, the majority of constituent nodes were cortical areas that match GNW nodes from a previous publication [243]. As DMN activity is known to be consistently suppressed by different anesthetics and a correlation was observed between the level of connectivity within the DMN and the severity of clinical consciousness impairment [2], we sought to link one of the brain networks to a DMN in monkeys. Certain regions of networks 1 and 2 are hubs of the default-mode-like network (DMN-like) of marmosets or macaques, e.g. the CCp and PFCdl [80]. While the former is in network 1, the latter is in network 2 (predominantly prefrontal). Moreover, network 1 also overlaps with the Dorsal Attentional Network (DAN) of the macaque (PCip, FEF) [208]. Overall, it is known that propofol-induced decrease in consciousness linearly correlates with decreased corticocortical connectivity in frontoparietal networks [25], which is consistent with the decrease in BNA in networks 1 and 2 for both datasets. BN3 is difficult to classify specifically as a primary or higher-order network. In fact, it contains the auditory cortex and the second somatosensory network, which tends to classify it as a primary network, but also a specific region, the insula. The insula is a functionally heterogeneous brain area that participates in a wide variety of behaviors involving interoceptive awareness and the mediation of emotion, functions that have placed this structure at the center of the Saliency Network (SN) which also contains anterior cingulate cortex (CCa) as a hub [16]. This network, and the insula in particular, has been proposed to serve as a sentinel, detecting salience from a vast array of constant

streams of stimuli, and, perhaps, serving as a signal to initiate engagement of various cognitive control networks, with concomitant disengagement of the DMN. In our case, the cinguloinsular SN network is not apparent: Cc and insula appear in BN1 and BN3 networks respectively. Although Hutchison et al. [105] also found a cinguloinsular component in anesthetized macaques, as well as Belcher et al. [16] in marmoset data, Mantini et al. (2013) [157] identified a salience-like network in humans (a "cinguloinsular" component) that had no correspondence in their monkey results. The authors suggest that the absence of a salience-like network in rhesus monkeys could be related to the conditions under which the resting-state experiments were conducted: the monkeys received liquid reinforcement when they paid attention to the stimuli on a screen in front of them. Because of this important methodological difference between their human and monkey acquisition paradigms, it is perhaps not surprising that they reported differences in network patterns within and between the species. We should point out that the same effect could be present in our data, with awake macaques also receiving a reward, while anesthetized monkeys did not, which could explain why SN does not emerge clearly in our data.

On the DBS dataset, networks 1 and 2 are the ones used to differentiate states of consciousness. They are more difficult to compare with known networks, as they each contain more ROIs, but clearly show a front vs. back of the brain separation. For this dataset, it would seem appropriate to repeat the analysis without averaging over the voxels impacted by the electrode. This would free us from this source of variability and potentially allow us to obtain more networks. However, the study of DBS data is limited by the data acquisition scheme (only one NHP has all the conditions (effective and control) for DBS stimulation). We could also repeat this analysis only on the half of the brain without the electrode. Given the symmetry of the networks on the anesthesia dataset, it would be interesting to see how the electrode affects the networks.

Sensitivity analyses

Recognizing the limitation that data might be sparse, it could be useful to provide analyses showing that if the model is re-run on a few individual NHPs, qualitatively similar results in each one, emerge. This remains complicated to achieve because of the sparsity of the conditions, but an experiment in which one subject is removed shows similar results and reinforces the method's stability.

Limitations

Non overlapping networks With the MHA model, the joint participation of ROIs in several networks is not possible, so the discovered networks are necessarily non-overlapping, as imposed by the optimization constraints. Overall, the MHA approach yields few tailored brain networks and associated BNAs, which promotes

interpretability.

Correlation inter-networks Our network-based analysis is not informative about inter-network correlation. It might be interesting to extend the analysis by studying inter-network correlation according to acquisition conditions, for example, by calculating the correlation between the BNs' FC. This analysis could be performed on sFC or dFC to assess network dynamics. This would make it possible to assess the temporal evolution of interaction between BNs.

Limiting overfitting Given the typically small number of subjects in fMRI acquisition for non-human primates, the risk of overfitting is acknowledged. Overfitting arises when a model becomes overly complex, fitting noise or random fluctuations in training data rather than the underlying patterns or relationships of interest. In datasets with a limited number of individuals, overfitting can be especially problematic due to insufficient examples to capture the true data distribution. This small sample size may constrain the statistical power, complicating the detection of true effects and increasing the risk of false negatives.

To address this limitation and enhance the dataset size without additional acquisition, in the future, we could explore the benefits of incorporating simulated data [128, 56] (cf. section 2.3.1).

There are also alternatives, such as Prime-DE (cf. section 2.3.2), which provide free access to an increasing number of fMRI recordings.

Potential clinical applications of the proposed framework

The brain is a highly interconnected system consisting of multiple regions that communicate and interact with each other. The proposed framework tests a new method that is added to the state-of-the-art data-driven strategies. It decomposes the FC into brain networks. This provides valuable information about the functional organization of the brain and allows the study of individual differences in brain activity. In fact, each individual has a unique pattern of brain connectivity. Characterizing these individual differences can provide insights into variations in cognitive abilities, behavior, and susceptibility to brain disorders. In addition, understanding individual differences in network connectivity may facilitate the development of personalized treatment approaches, where interventions can be tailored to target specific network dysfunctions in a given individual. The study of network-level properties offers the potential to identify specific biomarkers that can be used for diagnostic purposes, disease monitoring, or prediction of treatment outcomes. Indeed, many neurological and psychiatric disorders are characterized by alterations in brain connectivity. By comparing network properties between healthy individuals and patients, it is possible to identify aberrant connectivity patterns associated with specific disorders. This approach can lead to a better understanding of the

underlying mechanisms of the disorders and potentially help to develop diagnostic or therapeutic strategies.

Teenagers are told, warned and "unfiltered" movements have even sprung up. What we see on networks is not reality. The MHA is certainly not a social network. But like social networks, it shows a fixed reality. Even if, unlike them, MHA is not based on an image taken at a given moment t , chosen and most of the time a biased reflection of reality, the fact remains that the data is averaged, and stripped of its temporal component. Reality is, therefore, distorted, partial, and incomplete. In the next chapter, we propose to integrate the temporal dimension with the spatial dimension, by working on dFC matrices. We move slightly away from networks in an attempt to model transitions between different functional connectivity patterns. We're looking to see how the different connectivity patterns fit together, how they move from one to another. These patterns have already been shown to be markers of consciousness, and the importance of exploiting them no longer needs to be proven. Considering them as a group, as a cluster, may initially seem to facilitate analysis, but it doesn't seem to be the most relevant way of characterizing these cerebral patterns that come and go.

The image on the next page is a reproduction of René Magritte, *La condition humaine*, 1933

Part III

Seeing beyond reflection: latent variable models for studying states of consciousness



4 - From networks to dynamic functional connectivity analysis

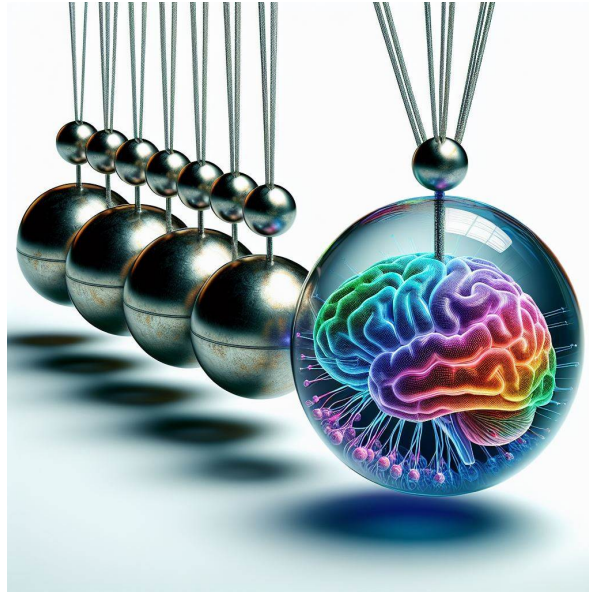
Contents

4.1	Introduction	109
4.1.1	Material and Method	110
4.1.2	Application to consciousness : state-of-the-art	111
4.1.3	Overcoming the limitations of K-means clustering	114
4.2	Material and Method	118
4.2.1	Dataset and data partitioning	118
4.2.2	Model architecture	118
4.2.3	Maps of Predictive Connections	120
4.3	Results on Anesthesia dataset	120
4.3.1	Performance of the classifier	120
4.3.2	Towards modeling the brain patterns dynamic	123
4.3.3	Maps of predictive connections	124
4.4	Conclusion	126
4.5	Discussion	126

In the previous study, it was assumed that the functional connectome remains constant (i.e., temporally stationary) during the recording period in RS-fMRI, but functional connectome between regions are actually linked to dynamic brain activity over time [106, 7, 14, 99, 245]. According to the pioneering study by *Allen et al. (2014)* [7], the hypothesis of temporal stationarity is indeed practical but represents an oversimplification of whole-brain connectivity analysis. This study, which inspired the method used by our team in previous works, proposes to explore resting-state functional dynamics with a data-driven approach, without a priori assumptions. This allows for the unveiling of stable connectivity patterns and changes directly from the data.

4.1 . Introduction

Dynamic network analysis using RS-fMRI offers valuable insights into the inherently dynamic characteristics of the brain, thus providing an effective approach for automated biomarker identification. Specifically, studies in humans and non-human primates, both within our team and elsewhere [14, 245, 60, 234], have



explored dynamic RS-fMRI analysis to discover markers of consciousness. First, we will describe the methodology employed in previous research conducted by our team to explore the dynamic connectome, followed by its application in the context of consciousness modulation.

4.1.1 . Material and Method

The following straightforward data-driven method for assessing dFC relied on established techniques such as sliding time-window correlation and K-means clustering of windowed correlation matrices [7].

The observed dFC states exhibit high replicability and partially deviate from stationary connectivity patterns, challenging existing descriptions of interactions among large-scale networks. Moreover, the differential occurrence of specific FC states over time inspires theories regarding their functional roles and relationships with various conditions and states of consciousness [7].

Dataset

Here, the "Dynamic Anesthesia" dataset is used (see Table 2.3 and 2.2.4 for associated methodology).

Unsupervised clustering to find brain patterns

Then, K-means clustering was employed to detect recurrent patterns of FC over time and across subjects [146]. These clusters are conceptualized as "brain states" or "Brain Patterns" (BPs) akin to EEG microstates, which denote brief periods characterized by quasi-stable scalp topography [136, 184, 7].

The L1 distance function (Manhattan distance) was utilized, as implemented in MATLAB (MathWorks), based on evidence suggesting its effectiveness as a similarity measure for high-dimensional data compared to the L2 (Euclidean) distance [7]. Covariance values between all ROIs were considered, resulting in $[82 \times (82 - 1)] / 2 = 3,321$ features per matrix (keeping only the upper triangular matrix, because the matrix is symmetric). Before clustering, the Fisher transformed $Z_{r,w}$ matrices were subsampled along the time dimension w , similarly to EEG microstate analysis, to reduce redundancy between windows and computational demands [184, 7]. Sub-sampling involved selecting connectivity matrices ($Z_{examples_r}$) corresponding to windows exhibiting local maxima in functional connectivity variance, where the absolute normalized variance exceeded 0.5 Standard Deviation. The clustering algorithm was applied to $Z_{examples_r}$ and repeated 500 times with random initialization of centroid positions to increase the likelihood of avoiding local minima.

The resulting centroids or median clusters (termed Brain Pattern (BP) (BP_n), where $n \in [1-7]$; each sized 82×82) were then used to initialize clustering of all data (i.e., not just examples but entire $Z_{r,w}$ matrices) across different experimental conditions, yielding a brain pattern matrix $B_{r,w}$. For a given run r , this matrix comprises a vector of length 464 with values ranging from 1 to 7 (predefined number of clusters), as each matrix in $Z_{r,w}$ is assigned a BP_n . The number of brain patterns was predetermined as seven, following previous studies [7], with additional exploratory analyses confirming consistent and robust results over a range of k values [14, 245, 234]. In the end, we obtain a repertoire of brain patterns whose richness can be evaluated by condition, as well as their similarity to anatomy.

Similarity with anatomical connectivity

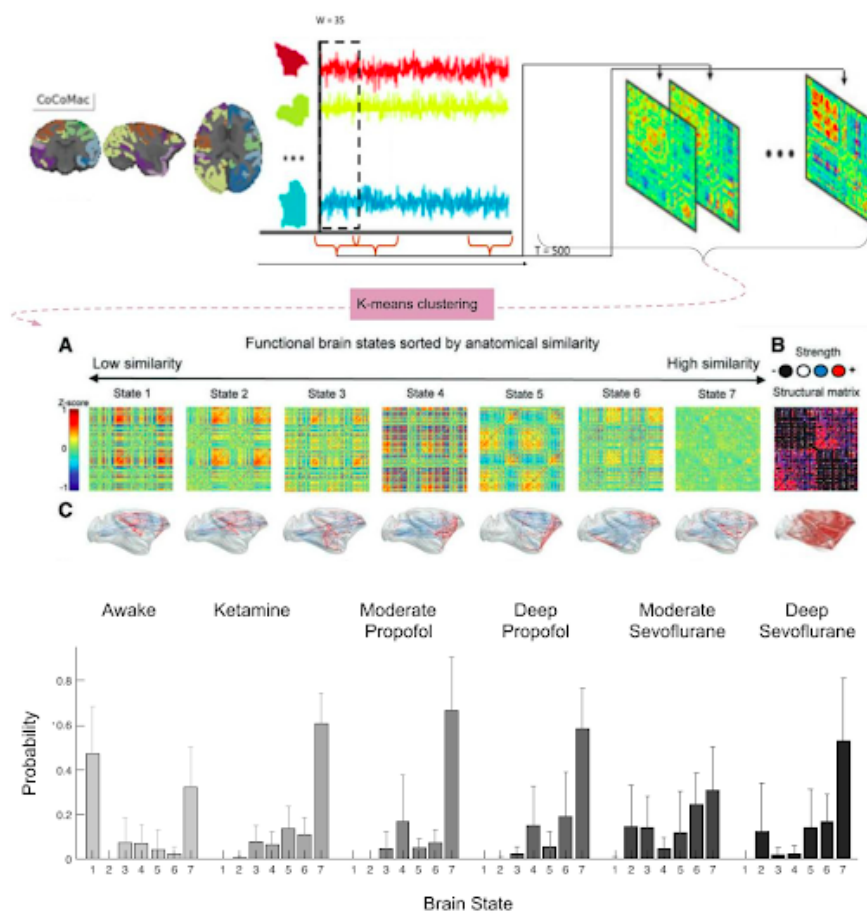
The BP_n matrices obtained were compared to the anatomical connectivity matrix derived from the CoCoMac atlas. This matrix expresses the strength of the anatomical connection between any pair of cortical areas (cf. section 2.2.3 for detailed methodology). To investigate the relationship between brain dynamics and arousal levels, a metric of similarity between anatomical connectivity and FC was established to categorize all brain patterns along this dimension. This similarity measure was computed by calculating the correlation coefficient between the vectorized structural matrix (with dimensions $3,321 \times 1$) and each vectorized brain pattern or centroid derived from the clustering analysis. Utilizing this similarity score, all brain patterns were arranged in ascending order based on their resemblance to the structural connectivity [14, 245, 234].

4.1.2 . Application to consciousness : state-of-the-art

The methodology developed above provides representatives of brain dynamics. These are a valuable means of discovering biomarkers linked to the state of alertness. Several signatures of consciousness have been obtained through numerous

studies, in humans and NHP, summarized below [7, 14, 245, 60, 234, 37] (cf. Fig 4.1):

- wakefulness is associated with a rich repertoire of states, where flexibility between states is high. Anatomy-function similarity is low, i.e. the states most represented in wakefulness are not particularly close to the anatomical matrix.
- loss of consciousness is associated with a poor repertoire of states and reduced flexibility. Anatomy-function similarity is high, and the most-visited state is systematically the one closest to anatomy.



State 1 only exists in the awake state

State 7 is predominant under all anesthetics

Figure 4.1: Representation of the dFC analysis framework and its main results concerning consciousness. Adapted from [245]

Transitioning from wakefulness to anesthesia, the underlying anatomical connections between brain regions emerge as the principal guide of the repertoire of

functional states. Subjects under anesthesia lose the capacity to produce adaptable functional brain patterns that transcend brain anatomy [245].

Allen et al (2014) [7] are the first to propose this approach in awake humans. They note that one brain pattern is systematically more visited at the end of acquisition. They hypothesized that this was linked to the participants' level of arousal: participants dozed off or even fell asleep during acquisition. This hypothesis links the dynamics of cerebral states to the condition of vigilance.

This is borne out by studies carried out in the awake state and with loss and recovery of consciousness, both in the NHP and in humans.

Barttfeld et al. (2015) [14] by proposing this approach in awake and anesthetized NHPs show that BPs are indeed consciousness markers. In the awake state, brain patterns exhibited similar probabilities of occurrence, independently of the resemblance of functional networks to structural connectivity. However, under sedation, the probability distribution of brain patterns underwent significant reshaping: those resembling structural connectivity became more probable. Sedation altered the composition of brain patterns, with some, particularly states 1 and 2, having such low occurrence probabilities that they rarely manifested during sedation. Conversely, brain pattern 7, bearing the closest resemblance to anatomical structure, emerged as the dominant state, its occurrence probability strongly influenced by the level of vigilance. Additionally, the average duration of each brain pattern, indicative of dynamical connectivity stability, was analyzed. Sedation led to an increase in the average duration of brain patterns, even after accounting for the increased presence of specific states. This phenomenon was primarily driven by the prolonged duration of brain pattern 7, which prevails during deep sedation. Likewise, as anesthesia deepens, brain activity in rats explores fewer distinct states and undergoes fewer transitions [103].

Uhrig et al (2018) [245] add acquisitions with other anesthetics to the above study. Their results confirm that the state repertoire-related signatures of consciousness previously observed are not due to a specific anesthetic agent but to anesthesia-induced loss of consciousness globally. Regardless of the molecular mechanism involved, anesthesia triggered a profound reorganization of the repertoire of functional brain patterns, primarily influenced by brain anatomy (high function-structure similarity).

Demertzi et al (2019) [60] propose to replicate this method in patients with DoCs. Their results confirm that this pattern is not only related to anesthesia-induced loss of consciousness, but more generally to loss of consciousness.

During effective stimulation by CT-DBS (i.e. stimulation that induced arousal in an on/off manner) in anesthetized NHPs, the state repertoire reorganizes in a manner similar to the awake state, with a broad dynamic repertoire of spontaneous resting-state activity, previously described as a signature of consciousness [234].

Finally, this signature was also observed in humans under anesthesia (propofol) and during deep sleep in a recent study [37], confirming a signature robust to the

type of loss of consciousness.

It would also be interesting to propose such a study in REM sleep, where vigilance is absent but awareness is not, to determine whether this is a signature of arousal, awareness or both.

Interestingly, all these studies do not employ exactly the methodology initially proposed by *Allen et al. (2014)* [7].

In particular, Phase Synchronization [82] (also named Phase Coherence Connectivity [29] or Dynamic Functional Coordination Analysis [60]) is also used to avoid the overlapping induced by the sliding-windows method. Indeed, in the original study, functional connectivity was computed over a sliding time window. Various window widths were tested, but reducing the temporal window below a critical sample size compromised the reliability of the correlation values; shorter windows (i.e., with too few samples) led to biased estimates. On the other hand, longer time windows improved reliability but at the cost of temporal resolution [82]. To address this trade-off between temporal resolution and reliability, phase synchronization can be utilized as an instantaneous measure of dFC. The explanation of this method goes beyond the scope of this thesis, but details can be found in [29, 60, 37].

In the end, the two methods produced similar results, confirming that the loss-of-consciousness signature associated with the brain patterns repertoire is robust to a change in computation method.

4.1.3 . Overcoming the limitations of K-means clustering

The dFC analysis method, and in particular the use of the K-means algorithm in this context, has a number of limitations.

Optimal number of the brain patterns

The first drawback of this machine learning algorithm is that the number of clusters has to be chosen beforehand by the user. In previous papers, several numbers of k clusters were tested. The frequency of occurrence of each state was calculated to check that the overall trend was preserved whatever the number of states. However, from a mathematical point of view, in order to estimate the optimal number of brain patterns, several metrics can be employed. Most of them are based on the basic idea behind K-means, which consists in defining k clusters such that the within-cluster variations are minimum. We propose to define the number of clusters more objectively on the "Dynamic anesthesia" dataset, to overcome the lack of use of mathematical methods for this question [91]. Among these methods, we have retained the Elbow [123], the Silhouette [123, 206], and the Calinski and Harabasz [31] scores. The Elbow value is defined as the Within-Cluster Sum of Squared (WCSS) error for different values of k . The optimal k corresponds to the point of inflection on the WCSS versus k curve. Note that the Elbow score is more a decision rule than a metric. The Silhouette value measures how similar a sample

is to its cluster compared to others and is bounded between $[-1, 1]$. -1 corresponds to incorrect clustering and $+1$ to highly dense clustering. Scores around zero indicate overlapping clusters. The Calinski and Harabasz value is defined as the ratio between the within-cluster and the between-cluster dispersions. The last two scores are higher when clusters are dense and well separated.

From the proposed experiment, we found no clear rule to select the optimal number of brain patterns for the K-means clustering (Fig. 4.2). The Silhouette and the Calinski and Harabasz indicators would tend to select the smallest k value. This is not in line with previous studies, which generally choose a k between 4 and 7 [7, 14]. But the most important conclusion to be drawn from these experiments is that the boundary between the clusters is not clearly defined. It is likely that they form a continuum and overlap. It would be wise to explore further a latent representation of our lower-dimensional dataset to check the relevance of using K-means.

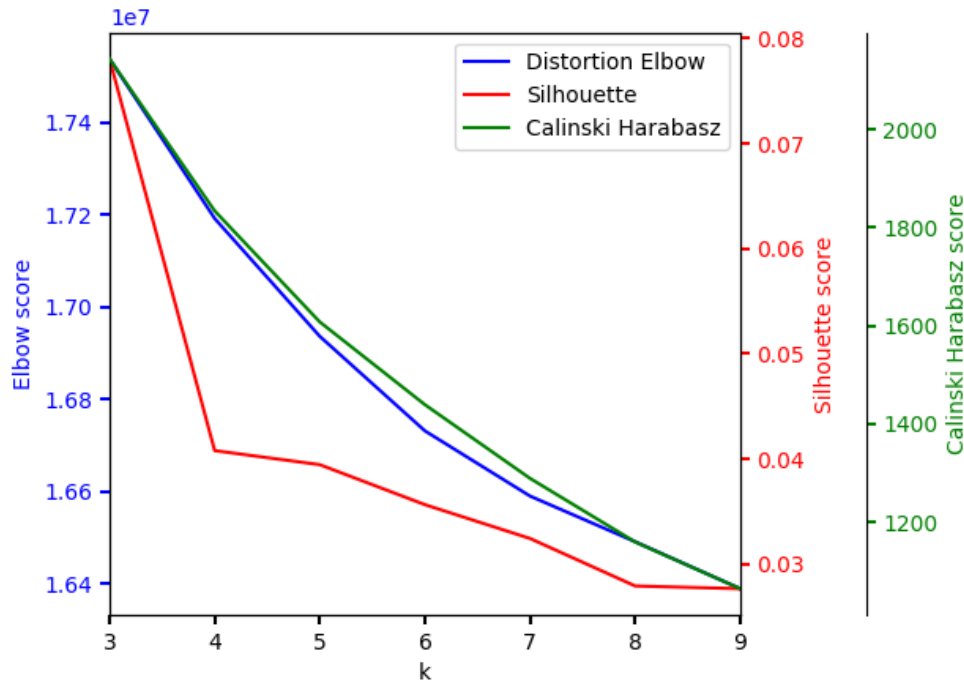


Figure 4.2: Determining the optimal number of brain patterns for the K-means clustering using the Elbow, the Silhouette, and the Calinski and Harabasz scores.

Quickly predict new states

The methodology of computing brain patterns $\{BP\}^k$ described above (cf. 4.1.1) includes the entire dataset for the extraction and annotation of $\{BP\}^k$. Although the set of $\{BP\}^k$ is robust and stays stable when a new RS-fMRI sequence is added to the database, the analysis process resumes the whole clustering, with

potentially an appropriate initialization. It would be desirable to have a faster process that does not require recomputing the $\{BP\}^k$ set to classify the $\{BP\}^k$, especially in the case of a real-time approach, which is particularly interesting for use in neurofeedback. Although the K-means algorithm offers the possibility of predicting a label on new data, clustering algorithms are descriptive analytics as opposed to predictive analytics. By itself, clustering is not really intended to forecast. For that, it is better to use a classifier algorithm.

Fighting the curse of dimensionality

Finally, in very high-dimensional spaces, euclidean and manhattan distances tend to become inflated (this is an instance of the so-called “curse of dimensionality”). It is more reasonable to consider some compact and representative features instead of the whole feature space. Running a dimensionality reduction algorithm such as PCA prior to K-means clustering can alleviate this problem and speed up the computations [187]. But PCA is linear and doesn't capture non-linear manifolds. To overcome this problem, an emerging branch of the literature proposes to learn a latent representation of the data using deep learning.

Deep representation learning

Deep representation learning (also known as self-supervised learning) allows us to move away from feature space. By using representation learning, you don't need the label to learn a representation of the data. The network learns like in supervised learning but the labels are got without human intervention, in an automatic way. It is not the same than unsupervised learning, such as K-means algorithm, because in self-supervised learning, the labels are learned and then used like in supervised manner. In unsupervised learning there is no label, nor correct output. This means that learning is not constrained. In our case, our labels come from a machine learning algorithm, so they have no neuroscientific, nor clinic reality. We'd like the representation we learn to be influenced as little as possible by these labels. We therefore try to learn an unbiased representation, to which we add a label-based classification task. The label-dFC association is no longer based solely on the original feature space, but on its latent representation. To learn a representation without label, most of self-supervised framework need a pretext task. It is a task that will allow to learn some features and representations to be used in the principal task. For example, predicting rotations is a famous pretext task : the network need to predict what sort of rotation was applied to the image (4 class classification). Why it works ? To predict well the rotation, the model needs to learn the representation of the image.

Deep contrastive learning

Within this branch of literature, contrastive methods are in vogue [35, 38, 256]. In SwAV [35], prototypes/clusters are discovered while enforcing consistency be-

tween cluster assignments of contrasted samples. Generic contrastive methods can also tackle the downstream clustering problem by applying a linear classifier on top of the fixed learned representation. For example, in SimCLR [38], input-distortion invariance is used as a pretext task to learn an appropriate representation of the data. Specifically, it encourages two augmented samples to be close in the representation space. An alternative option, computationally efficient and stable, has been proposed in Barlow Twins [256]. The learned representation is discovered by estimating the empirical cross-correlation matrix from contrasted samples. From a classification point of view, no constraint on class collision is enforced, thus allowing several samples sharing similar semantic content to be pulled apart. However, in practice, this problem is limited, and Barlow Twins produces clustering results as good or better than SwAV and SimCLR on ImageNet, Places-205, VOC07, and iNat18 [256].

The rest of this chapter presents the method and results of the preprint "Predicting Cortical Signatures of Consciousness using Dynamic Functional Connectivity Graph-Convolutional Neural Networks" (bioRxiv 2022) [91]. We propose to use a self-supervised contrastive machine learning method based on artificial neural networks to predict functional brain patterns across levels of consciousness from RS-fMRI.

First, we propose to use a non-linear contrastive model adapted to connectivity matrices. Indeed, the circularity between the targets composed of pseudo-labels derived from the K-means and the predicted probabilities issue is problematic, and without particular care, the weights of the model may more reflect the clustering algorithm than the variability of the data. For this purpose, recent contrastive learning techniques are implemented, more specifically, the Barlow Twins strategy [256]. We then classify the resulting representations and compare the results with those of a simple linear classifier. To check the robustness of these results to a different number of brain patterns, we train these networks for different values of k .

We will initially validate the utility of $\{BP\}^k$ as descriptors for characterizing the sequence of states occurring within a specific arousal condition. Our approach involves training a classifier capable of predicting individual brain patterns from short RS-fMRI temporal windows in entirely new samples/runs [124]. Specifically, we will employ a graph-Convolutional Neural Network (gCNN) classifier known as BrainNetCNN [117].

Secondly, we will challenge the data and learning process with capturing the dynamic transitions between different brain patterns.

Finally, we will utilize the inherent automatic differentiation capabilities of the deep learning framework to identify the connections that contribute to the classification of brain patterns. These connection maps, which complement the $\{BP\}^k$, offer insights into both brain patterns and the cortical correlates of conscious-

ness. They can be viewed, to some extent, as proxies for cortical signatures of consciousness.

4.2 . Material and Method

4.2.1 . Dataset and data partitioning

Here, we work with "Dynamic Anesthesia" only (see Table 2.3). Our approach involves reserving one monkey as the test set (35 runs) and using the remaining four monkeys as the training set (121 runs). Additionally, we ensured that both the training and test sets maintain consistent proportions of arousal conditions and brain pattern labels. The test set was kept separate and not utilized during the training phase. Furthermore, the training set was divided into validation and training folds through a 3-folds stratified cross-validation method.

4.2.2 . Model architecture

Traditional CNNs lack the capability to capture the spatial relationships among the ROIs used in constructing the FC matrices. BrainNetCNN incorporates specialized convolutional filters for edge-to-edge, edge-to-node, and node-to-graph connections, ensuring a more accurate representation of the topological proximity among the ROIs.

The BrainNetCNN [117] works specifically with network data and will enforce the FC patterns. The model architecture is composed of convolutional layers followed by fully connected layers. Among the proposed configurations, we selected the E2Enet-sml, which consists in removing one edge-to-edge layer and two of the fully connected layers (Fig. 4.3). This configuration has shown excellent performances with a restricted number of parameters. Precisely, E2Enet-sml has an edge-to-edge layer composed of $32 \ 1 \times 82$ and $32 \ 82 \times 1$ filters producing feature maps of size $32 \times 82 \times 82$, followed by an edge-to-node layer with $64 \ 1 \times 82 \times 32$ filters yielding feature maps of size $64 \times 82 \times 1$, a node-to-graph layer with feature maps of size $1 \times 1 \times 30$, and a fully connected layer with an output of size k . Increasing the number of feature maps with each layer is a common strategy for CNNs to compensate for the reductions along the other dimensions. Every layer uses very leaky rectified linear units as an activation function with a negative slope of $1/3$.

We implement the Barlow Twins approach [256], wherein the fully connected layer of BrainNetCNN is substituted with a projection head comprising two hidden layers, resulting in an embedding space of dimensionality 120. The objective of the model is to minimize the discrepancy between the empirical cross-correlation matrix, derived from twin embeddings (i.e., outputs of the network fed with augmented or distorted versions of a sample), and the identity matrix (Fig. 4.3). To induce distortion in the FC matrices, a random connection erasing scheme is employed,

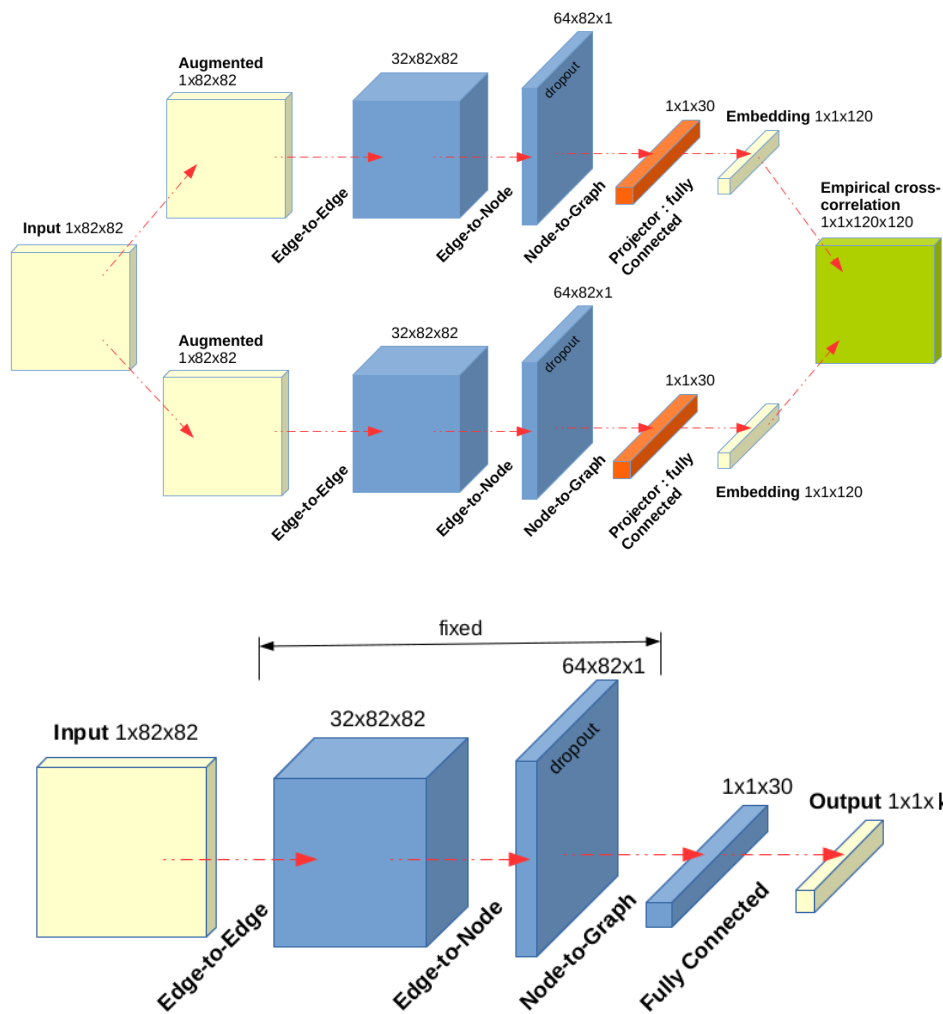


Figure 4.3: Schematic representation of the proposed training. The BrainNetCNN E2Enet-sml layers are represented in blue. In the first row the Barlow Twin contrastive approach is described. The fully connected layer of the BrainNetCNN is replaced by a projection head (in orange), and empirical cross-correlation matrix (in green) is computed from augmented/distorted samples (in yellow). In the second row the connected layer of the BrainNetCNN is trained using the K-means pseudo-labels by fixing the other weights.

wherein the proportion of zeros is randomly selected within the interval [0.3, 0.7]. Pseudo-labels obtained from K-means clustering are solely utilized for training a linear classifier on the fixed representations learned by the Barlow Twins. All other parameters are frozen to mitigate circularity between the targets and the predicted probabilities.

For training, we employed a dropout of 0.5 before the node-to-graph layer as shown in Fig. 4.3, and we followed the optimization protocol described in [256]. We use a LARS optimizer, a mini-batch of size 1024, a weight decay of $1e-6$, a learning rate of 0.2 for the weights and 0.0048 for the biases and batch normalization parameters, and train the model for 1000 epochs. We employ a learning rate warm-up period spanning 10 epochs, following which the learning rate is decreased by a factor of 1000 using a cosine decay schedule. The biases and batch normalization parameters are excluded from LARS adaptation and weight decay. The linear classifier is trained based on the fixed representation, optimizing the cross-entropy loss over 100 epochs with an SGD optimizer, a weight decay of $1e-6$, a learning rate set to 0.01, and a cosine annealing schedule.

4.2.3 . Maps of Predictive Connections

Saliency maps serve as a popular visualization tool for understanding the rationale behind decisions made by deep learning models, such as images or FC matrices classification [227, 230]. Through a single backpropagation, gradients of the target class concerning the input FC matrix are computed from the initial convolution layer. These gradients are then utilized to generate an FC-specific class saliency map. FC matrices that are poorly predicted are disregarded, and the resulting saliency maps are averaged on a brain-states-wise basis. These markers of consciousness are depicted using a circular graph layout. The 41 cortical regions within each hemisphere are grouped, and the top $p = 15\%$ largest positive and negative connections are showcased. This threshold is calculated at the brain-state level or across all brain patterns.

4.3 . Results on Anesthesia dataset

4.3.1 . Performance of the classifier

Given that no optimal number of brain patterns is readily apparent, we opted to evaluate the prediction accuracy of BrainNetCNN across various values of k (or targeted brain patterns $\{BP\}^k$), as illustrated in Fig. 4.4. In a k -classes classification scenario, the theoretical chance level is $1/k$. This threshold is true for an infinite number of samples, and the smaller the sample size, the more likely it is for chance performance to deviate from this theoretical chance level [46]. In our context, with a large sample size (56,144 samples), we assume that the chance level closely approximates its theoretical value. Across all k values, we observed

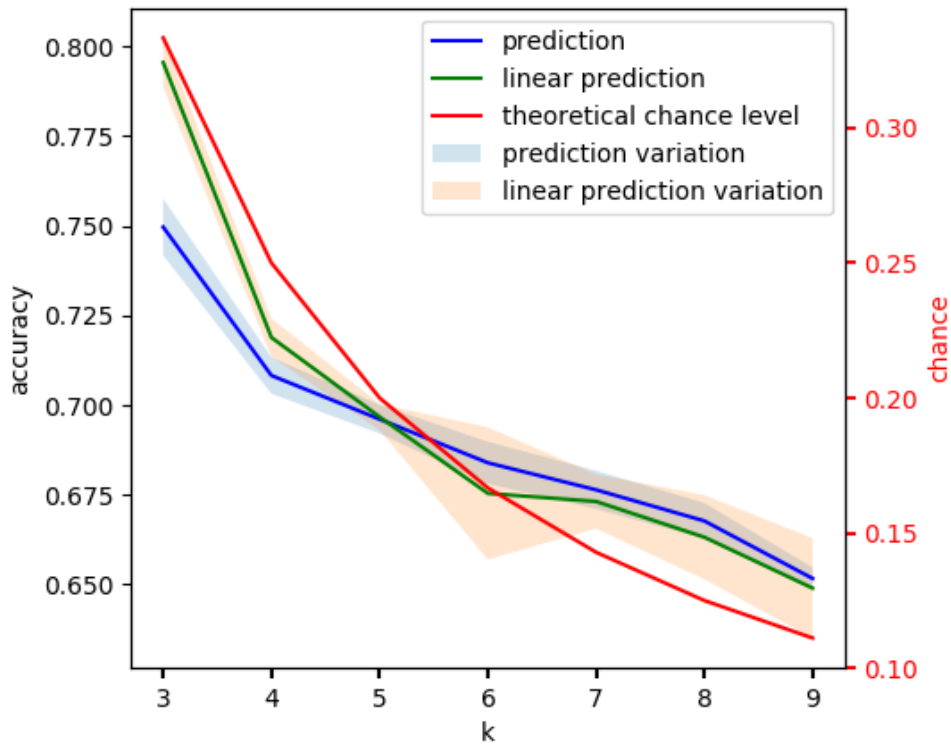


Figure 4.4: Determining the optimal number of brain patterns for the BrainNetCNN prediction using the accuracy as a reference metric: the mean and standard deviation computed across the training folds are displayed. A linear SVC prediction, as well as the theoretical chance levels are displayed as baselines. Note that results are always much higher than random chance.

that the accuracy significantly exceeded the theoretical chance level. The overall predictions fall within the range of $[0.655, 0.759]$, which is deemed satisfactory. To contextualize these findings, we compared them against those obtained by training a linear Support Vector Classifier (SVC) directly on the input upper triangular FC data (Fig. 4.4). Both methods yielded similar results, albeit with a lower variability observed in the deep learning prediction as k increased. While this outcome may initially seem disappointing, it sheds light on the simplicity of the downstream classification task at hand. Specifically, the annotations are derived from K-means clustering, which inherently involves linear decision boundaries. In contrast, although the proposed network employs a cascade of nonlinear processing units for feature extraction, the performance of the linear clustering task is likely constrained by the empirical nature of these annotations.

To delve deeper into the analysis, we employed a Uniform Manifold Approximation and Projection (UMAP) to visualize the feature spaces derived from both machine learning and deep learning approaches [159]. These spaces have dimensions of 3403 (representing the upper triangular elements of the FC matrices) and

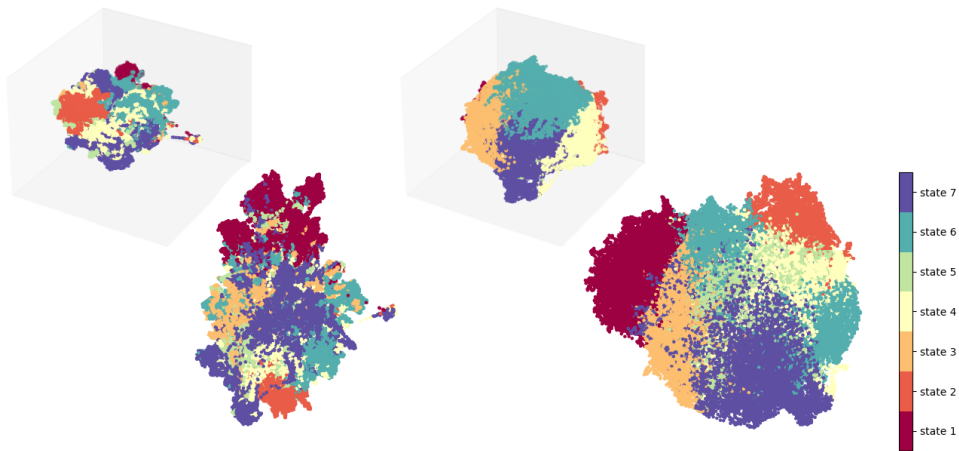


Figure 4.5: UMAP 2-d and 3-d projections of the machine learning (left) and deep learning (right) feature spaces: the upper triangular elements of the FC matrices (3403) vs the latent space dimension of the BrainNetCNN (30). The samples are colored using the $k=7$ brain pattern labels.

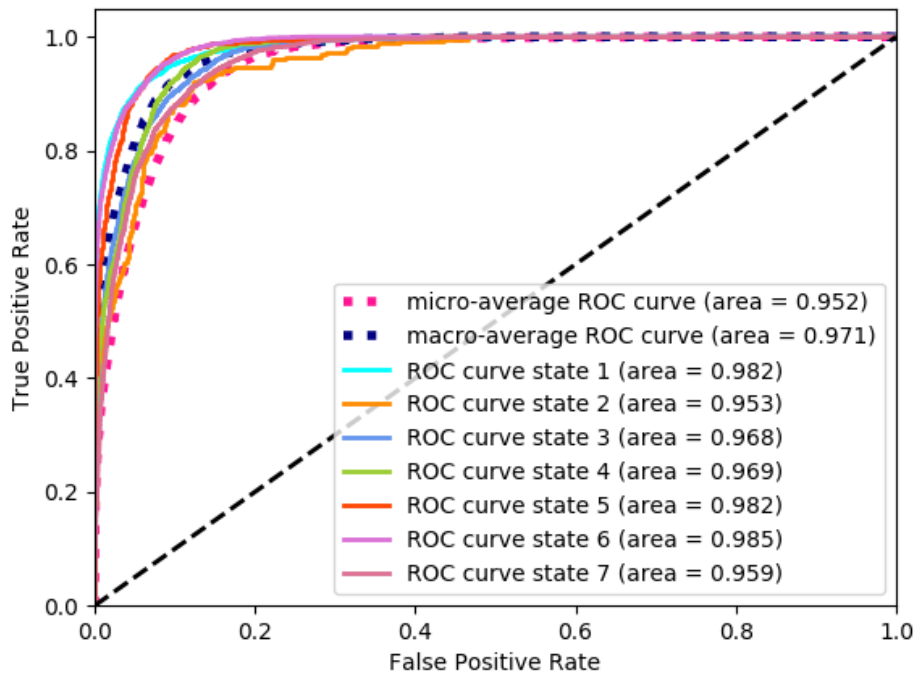


Figure 4.6: Evaluation of the BrainNetCNN classifier outputs quality using ROC curves for $k = 7$.

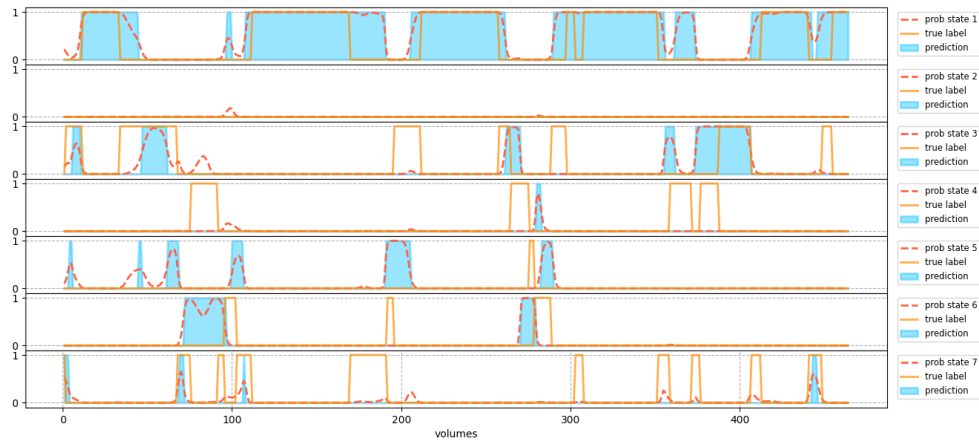


Figure 4.7: *The BrainNetCNN predicted transitions with the associated probabilities for a test set run acquired in the awake condition for $k = 7$.*

30 (representing the latent space dimension of the BrainNetCNN), respectively. We posit that the low-dimensional features learned by the BrainNetCNN encapsulate all pertinent information, facilitating interpretation, as depicted in Fig. 4.5.

For the rest of this chapter, we maintained the same number of brain patterns as in the previous works done in our team [14, 245], namely $k=7$. We opted for accuracy as the primary metric to determine the best model. Then, we assessed the quality of the BrainNetCNN classifier output using the Receiver Operating Characteristic (ROC) metric. ROC curves plot the true positive rate on the y-axis against the false positive rate on the x-axis. An Area Under the Curve (AUC) value of 1 signifies an ideal scenario where the false positive rate is zero, and the true positive rate is one.

In the context of multi-label classification, the predicted brain pattern outputs need to be binarized. Initially, a separate ROC curve is generated for each label, followed by micro and macro-averaging. Through this approach, we illustrate that the BrainNetCNN consistently predicts brain patterns with high reproducibility ($AUC > 0.92$), as depicted in Fig. 4.6.

4.3.2 . Towards modeling the brain patterns dynamic

Dwell time, i.e. the time spent continuously in a state, is an important feature when studying brain patterns. Here, as the classification labels are the same as those used in previous studies, it's not really relevant to recalculate it. However, unlike K-means clustering, which forces the association of a dFC with a pseudo-label (hard clustering), the classifier gives a probability of belonging or not to the class in question. However, the dFCs for transitions between two states reflect both states at the same time. We hypothesize that if this transition is well encoded in

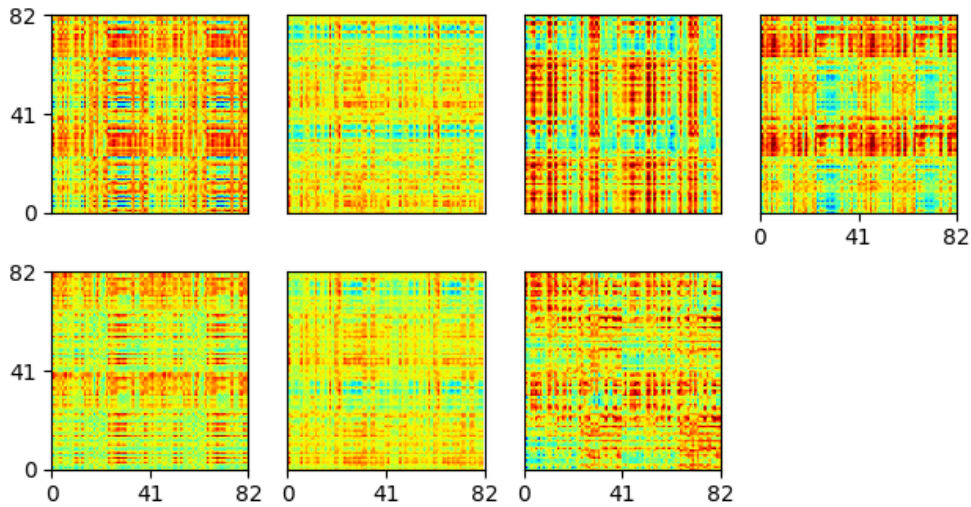


Figure 4.8: Resulting mean saliency maps for each brain pattern.

latent space, classification should be worse at transitions than in the rest of the run.

By replaying the dynamic FC matrices movie for each run, we can directly access the transition probabilities between different states as the output of the trained network. Fig. 4.7 showcases both the predicted and true labels, along with the network-estimated probabilities for a specific run obtained under the awake condition. Notably, the network's decisions vary at state transitions. These chronograms provide insight into the monitored brain configurations and, prospectively, model the brain's dynamic oscillations from one state to another.

4.3.3 . Maps of predictive connections

We use saliency maps to uncover the connections learned by the BrainNetCNN to be predictive of consciousness. They are giving some intuition on connections within the input that contribute the most and least to the corresponding prediction. In our application, they are helpful to extract proxies of the cortical signature of consciousness.

Proxies of the cortical signatures of consciousness are computed as described in section 4.2.3. The resulting mean saliency maps for each brain pattern are displayed in Fig. 4.8. The top $p = 15\%$ positive and negative connections using a specific or global threshold across brain patterns are presented in Fig. 4.9. The Co-CoMac regions are grouped into seven locations comprising the cingulate, frontal, gustatory, insular, occipital, temporal, and parietal cortex.

4.4 . Conclusion

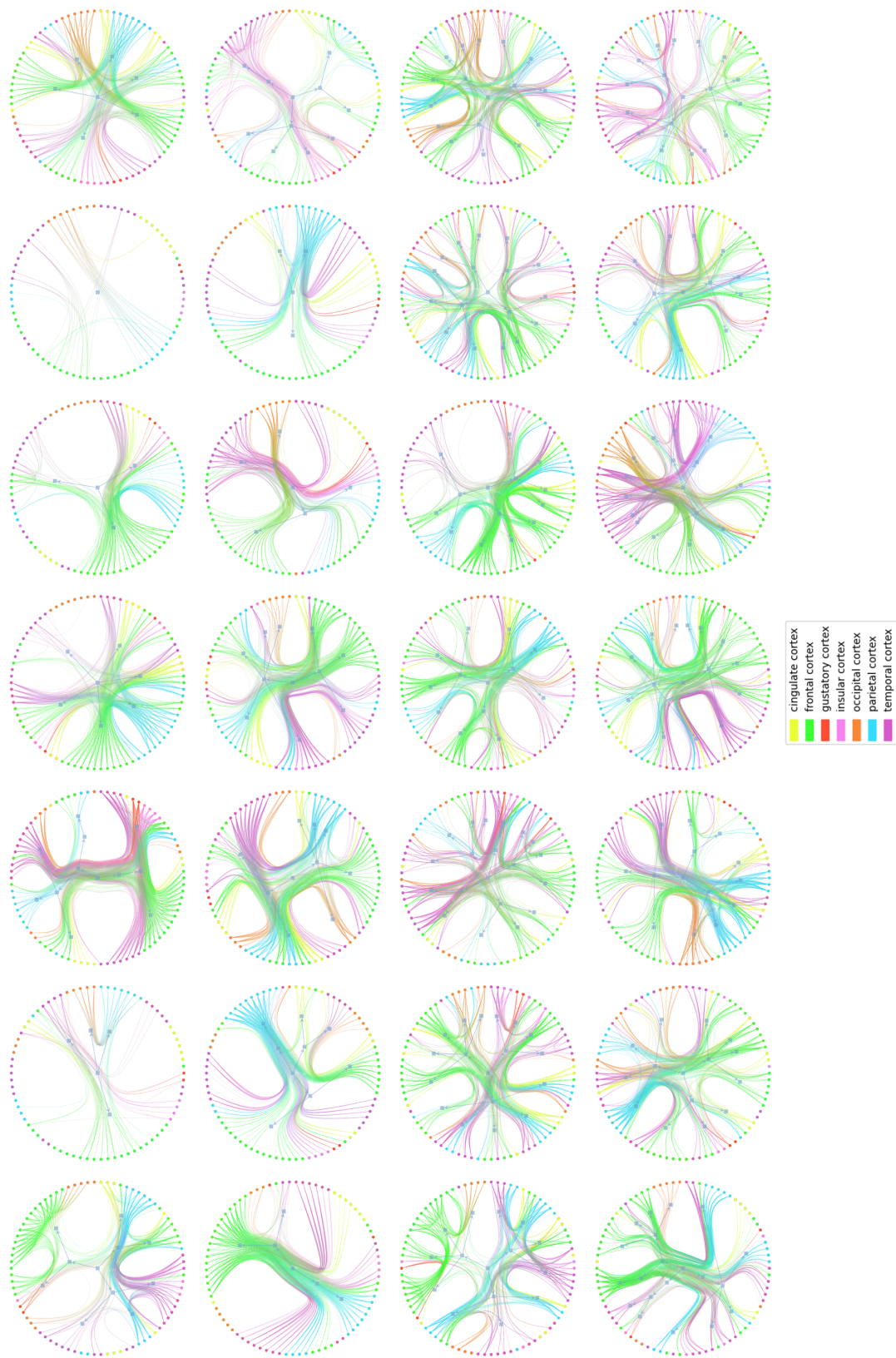


Figure 4.9: Within-brain patterns marker of consciousness for $k=7$, rendered using a circular graph where the top $p = 15\%$ positive and negative connections using a specific or global threshold across brain patterns are presented. Top to bottom rows contain the global-positive, global-negative, specific-positive, and specific-negative results, respectively.

The BrainNetCNN gCNN model demonstrates excellent reproducibility and accuracy in predicting brain patterns. Its performance closely matches that of a linear SVC when applied to the input dynamical FC data, highlighting the simplicity or constraint imposed by the downstream classification task driven by the K-means pseudo-labels. However, the learned latent space has the capability to handle more intricate tasks by acquiring complex representations. Employing a self-supervised contrastive learning strategy helps mitigate circularity associated with the pseudo-labels. Interestingly, the predictions made by BrainNetCNN diverge from those of K-means during brain pattern transitions. Beyond serving as a mere prediction tool, the proposed network has the ability to model the dynamic oscillations of the brain as it transitions between states, generating state signatures represented by sets of predominant connections. By mapping out the most influential connections in predicting a specific brain pattern, it becomes possible to discern which connections are crucial for discriminating between different levels of wakefulness, thus providing valuable insights into brain patterns. These maps are expected to aid in understanding the signature of consciousness within different brain patterns.

4.5 . Discussion

Caution about fluctuations driving by arousal

This study, along with previous research (see section 4.1.2), suggests that fluctuations in arousal could account for a significant portion of the variability in dFC. This could present challenges when comparing individuals or groups with varying levels of drowsiness (e.g., Parkinson's disease; [122]), underscoring the importance of integrating sleep assessments and measures of arousal in studies examining both static FC and dFC [153].

Augmentation data strategies

The framework might encounter limitations due to its coarse augmentation strategy. Alternative methods utilizing data-based generative models such as GANs have been proposed [139, 13], along with approaches based on dynamical systems [192] (see Section 2.3.1 for details).

Modeling states and transitions

The classifier provides finer distinctions regarding transitions between states, thanks to the prediction probabilities. It remains to be determined whether what the classifier perceives is due to actual transition states in the data or to our method of calculating dFC (redundancy induced by sliding windows). It would be interesting to replicate this classification using dFC calculated with Phase Synchronization, for example.

Limitations of K-means labeling and perspectives

Numerous empirical studies of dFC also aim to assess transient "brain patterns" and their transitions. In this framework, each state characterizes a distinct pattern of whole-brain activity or functional connectivity. However, determining the "ground truth" of the fluctuating neural interactions that underlie dFC is often challenging, if not impossible (and maybe even undefined) [153].

Hard clustering Different models impose varying constraints on the estimated states, such as whether they occur in isolation (one state per time point) or in combination (a mixture of states at each time point). With the K-means algorithm, our states definition is "hard," meaning we consider that each time point exhibits a single state. Alternatively, we could adopt a "soft" configuration by using overlapping clustering, such as fuzzy K-means, where each sample belongs to two or more clusters with different degrees of membership, or probabilistic clustering, such as mixtures of Gaussians, where each cluster is represented by a parametric distribution [91, 138, 163].

Distance metric Formal model selection and comparison, such as utilizing information-theoretic criteria, enables the assessment of which models offer the most accurate description of the observed data [153]. The choice of distance functions (such as correlation, Euclidean, and cosine) appears to have little impact on the results [7]. However, preliminary work by Aurélien Stumpf Mascles during a 6-month internship at Neurospin suggested that the choice of distance could influence the results. When clustering similar connectivity matrices, it is more interesting to examine the similarity of the networks. This emphasizes our focus on the patterns within the matrices rather than their actual connectivity values, which may be noisy. Consequently, the Euclidean distance is not the optimal metric choice, as it fails to distinguish between a lightened connectivity network (with the same pattern but lighter weights) and random noise added to this network. The correlation metric proves to be a much better choice as it captures the common patterns between matrices.

In another preliminary study conducted with Olivier Cornelis, a 3-month intern at Neurospin, we conducted a benchmark of clustering methods for dFC analysis using synthetic and anesthesia datasets. We assessed the sensitivity of clustering methods to noise and geometric space. Specifically, we aimed to evaluate methods where the algorithm considers outliers (such as DBSCAN and OPTICS) and hierarchical clustering based on the fusion of nearest clusters starting from clusters containing only one sample.

Temporal ordering disregarded The dFC analysis pipeline utilized in our study incorporates stages that alternatively consider and disregard temporal ordering. Initially, we estimate sliding-window correlations, calculated using time series

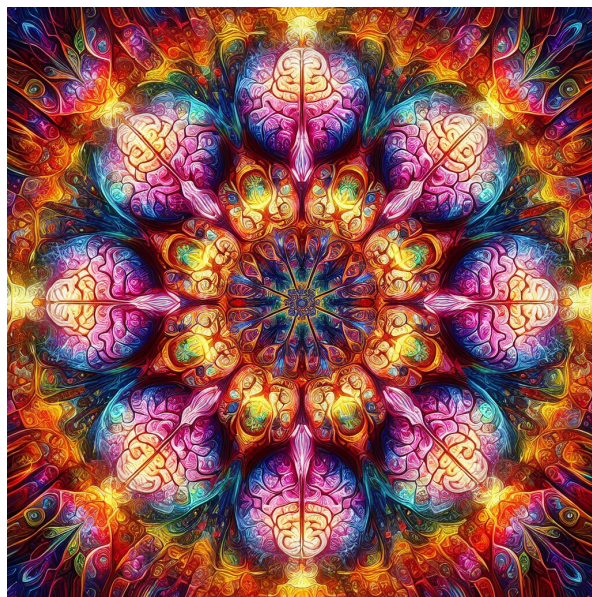
with time points ordered as observed. Subsequently, we apply K-means clustering to the resulting dFC matrices, where K-means ignores the temporal ordering of the windows. Finally, we evaluate state properties such as dwell times and transition probabilities, which again take into account the temporal order of time points. This is a common issue found in several studies [7, 14, 245].

Deep self-supervised clustering Recent studies propose leveraging self-supervised deep learning techniques to unveil semantically relevant groups of samples [34, 257]. Notably, Deep Cluster [34] employs a process that alternates between a clustering phase and a back-propagation through a classification encoder driven by pseudo-labels. Specifically, during each epoch, previous clustering assignments generated by K-means on the embeddings are utilized as pseudo-labels for minimizing the cross-entropy. These self-supervised methods have demonstrated the ability to generate semantically valid clusters without the need for manual annotation. However, training such models presents challenges. The transition between pseudo-label generation and network training introduces clustering degeneracy and inconsistency. While degeneracy issues are often addressed with heuristics, such as reassigning out-of-bounds samples/clusters, inconsistency significantly disrupts training. For example, in Deep Cluster [34], the clustering head is re-initialized at each epoch due to the label shuffling inherent in K-means. This training approach is also prone to local minima due to its circular entanglement with the pseudo-labels [91].

5 - Unravelling brain dynamics: deciphering brain states

Contents

5.1	Introduction	130
5.2	Material and method	131
5.2.1	Dataset	131
5.2.2	Low-dimensional generative models	131
5.2.3	Model evaluation	134
5.2.4	Latent space exploration	135
5.2.5	Connection-wise simulations	136
5.3	Results on Anesthesia dataset	138
5.3.1	Model evaluation	138
5.3.2	Latent space exploration	140
5.3.3	Connection-wise simulations	142
5.4	Discussion	144



5.1 . Introduction

"The stream of our consciousness, [. . .] like a bird's life, seems to be made of an alternation of flights and perchings", said the philosopher William James [109]. A fundamental observation, that still puzzles many scientists. Like the seasons that transform our landscapes, we have seen that the spontaneous fluctuations of the brain reveal very different brain configurations. Brain activity at rest is commonly characterized by the spontaneous fluctuations of regional brain fMRI signals and can be studied with dFCs. As discussed earlier, temporal analysis of the dFCs shows that wakefulness and loss of consciousness exhibit a reorganizing repertoire of brain patterns. The conscious brain is the site of rapidly changing dynamics, within a rich repertoire of brain patterns. Conversely, during anesthesia and disorders of consciousness, brain activity is expressed according to a more rigid and poorer repertoire of brain patterns (i.e., transitions between brain patterns are rare, and some brain patterns are almost never visited). In this case, the brain dynamic connectivity is reduced to the underlying anatomical connectivity [14, 60, 245].

The representation and interpretation of brain patterns is still an area of ongoing research. Previous dynamic studies have examined the frequency of occurrences or stability of each brain pattern [7, 14] and have proposed to project fMRI data into two- or three-dimensional space [191]. However, they either do not take into account the spatiotemporal nature of the data or focus on task fMRI rather than rs-fMRI [165, 79]. Consequently, it may be interesting to explore such a low-dimensional space to model a fine-grained representation of brain patterns.

Some works in the literature support the choice of a low-dimensional model to study brain dynamics. For example, dFCs have been shown to reflect the interplay of a small number of latent processes using clustering or PCA-based reduction techniques [7, 168] and latent linear models can also be used to estimate these underlying processes [48]. However, linear models may be inadequate if the mapping is nonlinear or, equivalently, if the learned manifold is curved.

The emergence of deep learning-based generative models has spread to many disciplines, including medicine and neurosciences [118, 144, 198, 217]. By learning and capturing the underlying probability distribution of the training data, generative models are able to generate novel samples with inherent variability. Three prominent families of generative models can be identified, namely generative adversarial networks, Variational Auto Encoders (VAEs) [119], and diffusion models. In line with the literature [191, 118, 259], we will focus on VAEs in this work. The probabilistic nature of such generative models holds great promise for exploring the data structure. Unlike discriminative models, VAEs are unsupervised models that do not require a labeled dataset.

In the proposed work, the choice of architecture is supported by the seminal work of Perl and colleagues [191]. They showed that when a VAE (which parameterized both the encoder and decoder using a Multi-Layer Perceptron (MLP)) is trained with simulated whole-brain data from awake and asleep healthy volunteers,

the learned representations showed faded states of wakefulness. The choice of the optimal latent space dimension remained an open question in their work. Overall, this choice is a trade-off between compressing only essential information and preserving data reconstruction.

The rest of this chapter presents the method and results of the preprint "Deep learning models reveal the link between dynamic brain connectivity patterns and states of consciousness" [86]. We proposed a new interpretability framework, called VAE for Visualizing and Interpreting the ENcoded Trajectories (VAE-VIENT) between states of consciousness (Figure 5.1). We took advantage of a previously acquired resting-state fMRI dataset in which non-human primates were scanned under different experimental conditions: awake state and anesthesia-induced loss of consciousness using different anesthetics (propofol, sevoflurane, ketamine) [14, 245]. After presenting the considered low-dimensional generative model, we showed that a 2D VAE has a balanced performance in reconstructing dFCs and classifying brain patterns. We then proposed a discrete and continuous characterization of the latent space. Finally, we showed that this model can translate some virtual modifications or inactivations of inter-areal brain connections into a transition of consciousness.

5.2 . Material and method

5.2.1 . Dataset

Here, we work with the "Dynamic Anesthesia" dataset only (see Table 2.3).

5.2.2 . Low-dimensional generative models

The Gaussian VAE VAE training involves learning both an encoder to transform data as a distribution over the latent space and a decoder to reconstruct the original data (Fig.5.1). The training minimizes the mean squared error reconstruction term, making the encoding/decoding scheme as effective as possible. Latent space regularity is enforced during the training to avoid overfitting and to ensure continuity (two nearby points in the latent space give similar content once decoded) and completeness (a code sample from the latent space should provide relevant content once decoded). These properties are at the core of the generative process. In practice, a regularization term constrains the encoding distributions to be close to a standard normal distribution using the Kulback-Leibler (KL) divergence.

Let's consider a dataset $D = \{X^{(1)}, \dots, X^{(n)}\}$ with $n = 72,384$ dFC samples, where each sample $X^{(i)} = [x_1^{(i)}, \dots, x_d^{(i)}]$ is a vector of $d = 3321$ dimensions (the dFC upper triangular elements). An autoencoder learns an identity function in an unsupervised way as follows:

$$\tilde{X}^{(i)} \approx f_{\theta}(g_{\phi}(X^{(i)})) \quad (5.1)$$

where $g_{\phi}(\cdot)$ denotes the encoder, $f_{\theta}(\cdot)$ the decoder, and $\tilde{X}^{(i)}$ is the network reconstruction of $X^{(i)}$. The reconstruction loss, expressed as a Mean Squared

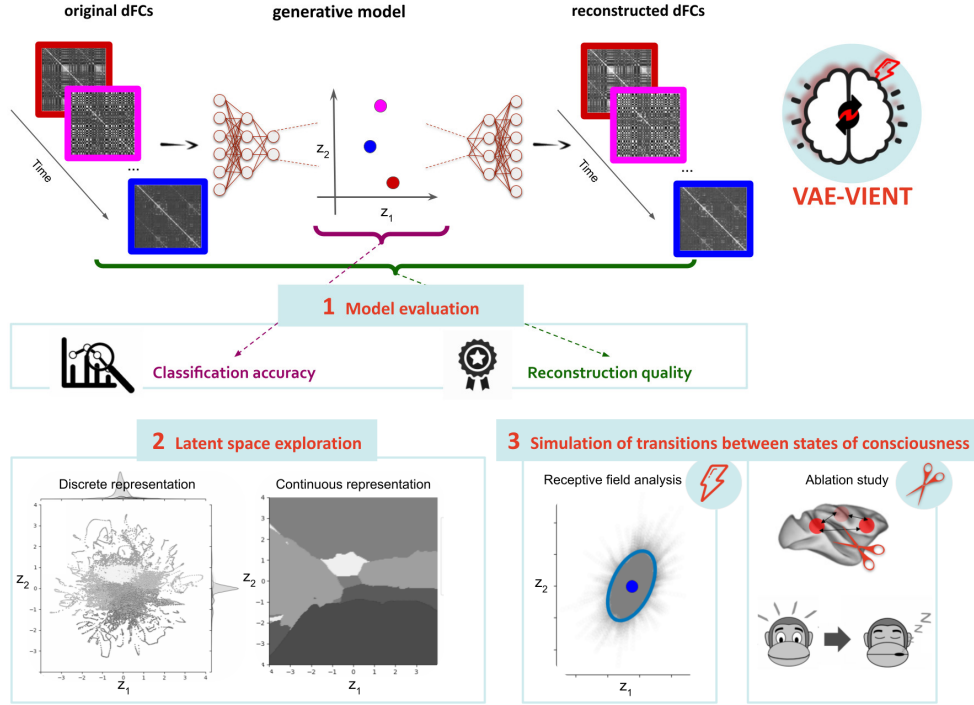


Figure 5.1: Illustration of the proposed VAE-VIENT framework. A VAE learns 2D latent representations $z = (z_{1_i}, z_{2_i})$ from dFC matrices, leading to 1) evaluation of the proposed model against other generative models implementing different latent dimensions, 2) exploration of latent space with the ability to view discrete or continuous representations (here we observe how brain patterns are organized in latent space), and 3) two simulation paradigms, including a receptive field analysis that generates tensor representations to study the effect of perturbing input dFCs, and an ablation study of Global Neuronal Workspace (GNW) connections to study the transition from wakefulness to unconsciousness.

Error (MSE), can be written as:

$$L_{MSE}(\theta, \phi) = \frac{1}{n} \sum_{i=1}^n (X^{(i)} - \tilde{X}^{(i)}) = \frac{1}{n} \sum_{i=1}^n (X^{(i)} - f_{\theta}(g_{\phi}(X^{(i)}))) \quad (5.2)$$

In this work, the VAE relationship between the input dFC data $X^{(i)}$ and the latent encoding vector $z^{(i)}$ is defined with a prior $p_{\theta}(z^{(i)}) \sim \mathcal{N}(z^{(i)}; 0, 1)$, the likelihood $p_{\theta}(X^{(i)}|z^{(i)})$, and the posterior $p_{\theta}(z^{(i)}|X^{(i)})$. Unlike (finite) Gaussian mixture models, the posterior $p_{\theta}(z^{(i)}|X^{(i)})$ is intractable. Therefore, we use a posterior approximation $q_{\phi}(z^{(i)}|X^{(i)})$ that outputs what is a likely code given an input $X^{(i)}$. It plays a similar role as $g_{\phi}(z^{(i)}|X^{(i)})$. In our case of Gaussian VAE, $q_{\phi}(z^{(i)}|X^{(i)}) = \mathcal{N}(z^{(i)}; m_{\phi}(X^{(i)}), s_{\phi}(X^{(i)}))$, where m_{ϕ} and s_{ϕ} are expressive parameterizations of the conditional mean and variance of $q_{\phi}(z^{(i)}|X^{(i)})$. The distributions returned by the encoder are further constrained to follow a standard normal

distribution as follows:

$$L_{KL}(\theta, \phi) = D_{KL}(q_{\phi}(z^{(i)}|X^{(i)})||p_{\theta}(z^{(i)})) \quad (5.3)$$

where D_{KL} is the KL divergence. To learn disentangled representations and increase interpretability, a regularization parameter β is further introduced [98, 28]. The idea is to keep the distance between the real and the estimated posterior distribution small while maximizing the probability of generating real data. A high β value emphasizes statistical independence over reconstruction. The final VAE loss is expressed as follows:

$$L_{VAE}(\theta, \phi) = L_{MSE}(\theta, \phi) - \beta L_{KL}(\theta, \phi) \quad (5.4)$$

The considered generative models In this work, we consider a VAE with a one (VAE₁), two (VAE₂) or three (VAE₃) dimensional latent space, adapting the architecture proposed in [191]. The input is the upper triangular dFCs (as each dFC is symmetric). Then, the encoder part uses two hidden fully connected layers (512 and 256 units, respectively) with ReLU activation functions, and the decoder part is implemented with the same structure. The dimension of the latent space corresponds to common neurobiological assumptions made when studying disorders of consciousness [194, 131, 59]. Furthermore, we compared our models with the sparse VAE (sVAE) [8], initialized with thirty-two latent dimensions. The sVAE implements a variational dropout to enforce parsimony and interpretability in the latent representations. A threshold on the dropout rates is used to select the optimal number of latent dimensions. We also apply a baseline machine learning model, the Probabilistic PCA (PPCA) [237], and compare the results to those obtained with VAE/sVAE. In fact, PPCA can be considered as a latent variable model. Its assumptions are Gaussian distributions and linear decomposition. The purpose of adding this model is to assess the interest in non-linear models such as VAE or sVAE when working with small datasets.

Model training We train the VAE and sVAE using an Adam optimizer, with a learning rate starting at 0.001 and a 10% decay every 30 epochs. To limit overfitting during the training, we include early stopping. The model is trained on the training set until its error on the validation set increases, at which point the optimization stops. As a performance measure to monitor the stopping of training, we consider the sliding median using a 10 epoch interval. In addition, the patience argument allows training to continue for up to 15 epochs after convergence. This gives the training process a chance to get over flat areas or find additional improvements. Using cross-validation, we study the effect of the β regularization parameter for the VAE by performing a grid search to determine the better choice for $\beta \in [0.5, 20]$ with the following user-defined steps [0.5, 1, 4, 7, 10, 20]. With the intention of building an interpretable model, we keep 8 models: PPCA₁, PPCA₂,

PPCA₃, VAE₁, VAE₂, VAE₃ with 1, 2, and 3 latent dimensions respectively, sVAE, and PPCA with the same number of latent variables selected by the sVAE. We perform a leave-one-subject-out to create an independent test set, and a training set with an internal 5-fold cross-validation. In the cross-validation, the stratification of the arousal conditions further strengthens the distribution of the classes in each training split. In the end, only the weights associated with the best validation fold are evaluated on the independent test set.

The labels and pseudo-labels used are the arousal conditions (awake and the different anesthetics) and the brain patterns (BPs) ranked in ascending order of similarity to the structural connectivity (numbered 1 to 7), respectively. Briefly, the use of seven brain patterns has been shown to be effective in representing the different configurations of the brain [14, 245]. Note that choosing the optimal number of brain patterns is challenging. It results from balancing biological assumptions and computational evaluations. These labels are known to be unevenly distributed across experimental conditions. They are also known to be good descriptors of spontaneous fluctuations in brain activity.

5.2.3 . Model evaluation

Choosing an appropriate model is a trade-off between compressing only essential information and preserving data reconstruction. Thus, we evaluate the models using two distinct metrics. The first metric is a measure of the reconstruction quality. The second one is the relative entropy of the latent space, measured by considering a classification task. Both use as labels the seven brain patterns previously described [14, 245].

Reconstruction quality From the retained trained generative models, we computed the decoded dFC matrices $\tilde{X}^{(i)}$ associated with the test set. Instead of using the MSE training loss, we evaluated the Structural SIMilarity (SSIM) between the averaged decoded dFCs and true decoded dFCs associated with each label. The MSE calculation focuses on pixel values, while the SSIM measurement focuses on and analyzes the structural differences between two dFCs. Unlike the SSIM, the MSE can be very high just because some connection values have changed. Therefore, we prefer the SSIM because we want to study global dFC patterns. This metric ranges from 0 to 1, where 1 is a perfect match.

Classification accuracy From the retained trained models, we also compute the latent representations associated with the test set. In addition to the BP labels available from the dataset, we also match each test dFC latent space location to its nearest location in the train set and retained the corresponding matched label. Balanced accuracy (BAcc) is then used to compare the dataset and matched BP labels.

Consensus metric We propose a consensus metric \mathcal{M} , which is an average between SSIM and BAcc. The goal is to enforce a trade-off that imposes spatial coherence in the latent space without significantly degrading the reconstruction quality.

5.2.4 . Latent space exploration

Discrete and continuous descriptors All the information we can transfer in latent space is associated with encoder-generated representations and is discrete by nature. To build a comprehensive whole-brain computational model, semantically continuous representations are required. Fortunately, with the generative capabilities of the VAE (or generative models in general), it is possible to decode the entire latent space. Without losing generality, let us give the formula in 2D. Let's consider a discrete grid $G \in R^2$ with $g \times g$ latent samples and the associated decoded dFCs \tilde{X}_{lm} , with $l \in [1, g]$ and $m \in [1, g]$. Using the previously known information on the brain patterns, we can label each \tilde{X}_{lm} . To this end, and as suggested by Perl and colleagues [191], we compute the similarity between each \tilde{X}_{lm} and each brain pattern. To assess the strength of these associations, we use Pearson's correlation. At the end, the label assigned to \tilde{X}_{lm} is the number of the most correlated brain pattern. The obtained continuous labeling reflects the functional reconfiguration of the brain.

Confidence level of continuous descriptors In addition, by quantifying the best association strength, we propose to derive confidence and reliability maps associated with the continuous descriptor generation process. First, the confidence map \mathcal{CM} is derived at each latent space location by taking the average of the difference between the two largest associations and the correlation between the two closest brain patterns as follows:

$$\mathcal{CM}_{lm} = \frac{1}{2} \left((\mathcal{R}(\tilde{X}_{lm}, \bar{BP}_1) - \mathcal{R}(\tilde{X}_{lm}, \bar{BP}_2)) + \mathcal{R}(\bar{BP}_1, \bar{BP}_2) \right) \quad (5.5)$$

where \mathcal{R} is the Pearson correlation, and \bar{BP}_1 and \bar{BP}_2 are the brain patterns with the first and second highest correlation, respectively. This metric takes into account both the reluctance to label and the objective nature of that reluctance. The model is reasonably confident when $\mathcal{CM}_{lm} \approx 0.5$, and overconfident when $\mathcal{CM}_{lm} \approx 1.0$. Second, the reliability map \mathcal{RM} is expressed at each latent space location by decoding the dFC and targeting the brain pattern with the highest Pearson correlation:

$$\mathcal{RM}_{lm} = \max_{k=1; k \leq 7} \mathcal{R}(\tilde{X}_{lm}, BP_k) = \mathcal{R}(\tilde{X}_{lm}, \bar{BP}_1) \quad (5.6)$$

The higher \mathcal{RM}_{lm} , the more reliable the model is.

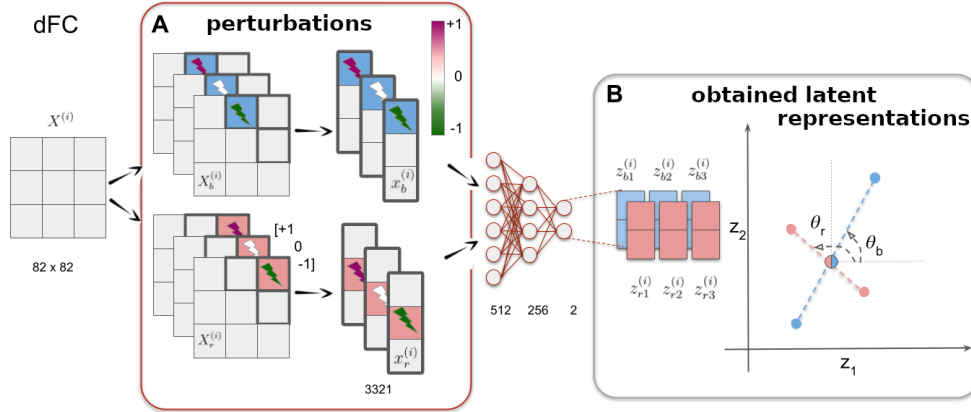


Figure 5.2: Illustration of the two steps involved in the receptive field analysis. A) From a dFC matrix $X^{(i)}$, or equivalently its upper terms $x^{(i)}$, two perturbations are performed on connections b (blue) and r (red) by swapping one connection with the following three correlations $[-1, 0, 1]$ ($p = 3$). B) Corresponding $z_b^{(i)}$ and $z_r^{(i)}$ latent representations ($N = 2$) follow lines and are summarized by their inclination angles with the x -axis θ_b and θ_r , respectively.

5.2.5 . Connection-wise simulations

The receptive field analysis If we consider brain dynamics as a physical process characterized by gradual changes in the FC space, then the dFCs used so far are samples of these changes. Physicists tried to model these processes in a principled way by analytically identifying prior knowledge about the underlying processes, e.g. by using differential equations [56]. Instead of incorporating physical knowledge into a deep neural network, we propose a Receptive Field (RF) simulation paradigm to generate a tensor model of latent space and thus gain insight into its dynamics. Such a characterization of latent space is essential for building an interpretable model. Indeed, it can help to understand encoded latent trajectories between states of consciousness. Specifically, we propose to capture the latent space RF at the connection level. In this way, the proposed RF analysis could identify the connections that need to be disrupted in order to move from one state of consciousness to another. In the long run, such an analysis may be a tool to simulate the recovery of consciousness at the individual level.

In more detail, the RF analysis focuses on the trained VAE (or another generative model) encoder. A perturbation is simulated at each connection $j \in [1, d]$ of an input dFC matrix $X^{(i)}$ (Fig. 5.2-A). The effect of this perturbation on the encoded latent representations is tracked (Fig. 5.2-B) [144]. In particular, the simulation modifies a single connection value $x_j^{(i)}$ p times, by swapping its value with a correlation drawn uniformly in an interval $[-1, 1]$, while keeping the other connections fixed. In two dimensions, the latter simulation yields p latent encoded vectors $z_j^{(i)} = \{z_{j1}^{(i)}, \dots, z_{jk}^{(i)}\} \in \mathbb{R}^{p \times 2}$. The generated latent samples $z_j^{(i)}$ are distributed around a line of varying length. This specific behavior allows an in-

interesting parameterization of each perturbation, using polar (in 2 dimensions) or spherical (in 3 dimensions) coordinates, through the inclinations $\theta_j^{(i)}$ (Fig. 5.2-B). The perturbation of all connections return a cloud of points describing the RF. The resulting cloud has an ellipsoidal shape \mathcal{E} , estimated with a confidence interval of 0.01. \mathcal{E} can be parameterized by its sorted eigenvalues λ_i and associated eigenvectors \vec{e}_i , $i \in [1, N]$, where N is the latent dimension. Finally, since each connection can be related to a direction by the inclination angle $\theta_j^{(i)}$, it is possible to select the connections with high potential for action (i.e., generating the highest brain transitions when perturbed) by identifying the directions aligned with the first eigenvector \vec{e}_1 of the ellipsoid \mathcal{E} . The procedure described above can be applied to any dFC. That is, any dFC can be projected onto a point in latent space around which an ellipsoid representing the effect of all possible unit perturbations is computed.

The ablation analysis Based on known networks involved in consciousness, we also test the ability of the trained VAE to efficiently predict state transitions. For this purpose, we simulate specific ablations of functional connections between brain areas as a virtual experiment. The resulting dFC representations are then used to predict the state of consciousness. From the Global Neuronal Workspace (GNW) theory of consciousness, we previously identified key brain areas (referred to as "macaque GNW nodes") that account for the cortical signature of consciousness realizing a fronto-parieto-cingular network [243, 244]. The key brain regions identified whose associated connections are zeroed in this study are the posterior cingulate cortex (CCp), the anterior cingulate cortex (CCa), the intraparietal cortex (PCip), the frontal eye field (FEF), the dorsolateral prefrontal cortex (PFCdl), the prefrontal polar cortex (PFCpol) and the dorsolateral premotor cortex (PMCdl) of the left and right hemispheres [244, 245]. Thus, we propose a connection-wise ablation study, equivalent to a lesion perturbation, that removes the contribution of connections linked to these fourteen GNW regions. It's important to note that removing a GNW region actually removes all connections associated with that region. By zeroing these connections, which are known to have a strong influence on consciousness, we expect to shift dFCs acquired in the awake state to an anesthetized state.

In practice, to predict the awake and anesthetized states, we train an SVM classifier on the learned latent representations. We then evaluate the performance of the classifier in predicting the awake state using the balance accuracy (BAcc). To focus on the effect of the proposed ablation and to eliminate any unrelated source of variability, the analysis is performed on the training set only. As input to the trained SVM, we take only the raw or perturbed awake dFCs (i.e., awake dFC undergoing the ablation process) encoded with the VAE. We denote the corresponding prediction scores as $BAcc$ and \tilde{BAcc} , respectively. To assess the specificity of zeroing these nodes, we test the null hypothesis that removing

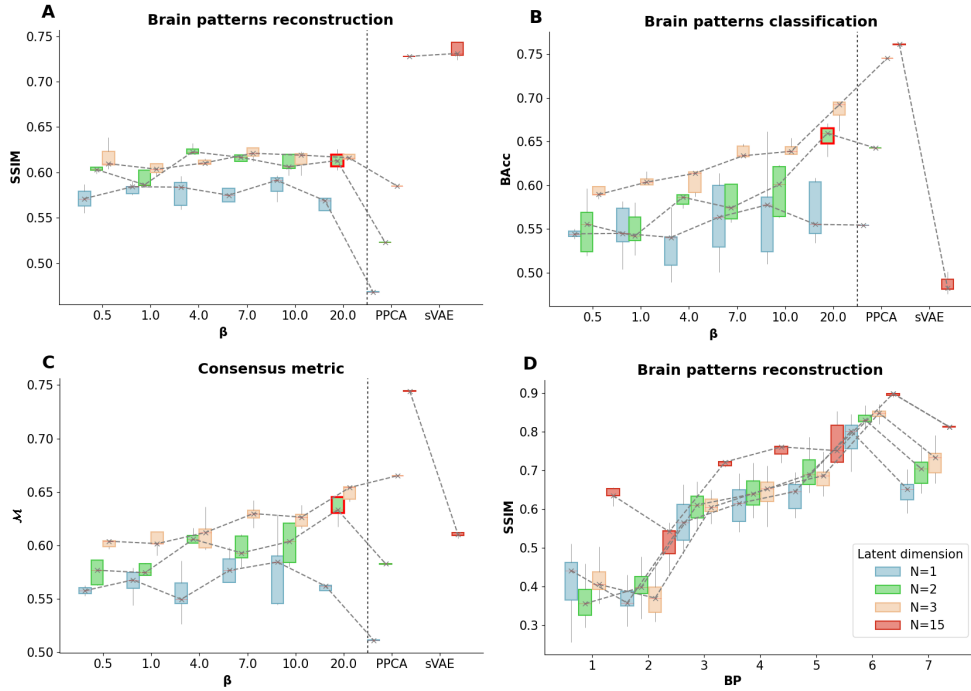


Figure 5.3: Brain pattern (BP) classification/reconstruction using VAE, PPCA and sVAE models: A) the SSIM of BP-wise averaged dFCs with respect to the model parameters, B) the balanced accuracy (BAcc) between the ground truth and the matched predicted label, C) the proposed consensus metric \mathcal{M} , and D) the SSIM recorded for each BP. In plots A, B and C, the selected VAE₂ is highlighted by a red bounding box. The dashed lines represent the trends obtained for each latent space dimension across the considered models.

random connections does not result in a significant loss of prediction compared to targeted GNW-associated connections. Let $G = 14$ be the number of GNW-associated connections. Our goal is to modify G connections that are not part of the GNW-related connections. The cardinality of the corresponding universe Ω of all possible combinations is large. Therefore, we draw a subset of $M = 1000$ samples from Ω without replacement. Finally, we evaluate the associated awake prediction performances $B\tilde{A}cc^i$, $i \in [1, M]$. Using this null distribution statistic, we compute a one-tailed empirical p-value for $B\tilde{A}cc$ by looking at the proportion of values less than or equal to the observed value when all GNW-related connections are removed [177].

5.3 . Results on Anesthesia dataset

5.3.1 . Model evaluation

In our experiments, the final number of epochs varies in the interval [159, 1359] when the early stopping criterion is applied. Note also that the variational dropout in the sVAE selects 15 of the 32 latent dimensions (see Appendix 5). To evaluate

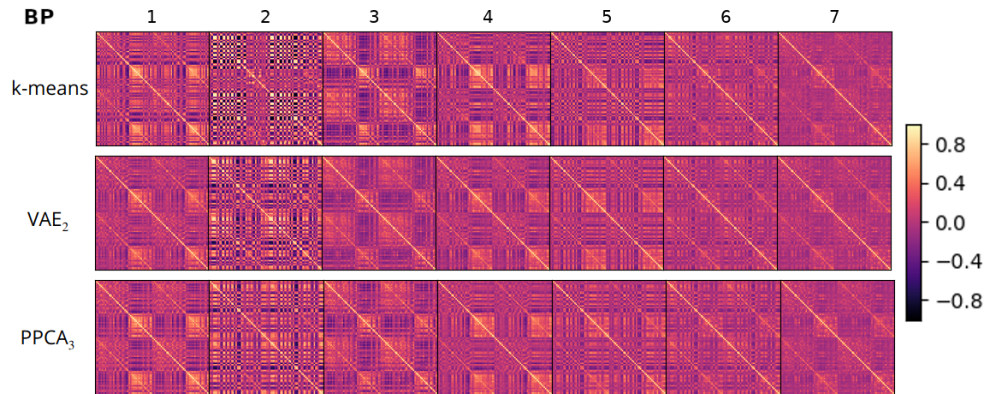


Figure 5.4: The reconstructed brain patterns (BPs) from the *k*-means clustering and the low-dimensional (2D or 3D) models maximizing the consensus metric \mathcal{M} : VAE₂ and PPCA₃.

whether a low-dimensional VAE can achieve reasonable assumptions, we compare several generative models with different parameters (PPCA₁, PPCA₂, PPCA₃, PPCA₁₅, VAE₁, VAE₂, VAE₃, and sVAE_{15/32}). This allows quantification of how the latent representations stratify the brain patterns and how a model can reconstruct the input dFCs from the low-dimensional representations.

Balancing reconstruction quality and regularization By looking at the SSIM for all models (Fig. 5.3-A), we observe that i) the chosen β has little effect on the VAE reconstructions, but ii) decreasing the latent space dimension degrades the reconstruction, and iii) nonlinear low-dimensional models have higher reconstruction quality. Higher dimensional models are expected to perform better because they capture more variability, resulting in a better reconstruction. We further quantify the decrease in reconstruction quality by comparing the VAE₂ with the sVAE_{15/32}. The cost of a low-dimensional, more interpretable model is approximately a 10% decrease in SSIM. It also appears that a nonlinear model can reconstruct better with fewer latent dimensions. Monitoring the SSIM brain pattern-wise also shows that not all brain patterns are reconstructed similarly with a SSIM in the [0.3, 0.9] range (Fig.5.3-D). Interestingly, the reconstruction quality increases with the number associated to each brain pattern. Thus, the models reconstruct more accurately the brain patterns closer to the structural connectivity with a simpler topology.

Balancing classification accuracy and regularization Monitoring the classification accuracy (Fig. 5.3-B), we observe that the BAcc i) increases as β increases, ii) decreases as the latent space dimension decreases, except in high dimensions (i.e., for the sVAE_{15/32}), and that iii) the linear PPCA baseline outperforms other models in high dimensions (PPCA₁₅). Overall, the classification

scores are relatively high for a seven-class classification problem. For all considered models, the BAcc scores range from 0.45 to 0.75 (to be compared to the theoretical chance level of 0.14). The classification accuracy metric favors the use of the highest regularization parameter ($\beta = 20$), which promotes coherence in the latent space. Furthermore, better performance (an increase of 6%) and lower interfold variance are observed for the 3D VAE (VAE₃) models. Notably, the sVAE_{15/32} performs poorly, suggesting that a few latent dimensions are preferable to encode the brain pattern information.

Balancing reconstruction and classification We have previously shown that as the number of latent dimensions increases, the model captures more variability (possibly noise). Furthermore, limiting the number of latent dimensions improves the brain pattern detection task. This trend confirms that dFCs reflect the interplay of a small number of latent processes [7, 168]. Looking at the consensus metric (Fig. 5.3-C), we specify the following model for the rest of the paper: a 2D VAE (VAE₂) with a $\beta = 20$ regularization parameter. Using these parameters, we enforce a trade-off that imposes spatial coherence in the latent space without significantly degrading the reconstruction quality. Finally, we show that the reconstructed brain patterns (as the reconstructed dFCs averaged over the different brain patterns) recover the dominant structures obtained with a k-means clustering of the dFCs (Fig. 5.4). The same model evaluation can be performed using arousal conditions as labels (see Appendix 6). To clarify the notation, the selected β_{20} -VAE₂ will be referred to as VAE in the following.

5.3.2 . Latent space exploration

To investigate the potential of latent representations to decode states of consciousness, we consider two types of descriptors: discrete and, by exploiting the generative properties of VAEs, continuous latent representations. Again, we focus on the stratification of latent representations according to brain patterns. We also consider the reliability of the generated continuous descriptors.

Stratification of brain patterns From the VAE encoder, we obtain discrete latent representations. Again, the ground truth labels are the brain patterns ranked in ascending order of similarity to the structural connectivity (numbered from 1 to 7). We examine the discrete composition of the latent space using the brain pattern labels (Fig.5.5-A) and the calculated lifetime (Fig.5.5-B). The lifetime is defined as the time spent continuously in a brain pattern (i.e. when no transition is observed). Therefore, all dFCs on this time axis have the same lifetime. Our focus is on three main properties of latent space. First, the resulting discrete representations form a cloud of points rather than a set of clearly separable clusters. Second, the generated latent representations are remarkably well stratified when looking at the brain pattern labels (Fig. 5.5-A). Each brain pattern is isolated while no constraint

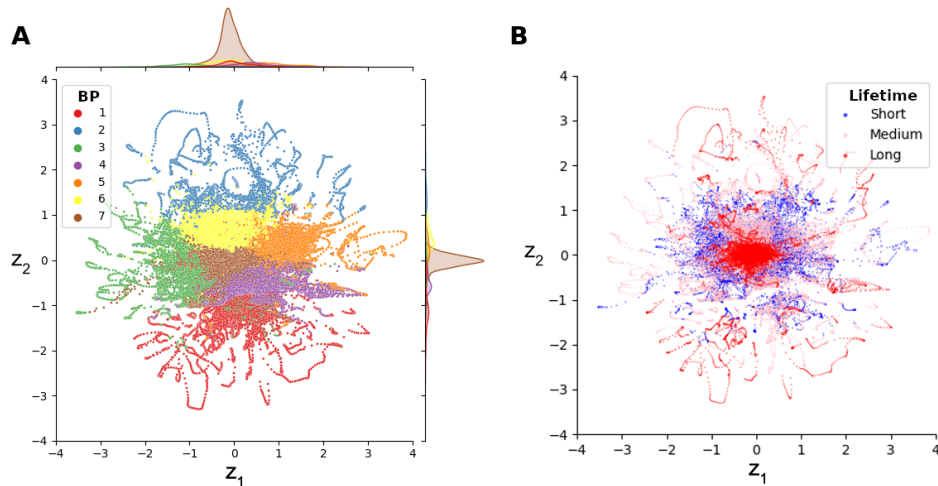


Figure 5.5: Discrete stratification of the latent space of the selected VAE into a base of A) Brain Patterns (BPs) - the centroids from a seven-class k -means clustering on the dFCs and B) lifetimes - the time spent continuously in the corresponding brain pattern. For the lifetimes, we discretize the values into three categories: the 25% longest (in red), the 25% shortest (in blue), and all others medium (in pink).

is enforced during training. To quantify the overlap between brain patterns, we choose the Dice similarity coefficient. The Dice metric yields values between 0 (no spatial overlap) and 1 (complete overlap) [64]. Overall, the average Dice metric remains relatively low ($< 0.37_{\pm 0.19}$), confirming that the spatial overlap between brain patterns is small (see Appendix 7 for details). Interestingly, brain pattern 7 (the one closest to the brain structure) occupies a central position in the representation space, and has the highest Dice coefficient. Third, the central locations, aligned with brain pattern 7, have longer lifetimes (Fig. 5.5-B). Note that we verify the absence of subject bias prior to analysis, and also illustrate the stratification of the learned latent space with respect to the acquisition conditions (see Appendix 8). We also verify that the proposed VAE reliably encodes the dFC time courses while no constraint is enforced during training (see Appendix 9).

Toward a whole-brain computational model By exploiting the generative capabilities of VAE, we obtain semantically continuous representations in the latent space, which promotes versatility. The generated continuous brain pattern labels cover the entire latent space. They also show a pooled organization of the brain patterns (i.e., each brain pattern is mostly composed of a single connected component) (Fig. 3.3-A). The accuracy of the brain pattern matching process is measured by the confidence \mathcal{CM} and the reliability \mathcal{RM} maps. Interestingly, the most striking trend is that brain pattern boundaries are less reliable than central locations (Fig. 3.3-C and D). With these maps, we gain confidence in using continuous descriptors in the latent space. Finally, decoding dFCs on a 19×19

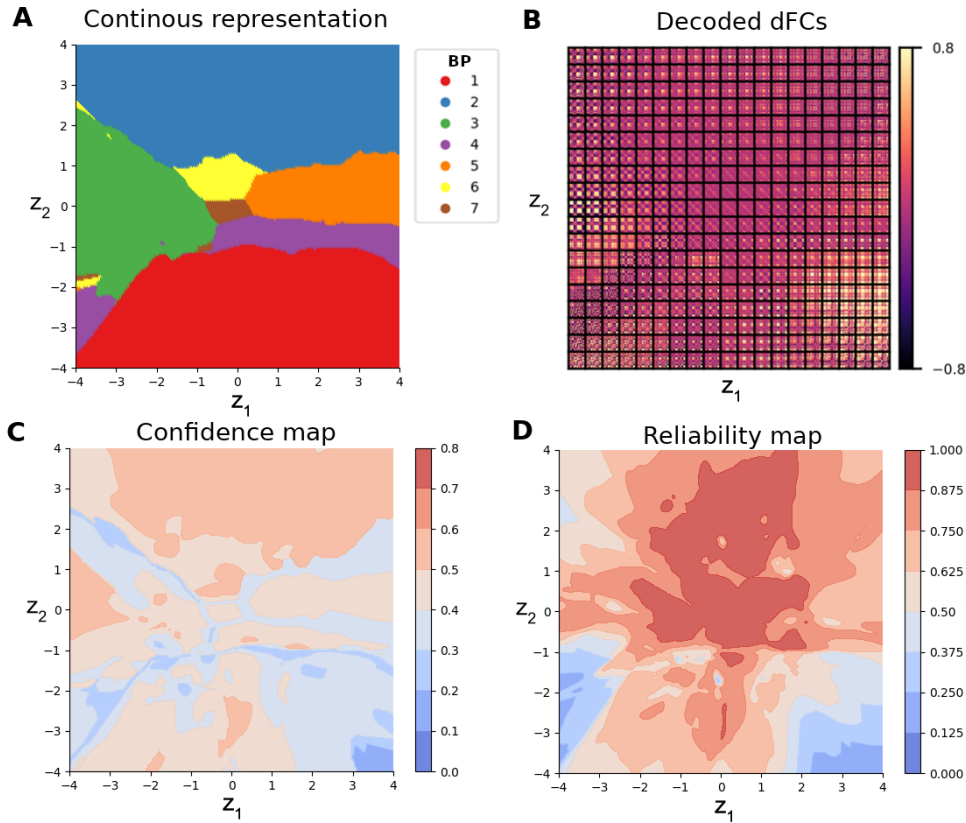


Figure 5.6: Continuous stratification of the latent space of the selected VAE and corresponding confidence and reliability maps: A) continuous representation of the Brain Patterns (BPs), B) decoded dFCs from a regularly sampled 19×19 grid in the latent space, C) estimated confidence map $\mathcal{C.M.}$, and D) estimated reliability map $\mathcal{R.M.}$.

regularly sampled grid in the latent space highlights the learned manifold structure. It noteworthy exhibits brain patterns gradient toward the origin (Fig. 3.3-B). Overall, the generated low-dimensional representations capture dynamic signatures of fluctuating wakefulness.

5.3.3 . Connection-wise simulations

We use external perturbations to further annotate the representation of different states of consciousness. To this end, we first study the shift in latent space induced by modifying a single connection of a dFC matrix. Using receptive field analysis, we can identify preferred directions for moving from one state to another. Second, we propose an ablation analysis to ensure that dimension reduction preserves critical information about consciousness. For the latter, specific connections related to the regions highlighted by one of the major theories of consciousness (the GNW) are zeroed, and the induced displacement in latent space is examined.

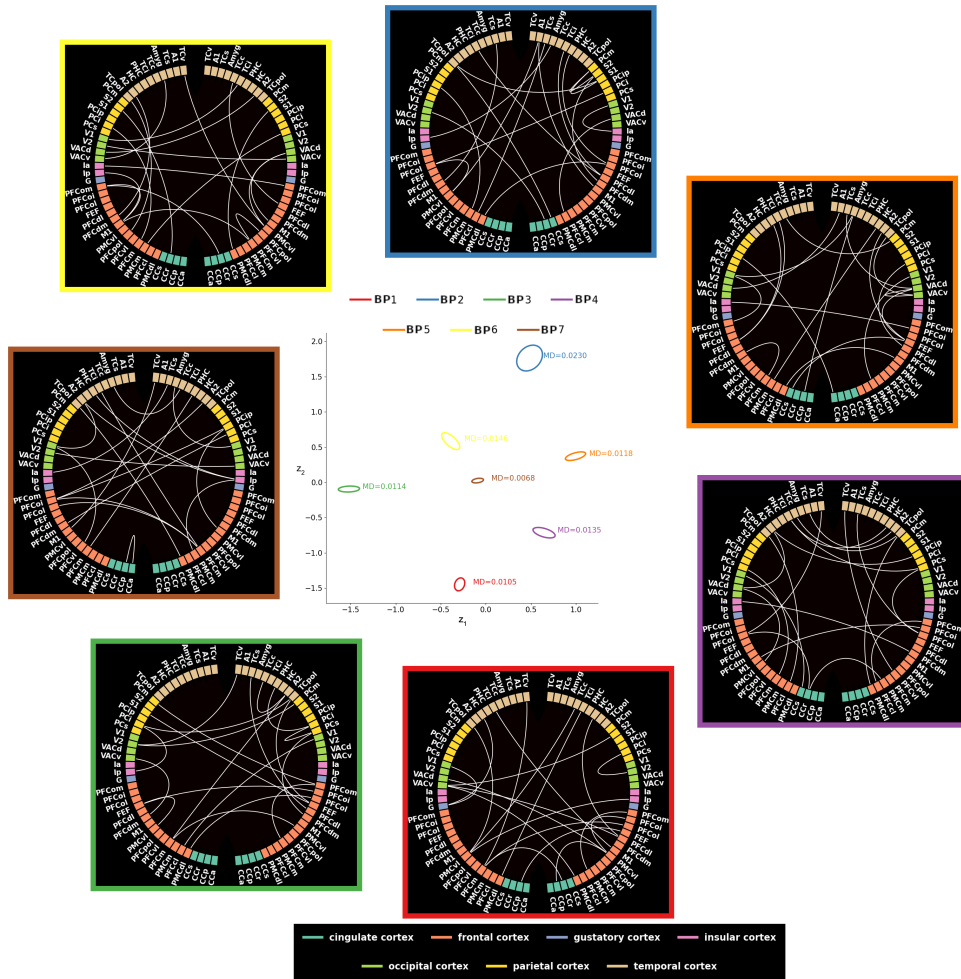


Figure 5.7: Results of RF analysis of the seven brain patterns and associated connections with a high potential for action. Using the proposed connection-wise RF analysis, a local perturbation model computed as an ellipse is derived at each encoded latent space location. Note that to improve readability, each ellipse is scaled. The associated MD is calculated. For each ellipse, the twenty connections that cause the most displacement in the latent space are displayed using a circular layout.

Perturbation of connections to study transitions Using connection-wise RF analysis, a tensor \mathcal{E} is estimated at each latent space location. We propose to focus on seven specific latent space locations that are obtained when encoding the seven brain patterns with the VAE (see central plot in Fig. 5.7). From each obtained tensor, we can characterize the overall potential for action (i.e., the chance of generating a brain pattern transition) by the Mean Diffusivity (MD) (obtained by averaging the tensor eigenvalues). We find that this potential for action is always present but is small, lying in the interval $[0.0068, 0.023]$. Nevertheless, all tensors obtained are anisotropic. Thus, it is possible to select the connections with the highest probability of generating a brain pattern transition. In this study, we keep

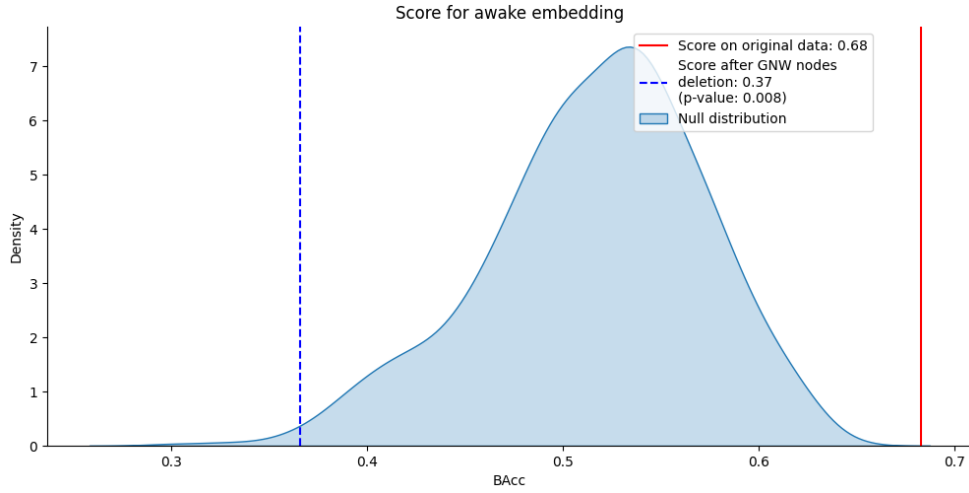


Figure 5.8: Ablation study performed from the GNW nodes. We evaluate the performance of a trained SVM classifier in predicting the awake state using the balance accuracy (BAcc). As input, we take only the raw or perturbed awake dFCs. We denote the corresponding prediction scores as $BAcc$ (vertical red dot line) and $B\tilde{A}cc$ (vertical blue dot line), respectively. We also display the histogram of $B\tilde{A}cc^i$ when random connections are removed.

twenty connections (see circular plots in Fig. 5.7). It is also interesting to note that the MD for BP_7 is minimal, making it a "stable" pattern (i.e., a perturbation of this pattern is unlikely to cause a shift in consciousness).

Ablation of connections for virtual experiments A connection-wise ablation study shows an apparent decrease in BAcc in wakefulness prediction when the GNW-associated connections are removed ($B\tilde{A}cc \ll BAcc$ in Fig. 5.8). Recall that the zeroed connections involve regions that are considered hubs in GNW theory and thus central to the processing of conscious information. We verify the significance of this decrease compared to random connection-wise ablations ($p_{val} = 0.008$). Thus, we show that a realistic state transition can be obtained by modulating a network involved in consciousness. The connection-wise ablation study highlights the relevance of the information captured in latent representations and supports the ability of the trained VAE to be an attractive computational model. To support this claim, we perform additional virtual ablation experiments (see Appendix 10).

5.4 . Discussion

We propose the VAE-VIENT framework as a tool for finding consciousness-related brain patterns, visualizing their organization, and their transitions. A VAE generative model has already been used to capture the different states of consciousness in a low-dimensional latent space. Here, we show that such a model

with tailored low-dimensional representations can be used to characterize brain dynamics over the dFCs. With low 2D-dimensional representations, the obtained performances are better than other linear (here, the PPCA) and nonlinear (here, the sVAE) generative models. However, this trend is not confirmed in higher dimensions (especially in 15D). It is generally accepted in neuroscience that simple models of neural mechanisms can be remarkably effective. In conclusion, we show that a 2D VAE model can i) generate a latent feature space stratified into a base of brain patterns, and ii) reconstruct new brain patterns coherently and stably despite the limited dataset size by exploiting the generative part of the model. Finally, we argue that the VAE-VIENT framework provides a simulation-based whole-brain computational model. Indeed, we show that the tensor fields generated from the RF analysis can model brain pattern transitions and that the proposed ablation analysis provides a unique way to non-invasively select target connections/regions. These findings pave the way for medical applications such as depth of anesthesia monitoring, coma characterization, and accurate diagnosis of disorders of consciousness in patients.

Dataset and preprocessing limitations This study has two major limitations. First, our dataset is relatively small, which increases the risk of overfitting. It will be necessary to perform tests with larger cohorts to validate our observations. One possibility is to use other studies of sleep and disorders of consciousness in humans [7, 60, 193]. Note that these datasets are also limited. However, they can be useful for validating the model with the ultimate goal of clinical translation. Second, we are working with sliding windows-based dFCs and not directly with time series. The former introduces hyperparameters that are not always easy to optimize [221, 209, 166]. Nevertheless, from a methodological point of view, we believe that working with sliding windows acts as a natural augmentation scheme that helps during the deep learning training on our limited dataset (5 monkeys - 156 runs - 72384 dFCs). Moreover, from a neuroscientific point of view, we aim to adhere to the dynamic representations of the brain originally described with dFCs [245, 14]. Note that identical conclusions have been reached in humans, using a phase-based dynamic functional coordination analysis, suggesting only a small bias (if any) induced by sliding windows [60].

Repertoire of brain patterns and arousal levels This work highlights that a 2D VAE preserves information related to brain patterns. Comparison of brain patterns using Pearson correlation similarity shows that similar brain patterns in the input space have closer latent representations (see Appendix 7, and Fig. 3.3-A). Looking at the last row of the correlation matrix between brain patterns, we see that BP_7 is highly correlated with BP_3 , BP_4 , BP_5 , and BP_6 . These patterns are also direct neighbors in latent space. Conversely, the less correlated BP_1 , and BP_2 are not direct neighbors of BP_7 in latent space. Thus, the global structure

of brain patterns can be revealed by the discovered latent space. Moreover, the performance of brain pattern classification is better than that of arousal level classification (awake vs. anesthetized) (see Appendix 6). We observe a 5% increase in classification performance. Given the difficulty of the task (i.e., a 7-class classification problem vs. a binary classification problem), the model seems to focus on the dynamic information shared between arousal levels. Similar conclusions were reached in [14, 245], where the brain pattern repertoire was described as a set of brain configurations that are unevenly distributed across arousal levels. In other words, compared to arousal levels, brain patterns provide a more detailed description of states of consciousness. On the one hand, this property may be inherited from the nature of the input data. Indeed, dFCs can be directly associated with changes in consciousness over time [7, 245]. On the other hand, the difference between levels of sedation (deep and moderate) in the present dataset is small (i.e., only a difference of one level on the monkey behavioral scale) [245]. Such a difference results in changes in reflexes (toe pinch, corneal reflex, shaking) but not in voluntary behavior (response to juice presentation). Therefore, establishing a direct relationship between a subject's level of sedation and his or her level of consciousness may be a more difficult task than characterizing overall brain dynamics. In addition, previous studies on the same dataset have shown that all three anesthetics (propofol, sevoflurane, ketamine), despite different pharmacological molecular mechanisms, imply the same dynamics of cortical activity measured with dFCs [245]. Thus, it remains to be seen whether we are unable to separate the different levels of consciousness because our data do not contain this information or because our modeling is inadequate.

Considering the time course A drawback of the current model is its inability to explicitly model the time course. We work with dynamic FC matrices, but do not consider their order in each run. However, inspired by [242], we investigate how temporal information is encoded by a 2D VAE model (see Appendix 9). Remarkably, the VAE-encoded latent variables have a coherent temporal structure that exhibits transitions characteristic of consciousness, even though no constraint is imposed during training. Other important features are the time spent consecutively in each pattern (previously called lifetime), the frequency of these steady states, and the associated transitions. Interestingly, and as described in the literature [14], the brain pattern closest to the structure (BP₇) is the most stable pattern with the longest lifetime. The latter also occupies a central place in the latent space around which other states are organized. The average lifetime is also significantly higher in the awake state than in all other anesthetized states. Similarly, the number of transitions is higher in the awake state than in all other anesthetized states. Furthermore, there is almost no difference in the number of transitions between different levels of anesthesia or between different anesthetics. In order to interpret such results with more confidence, a time-dependent model

seems essential. In the literature, some works have proposed modeling a time series with a VAE, where the encoder and decoder consist of LSTMs [142]. Other works abandon the generative property and the decoder. For example, CEBRA is a contrastive learning technique that allows label-informed time series analysis [215]. CEBRA jointly uses auxiliary variables and neural data in a hypothesis-driven manner to generate consistent, time-aware latent representations. In all cases, the goal remains the same: to obtain a consistent picture of the latent space that drives activity and behavior. In future work, we plan to directly consider the time course in the learning phase.

Performing virtual experiments An interesting finding is the ability of the VAE model to simulate transitions induced by selective ablation of connectivity between pairs of brain areas, or even ablation of connectivity within a larger network. Historically, ablation techniques have been used in animal models to directly test the function of brain areas. For example, ablation techniques have directly linked vision to the occipital lobe and auditory function to the temporal lobe [183, 171]. However, physical ablation/deactivation techniques are either irreversible or invasive and lack spatial resolution and specificity, highlighting the need for virtual ablation capabilities through the development of brain simulators [69]. Very few studies have been able to simulate the deactivation of global brain networks with the goal of suppressing consciousness. Here we present a model capable of simulating a virtual experiment in which deactivation of the "macaque GNW network" leads to suppression of consciousness. We believe that this simulation strengthens the capabilities of the model and opens up further virtual experiments that can, for example, test the specific effects of brain stimulation on consciousness [53].

Towards new biomarkers of consciousness The 2D VAE model demonstrates its ability to retain information about regions involved in conscious processing, showing that disruption of the "GNW nodes" causes a switch from a conscious to an unconscious state. It should be noted that only virtual inactivation of the entire "GNW network" (and not inactivation of individual node-related connections, see Appendix 10) causes a consciousness transition. We focus on the GNW theory of consciousness in this study because it has been translated from humans to monkeys [243]. In future work, we plan to explore other frameworks of consciousness, such as the Integrated Information Theory (IIT) [238, 12]. We can also imagine testing other networks simply by trial-and-error simulations. Setting all links connected to GNW nodes to zero is one of the limitations of the proposed ablation simulation. In reality, it is not realistic to set all connections to zero, and perhaps certain connections should be privileged (using a weighted modulation of true connection values). Conversely, it would be of great interest to show the opposite effect, i.e., to find the regions that should be stimulated to switch from an unconscious to a conscious state. As suggested by [193], this goal is challenging

and probably requires simulation. Further analysis of connection-wise RF latent space structure modeling will certainly be valuable in this context. In fact, the RF analysis relates the different patterns that reflect the dynamics of the brain (biological markers). The ellipsoids obtained in our work describe the most plausible connections to perturb in order to redirect trajectories and potentially restore wakefulness. Studying the sequences of different trajectories in latent space paves the way to a whole-brain computational model of conscious access. In other words, RF analysis provides the unique ability to identify the pairs of nodes involved in consciousness directly from the data. Changing one connection at a time is one of the limitations of the proposed RF simulation. In reality, changing multiple connections at the same time may have a more significant effect. Finally, we believe that the clinical and scientific applications are numerous. First, this approach allows the description of new biomarkers of consciousness and anesthesia-induced loss of consciousness. In addition, it is a unique tool to simulate the consequences of targeted modulation of specific brain regions for the recovery of patients suffering from disorders of consciousness. In this context, we hypothesize that the latent space structure will be essential for dissecting the exact mechanisms of DBS. It could help to build a general predictive model of the global brain effects of DBS.

Conclusions and Perspectives



To conclude this thesis, I return to the hypotheses formulated in the chapter about motivations (see 2.3.3), review a few limitations and propose a promising perspective.

In Chapter 3, we presented the results of 3 contributions.

In the papers, Grigis, Gomez et al., *Interpretable Signature of Consciousness in Resting-State Functional Network Brain Activity* (MICCAI 2022) [89] and Grigis, Gomez, et al., *Revisiting the standard for modeling functional brain network activity: application to consciousness* (preprint, 2024) [90], we were interested in applying a latent variable model to the "Anesthesia" dataset to identify spatial signatures linked to consciousness. Aligned with the GNW theory of consciousness, the brain network comprising the frontal, parietal, and cingulate cortices emerges as pivotal in distinguishing levels of consciousness. The MHA model facilitates the creation of an understandable brain decoding framework, presenting a hallmark of consciousness and anesthesia-induced loss of consciousness. The model yields outcomes that remain fairly unaffected by variations in different anesthetic agents. Hence, it could be inferred that we are deriving a universal signature of consciousness, distinct from potential indicators associated with a specific anesthetic impact.

In the paper Gomez, et al. *Exploration of the Neural Correlates of Consciousness Using Linear Latent Model* (ISBI 2023) [84], we replicated this analysis on the "DBS" dataset. This analysis underscored the significance of two pertinent networks in the processing of conscious information in anterior (prefronto-cingular) - posterior (parieto-cingular) networks, while affirming the beneficial influence of the

CT-DBS stimulation on the restoration of consciousness signatures. Additionally, the model identifies a network that captures the effects of stimulation. The model effectively disentangles the various sources of signal variability, offering valuable insights into evaluating the cortical impact of cerebral stimulation and identifying regions susceptible to collateral damage.

These three contributions allow us to assert that latent variables models can inform in an unsupervised way about cortical networks specifically related to conscious information processing, that was the first hypothesis.

In Chapter 4, we presented the results of one contribution.

In the paper Grigis, Gomez et al, *Predicting Cortical Signatures of Consciousness using Dynamic Functional Connectivity Graph-Convolutional Neural Networks* (preprint, 2022) [91], we employ a self-supervised contrastive learning strategy to predict brain patterns. The BrainNetCNN gCNN model showcases good reproducibility and accuracy in predicting BP. Its performance closely aligns with that of a linear SVC when applied to the input dynamical FC data, underscoring the simplicity or constraints imposed by the downstream classification task driven by the K-means pseudo-labels. However, the acquired latent space demonstrates capability to handle more complex tasks by acquiring intricate representations. Integration of a self-supervised contrastive learning strategy helps alleviate the circularity associated with the pseudo-labels. Notably, BrainNetCNN's predictions diverge from those of K-means during transitions between brain patterns. Beyond its role as a predictive tool, the proposed network exhibits the capacity to model dynamic brain oscillations as the brain transitions between states, thereby generating state signatures represented by sets of prominent connections. By delineating the most influential connections in predicting specific brain patterns, it becomes feasible to discern which connections are pivotal for discriminating between various levels of wakefulness, thereby offering valuable insights into brain patterns. These maps are anticipated to facilitate comprehension of the consciousness signature within different brain patterns.

This contribution confirms the second hypothesis that latent variables encode information about dynamics and transitions between states.

In the last chapter, Chapter 5, we presented the results of two contributions.

In papers Gomez et al., *Characterization of Brain Activity Patterns Across States of Consciousness Based on Variational Auto-Encoders*. (MICCAI 2022) [85] and Gomez et al. *Deep learning models reveal the link between dynamic brain connectivity patterns and states of consciousness*. (Preprint 2024) [86], we propose a VAE for finding consciousness-related dynamic brain patterns, visualizing their organization, and their transitions. A VAE generative model has previously been utilized to capture various static states of consciousness within a low-dimensional latent space. Here, we demonstrate that such a model, equipped

with customized low-dimensional representations, can effectively characterize brain dynamics across dFCs, surpassing the traditional categorical approach. Notably, with low 2D-dimensional representations, the achieved performance exceeds that of other linear (e.g., PPCA) and nonlinear (e.g., sVAE) generative models. However, this trend is not consistently observed in higher dimensions (e.g., 15D). In summary, we illustrate that a 2D VAE model can delineate a latent feature space stratified into a spectrum of brain patterns, and coherently and consistently reconstruct new brain patterns despite the constraints of limited dataset size, leveraging the generative aspect of the model. Furthermore, we contend that the VAE-VIENT framework furnishes a simulation-based computational model of the entire brain. Specifically, we demonstrate that tensor fields generated from the RF analysis can effectively model brain pattern transitions, and the proposed ablation analysis offers a non-invasive method for selecting target connections/regions.

These contributions validate the third hypothesis that the traditional categorical approach does not account for the continuum of the dynamics of states of consciousness. Although current studies in a therapeutic setting mainly aim at characterizing static states, this continuum of dynamic states may be important to consider. For example, consider the analogy of depicting a map of the world. Typically, geographical proximity dictates the spatial arrangement: neighboring countries are depicted close together on the map. This approach provides a useful initial approximation when examining cultural aspects. However, when delving into language distribution, this representation falls short in explaining phenomena like why English is spoken in Gibraltar while Spanish is prevalent in some South American countries and Portuguese in others. To elucidate such intricacies, one must delve into historical trajectories – in essence, analyzing temporal dynamics. Conversely, if our world map is based solely on the languages spoken in each country as of 2024, it ceases to represent geographical features accurately. Consequently, characterizing continents, or analogously, levels of consciousness, becomes challenging.

As for the fourth hypothesis, that reducing the observation space to a very low-dimension is sufficient to separate the levels of consciousness, while the ablation experiment partly validates it, it requires further experiments to show its generalizability to an external data set.

These contributions present original work, cross-referencing findings associated with theories of consciousness in an unsupervised way. However, the data used are limited, and the preliminary results found will need to be replicated on larger cohorts for more far-reaching conclusions. Deep learning methods were exploited as part of a computational approach to brain modeling, but we ultimately favored simpler approaches in order to better answer certain questions and not over-dimension our models. Black boxes undoubtedly have the potential to make brain function more transparent, but they require an increased amount of data to elucidate conscious-

ness.

In particular, an interesting perspective to this work comes from a method belonging to the VAEs family, the Contrastive Analysis Variational Auto-encoders (CA-VAEs). They are designed to distinguish between shared factors of variation present in both a background dataset (BG) (i.e. representing healthy subjects) and a target dataset (TG) (i.e. representing patients) and those unique to the target dataset. These methods partition the latent space into distinctive features specific to the target dataset (salient features) and those common to both datasets. These methods, first applied to natural images [220, 1], have also been used on medical images, in the context of autism [3], pneumonia radiographs [149] and schizophrenic patients [149].

To understand better the interest of these methods, let me quote the introductory example of Louiset et al. [149]. Let's consider two distinct datasets: 1) a collection of neuro-anatomical MRIs from healthy individuals (BG=background dataset), and 2) MRIs from patients diagnosed with Alzheimer's disease (TG=target dataset). Neuroscientists are often interested in discerning shared factors of variation, such as those related to aging, education, or gender, from specific markers of Alzheimer's disease, such as temporal lobe atrophy or an increase in beta-amyloid plaques. Until recently, disentangling the complex interplay of latent mechanisms underlying neuro-anatomical variability in neurodegenerative disorders was deemed challenging. This complexity arises from the overlapping effects of natural aging and the progression of neurodegenerative diseases. The intertwined nature of these processes has made it difficult to interpret potential discoveries of new biomarkers. The aim of developing a Contrastive Analysis method is to facilitate the separation of these processes. In the shared features space, patterns associated with aging should correlate with typical cognitive decline observed in healthy individuals, while salient features (i.e., Alzheimer's-specific patterns) should correlate with pathological cognitive decline characteristic of Alzheimer's disease. SepVAE [149], in particular, introduces two critical regularization losses: one that disentangles common and salient representations and another that classifies background and target samples within the salient space. It demonstrates superior performance compared to previous CA-VAEs methods.

In the case of studies of functional connectivity, such methods, transposed to the context of disorders of consciousness, also have their place. While our model highlighted major network differences between awake and anesthetized, stimulated and unstimulated brains, and confirmed the role of DBS in conscious arousal, we did not finely model the different sources of variability present in our data. However, we know that certain networks detected by the method are artifactual, notably the visual network during the study of consciousness on the "Anesthesia" dataset, or the network associated with stimulation, which is localized around the electrode used on the "DBS" dataset. As these networks are likely not involved in conscious-

ness, we would like the learned representations to ignore them, and better still, for the model to categorize this information as a bias automatically. In order to model this, we would like to apply contrastive analysis to our data. The general idea is to separate the latent space into two, with a first block of latent variables modeling the specific variability linked to awareness and another block capturing the general variability linked to acquisition conditions. This would enable us to focus on the mechanisms induced by the variations in consciousness generated by a DBS, while remaining invariant to the acquisition conditions. This method would help refine our first model by denoising our data, enabling highly accurate detection of variations caused directly by DBS or other factors. This would help refine the use of DBS for patients with disorders of consciousness, improving therapeutic targets.

Finally, and because we know that one of our project's main limitations is the lack of data, we wanted to emphasize that this thesis is part of a FAIR (Findable, Accessible, Interoperable, Reusable) research project. No new data has been acquired. We are re-using data, for an activity other than the one initially planned at the time of collection, which generates a certain number of challenges (lack of data, sparsity of conditions, challenges of preprocessing reproducibility, constraints of choices made beforehand). But reuse also saves time and money: the cost of creating, collecting and processing data can be very high. Reusing existing data rather than recreating it makes the time spent on acquisition more profitable. In the context of research using animal data, which is increasingly criticized, reusing data to extract the maximum amount of information before repeating experiments also seems fundamental to us. It also enables reproducibility to be tested and interdisciplinarity to be added: having larger databases, linked to several experiments, enables data to be searched, cross-referenced and visualized. Data management and sharing facilitate new research and the cross-fertilization of data from different disciplines, as here at the intersection of machine learning and neuroscience.

Appendix 1

Arousal scale based on the monkey sedation scale (behavior) and EEG

Arousal scale.		Monkey sedation scale					EEG	
Scale	Arousal	Behavior					Propofol	Ketamine
		Juice presentation	Spontaneous movements	Shaking/prodding	Toe pinch	Corneal reflex		
1	Alert/awake	+	+	+	+	+	Posterior alpha waves, anterior beta waves	Posterior alpha waves, anterior beta waves
2	Light sedation	-	+	+	+	+	Increased amplitude of alpha waves, anterior diffusion of alpha waves	Loss of alpha rhythm, decrease of amplitude
3	Moderate sedation	-	-	±	+	+	Diffuse and wide alpha waves, anterior theta waves	Persistent rhythmic theta activity, increasing amplitude, beta activity of low amplitude
4	Deep anesthesia/general anesthesia	-	-	-	-	-	Diffuse delta waves, waves of low amplitude, anterior alpha waves	Intermittent polymorphic delta activity of large amplitude, superimposed beta activity of low amplitude, increase in gamma power

Arousal scale based on the monkey sedation scale (behavior) and EEG for propofol, respectively ketamine anesthesia Note: Response to juice presentation: the experimenter presents a syringe with juice/water. (+) if the monkey drinks, (-) if the monkey fails to drink. Spontaneous movements: (+) if the monkey exhibits spontaneous movements, (-) if spontaneous movements are absent. Shaking/prodding: (+) if the monkey exhibits a response (body movement, eye blinking, eye opening, cardiac rate change) on Shaking/prodding, (-) if there is no response. Toe pinch: (+) if the monkey exhibits a response (body movement, eye blinking, eye opening, cardiac rate change) to toe pinch, (-) if there is no response. Corneal reflex: (+) if the corneal reflex is present, (-) if the corneal reflex is absent. From [244]

Appendix 2

Number and name of the 82 ROIs of the CoCoMac atlas

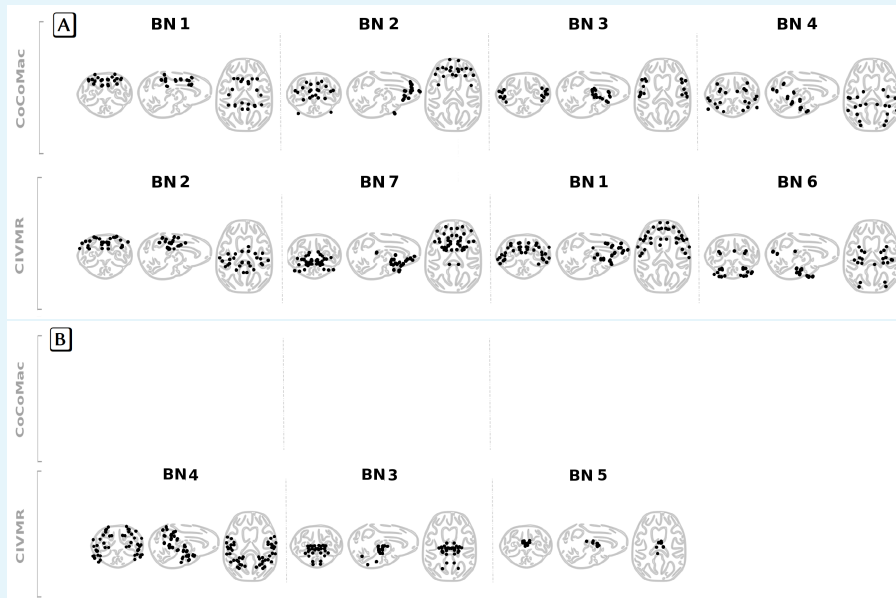
Supplementary Table 7: Number and name of the 82 ROIs

	Region Number	Region Name
Left Hemisphere	1	Templar polar cortex
	2	Superior temporal cortex
	3	Amygdala
	4	Orbito-inferior prefrontal cortex
	5	Anterior insula
	6	Orbito-medial prefrontal cortex
	7	Central temporal cortex
	8	Orbito-lateral prefrontal cortex
	9	Inferior temporal
	10	Parahippocampal cortex
	11	Gustatory cortex
	12	Ventrolateral premotor cortex
	13	Anterior visual area (ventral)
	14	Posterior insula
	15	Prefrontal polar cortex
	16	Hippocampus
	17	Subgenual cingulate cortex
	18	Ventrolateral prefrontal cortex
	19	Visual area 2
	20	Medial prefrontal cortex
	21	Ventral temporal cortex
	22	Anterior visual area (dorsal)
	23	Visual area 1
	24	Centrolateral prefrontal cortex
	25	Secondary auditory cortex
	26	Retrosplenial cingulate cortex
	27	Posterior cingulate cortex
	28	Anterior cingulate cortex
	29	Secondary somatosensory cortex
	30	Primary somatosensory cortex
	31	Primary auditory cortex
	32	Primary motor cortex
	33	Inferior parietal cortex
	34	Medial parietal cortex
	35	Dorsomedial prefrontal cortex
	36	Intraparietal cortex
	37	Superior parietal cortex
	38	Frontal eye fields
	39	Dorso-lateral prefrontal cortex
	40	Medial premotor cortex
41	Dorso-lateral premotor cortex	
Right Hemisphere	42	Templar polar
	43	Superior temporal cortex
	44	Amygdala
	45	Orbito-inferior prefrontal cortex
	46	Anterior insula
	47	Orbito-medial prefrontal cortex
	48	Central temporal cortex
	49	Orbito-lateral prefrontal cortex
	50	Inferior temporal
	51	Parahippocampal cortex
	52	Gustatory cortex
	53	Ventrolateral premotor cortex
	54	Anterior visual area (ventral)
	55	Posterior insula
	56	Prefrontal polar cortex
	57	Hippocampus
	58	Subgenual cingulate cortex
	59	Ventrolateral prefrontal cortex
	60	Visual area 2
	61	Medial prefrontal cortex
	62	Ventral temporal cortex
	63	Anterior visual area (dorsal)
	64	Visual area 1
	65	Centrolateral prefrontal cortex
	66	Secondary auditory cortex
	67	Retrosplenial cingulate cortex
	68	Posterior cingulate cortex
	69	Anterior cingulate cortex
	70	Secondary somatosensory cortex
	71	Primary somatosensory cortex
	72	Primary auditory cortex
	73	Primary motor cortex
	74	Inferior parietal cortex
	75	Medial parietal cortex
	76	Dorsomedial prefrontal cortex
	77	Intraparietal cortex
	78	Superior parietal cortex
	79	Frontal eye fields
	80	Dorso-lateral prefrontal cortex
	81	Medial premotor cortex
	82	Dorso-lateral premotor cortex

From [245]

Appendix 3

Brain networks derived from different atlas



The derived BNs consist of sets of unique ROIs represented by their centroids for the CoCoMac ($k=4$) and CIVMR ($k=7$) atlases.

Appendix 4

Listing of the networks inferred from the CoCoMac atlas with $k=3$

	name	hemi	location
CCp	posterior cingulate cortex	left, right	cingulate cortex
CCa	anterior cingulate cortex	left, right	cingulate cortex
S1	primary somatosensory cortex	left, right	parietal cortex
PCi	inferior parietal cortex	left, right	parietal cortex
PCm	medial parietal cortex	left, right	parietal cortex
PCip	intraparietal cortex	left, right	parietal cortex
PCs	superior parietal cortex	left, right	parietal cortex
M1	primary motor cortex	left, right	frontal cortex
FEF	frontal eye field	left, right	frontal cortex
PMCm	medial premotor cortex	left, right	frontal cortex
PMCDl	dorsolateral premotor cortex	left, right	frontal cortex
PMCVl	ventrolateral premotor cortex	left, right	frontal cortex
PFCdm	dorsomedial prefrontal cortex	left, right	frontal cortex
G	gustatory cortex	left, right	gustatory cortex

Listing of the network 1. The detected GNW areas are depicted in blue, and the associated sensory areas in green.

	name	hemi	location
TCpol	temporal polar	left, right	temporal cortex
Amyg	amygdala	right	temporal cortex
PFCoi	orbitoinferior prefrontal cortex	left, right	frontal cortex
PFCom	orbitomedial prefrontal cortex	left, right	frontal cortex
PFCol	orbitolateral prefrontal cortex	left, right	frontal cortex
PFCpol	prefrontal polar cortex	left, right	frontal cortex
PFCvl	ventrolateral prefrontal cortex	left, right	frontal cortex
PFCm	medial prefrontal cortex	left, right	frontal cortex
PFCcl	centrolateral prefrontal cortex	left, right	frontal cortex
PFCdl	dorsolateral prefrontal cortex	left, right	frontal cortex
CCs	subgenual cingulate cortex	left, right	cingulate cortex

Listing of the network 2. The detected GNW areas are depicted in blue, and the associated sensory areas in green.

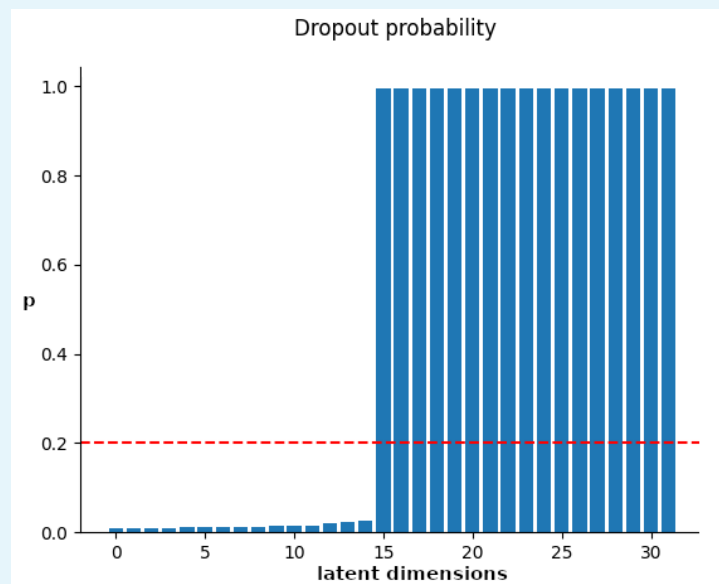
	name	hemi	location
TCs	superior temporal cortex	left, right	temporal cortex
la	anterior insula	left, right	insular cortex
lp	posterior insula	left, right	insular cortex
A2	secondary auditory cortex	left, right	temporal cortex
S2	secondary somatosensory cortex	left, right	parietal cortex
A1	primary auditory cortex	left, right	temporal cortex
TCc	central temporal cortex	left, right	temporal cortex
CCr	retrosplenial cingulate cortex	right	cingulate cortex
TCi	inferior temporal	left	temporal cortex

Listing of the network 3. The detected GNW areas are depicted in blue, and the associated sensory areas in green.

Appendix 5

Dropout regularization in sVAE training

During sVAE training, parsimonious and interpretable representations are enforced by variational dropout. Model selection in latent space can then be achieved using this technique. Here, the dropout rate after convergence is shown when the initial latent dimensions are set to 32 (Fig. 1). Note that the learned dropout rate is highly contrasted. For this reason, model selection can be done by keeping the latent dimensions that meet a suitable dropout rate threshold. We can see that it is possible to safely select the best model with a threshold $p < 0.2$ (as proposed in the original paper [8]).

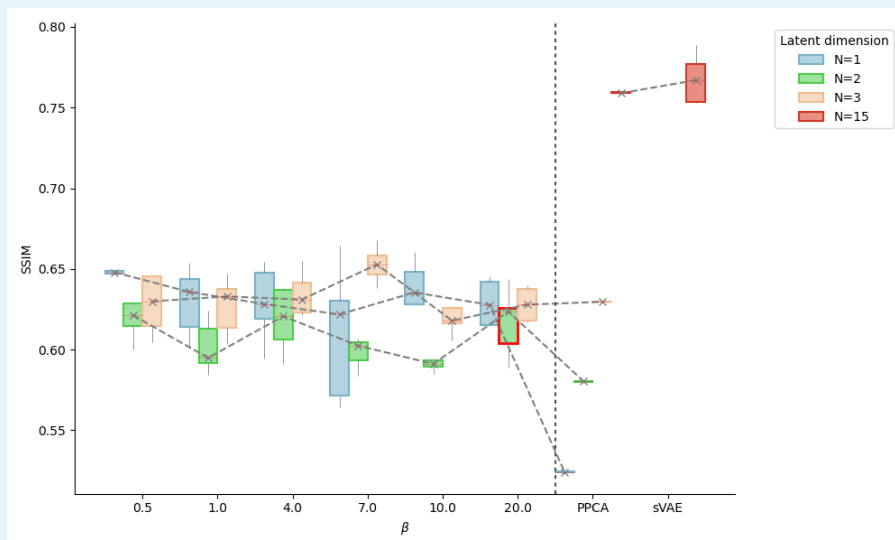


Appendix 5—figure 1: sVAE estimated dropout rate.

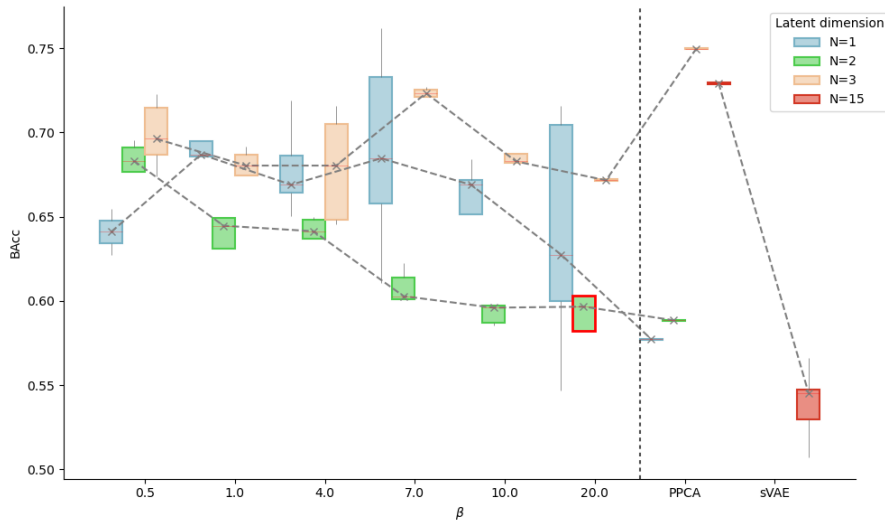
Appendix 6

Model evaluation using arousal conditions

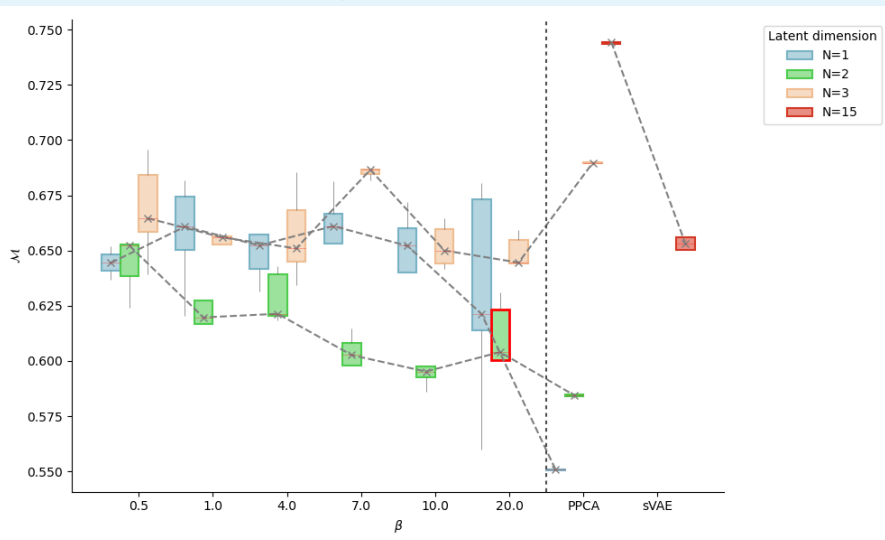
The proposed model evaluation relies on the brain pattern labels. Indeed, it has been shown that these labels can effectively represent the different configurations of the brain [14, 245]. Here, we re-evaluate the reconstruction quality, classification accuracy, and consensus metric using a different set of labels composed of the arousal conditions. In particular, we address a binary classification problem between the awake and the anesthetized data (where all the associated acquisition conditions are grouped together). First, the dFCs are well reconstructed for all models when looking at the reconstruction quality (using the structural similarity (SSIM) metric) (Fig. 1). We note that i) for the VAEs, the chosen β has little effect on the reconstruction, ii) for the PPCA baseline, increasing the latent space improves the reconstruction, but this is not the case for the nonlinear models, and iii) the sVAE with the chosen fifteen dimensions performs best. Second, it does not seem trivial to classify a dFC matrix as belonging to the conscious or unconscious category. This is consistent with previous findings showing that dFC matrices of one condition can be associated with different brain patterns [245]. Finally, looking at the consensus metric, the least constrained VAE models ($\beta = 0.5$ and $\beta = 1$) perform best in low dimensions (Fig. 3). Beyond three dimensions, the PPCA stands out. This may be due to the "relative" simplicity of our task.



Appendix 6—figure 1: VAE, PPCA, sVAE reconstruction quality: SSIM of label-wise averaged dFCs with respect to the β regularization parameters.



Appendix 6—figure 2: VAE, PPCA, sVAE classification accuracy: BACC between the ground truth and the matched predicted labels.



Appendix 6—figure 3: The proposed consensus metric \mathcal{M} .

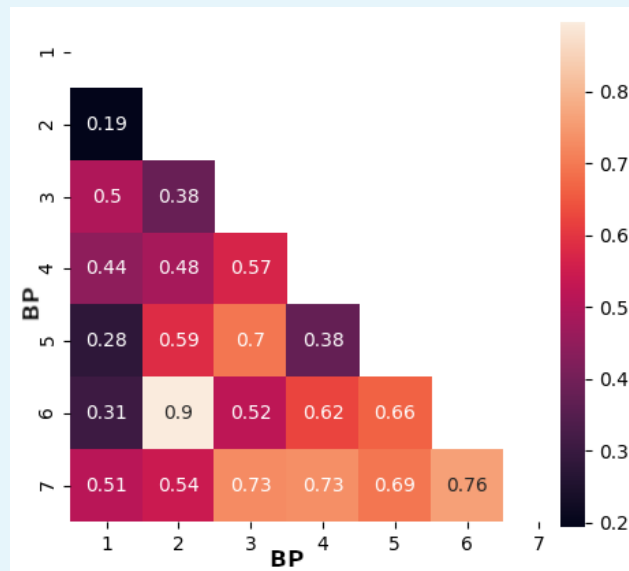
Appendix 7

Additional brain pattern analyses

The 2D latent representations obtained with a VAE₂ can be stratified using the brain pattern (BP) labels. To quantify the overlap between brain pattern locations in latent space, we use the Dice similarity coefficient (Tab. 1). The Dice metric yields values between 0 (no spatial overlap) and 1 (complete overlap) [64]. Unfortunately, the Dice metric can only be applied to array-like data. Therefore, we choose to perform a brain pattern-wise re-gridding (using a 60 × 60 grid) of the obtained latent representations, which produces a binary array-like support per brain pattern (numbered 1 to 7). The Pearson correlation matrix between the brain patterns further describes the studied brain repertoire (Fig. 1). Overall, these two experiments allow us to better characterize the learned latent space in terms of brain patterns.

BP	1	2	3	4	5	6	7
Dice	0.24±0.17	0.07±0.11	0.24±0.15	0.32±0.20	0.24±0.16	0.28±0.08	0.37±0.19

Appendix 7—table 1: Averaged across folds Dice coefficients and associated standard deviations for each BP embedding.

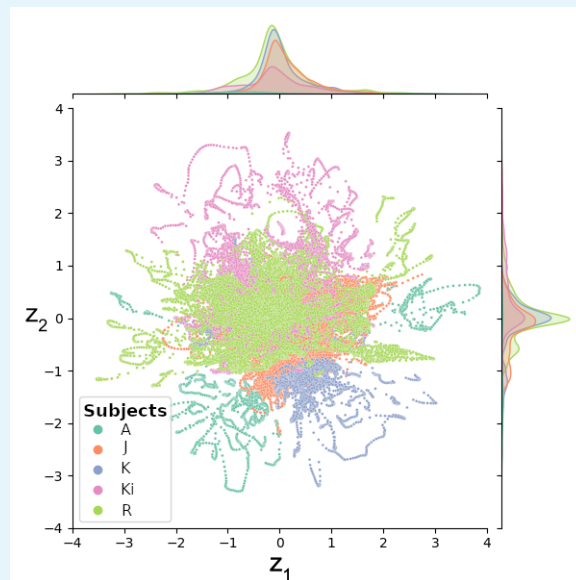


Appendix 7—figure 1: Correlation matrix between brain patterns.

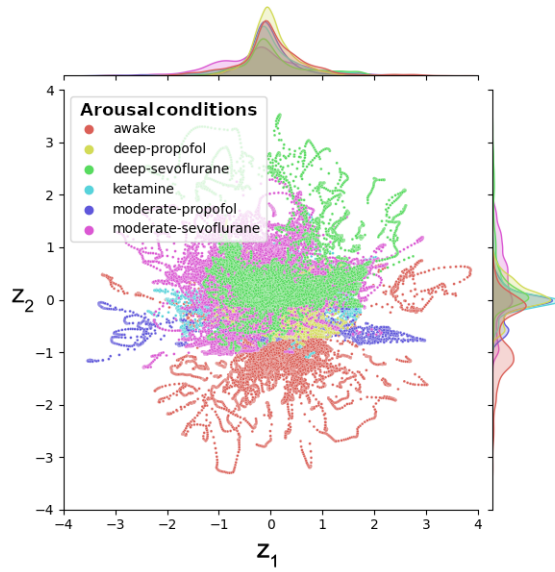
Appendix 8

Explore learned latent representations with different labels

The present study focuses on brain pattern labels. Indeed, it has been shown that these labels can effectively represent the different configurations of the brain [14, 245]. Two different sets of labels are considered below. First, the subjects to investigate whether the VAE₂ has incorrectly learned subject-specific information (i.e., there is some overfitting during training in our small dataset) (Fig. 1). Second, the acquisition conditions to verify that no anesthetic effect (known to have different vascular effects) can be observed in the latent representations (Fig. 2).



Appendix 8—figure 1: Stratification of VAE₂ latent representations by subject.

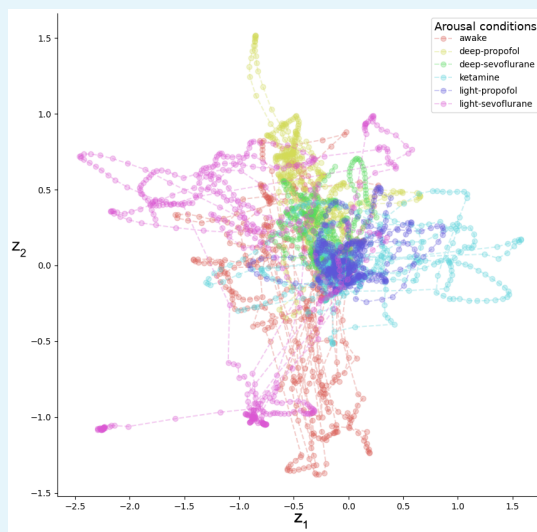


Appendix 8—figure 2: Stratification of VAE_2 latent representations by acquisition conditions.

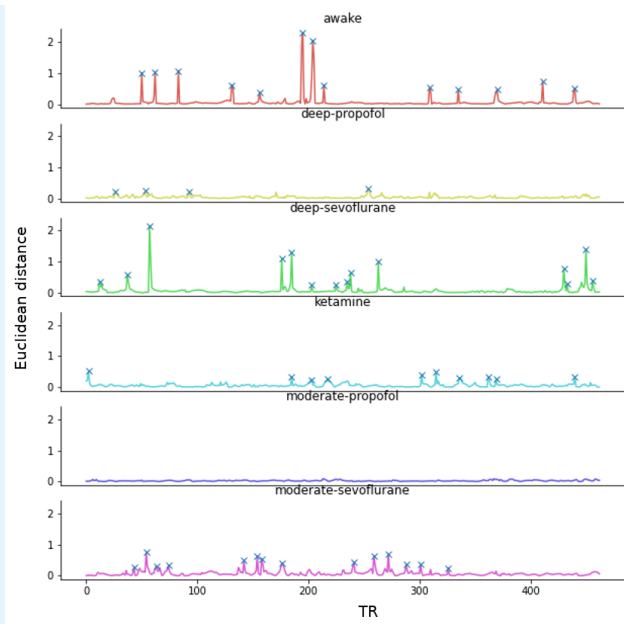
Appendix 9

Temporal structure encoded in latent space

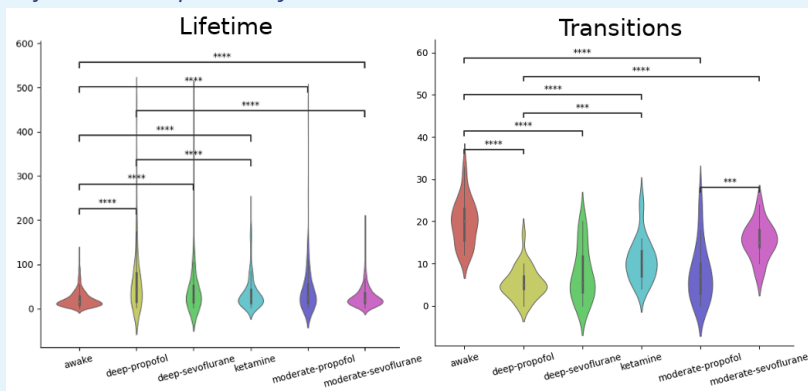
Following the idea of [242] to study meta-transitions between brain configurations, we examine the encoding of temporal information using the selected VAE_2 . While no modeling of the acquisition time course is enforced during training, the encoded latent variables have a coherent temporal structure (Fig. 1). This is a sign of successful modeling. It is then possible to examine the temporal transitions within the same run by calculating the Euclidean distance between each successive encoded time point (i.e., replaying the dFC movie) (Fig. 2). Stable periods have a Euclidean distance close to zero, and a transition occurs when a jump in the metric is observed (indicated by blue crosses). Transitions between brain configurations have stable periods, which we call meta-stable states. The average lifetime of a meta-stable state, i.e. the time spent continuously in this state, is significantly higher in the awake state than in all other anesthetized states (Fig. 3). Similarly, the number of transitions is higher in the awake state than in all other anesthetized states (Fig. 3). Interestingly, there is almost no difference between different levels of anesthesia or between different anesthetics.



Appendix 9—figure 1: Latent representations of six runs, one per arousal condition.



Appendix 9—figure 2: *The Euclidean distance between two consecutive time points for each run previously selected. Blue crosses mark transitions.*

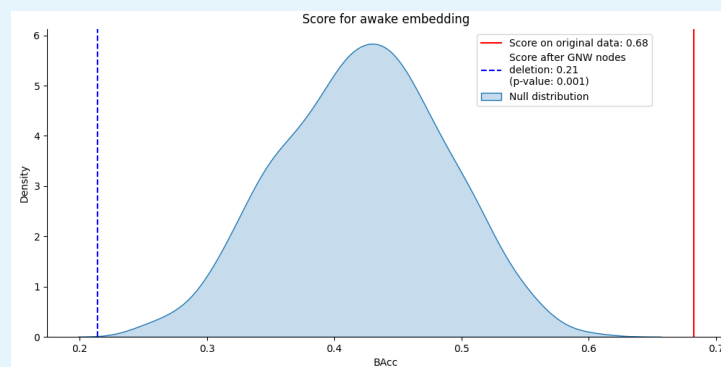


Appendix 9—figure 3: *Meta-stable state lifetime and transition occurrence distributions across acquisition conditions.*

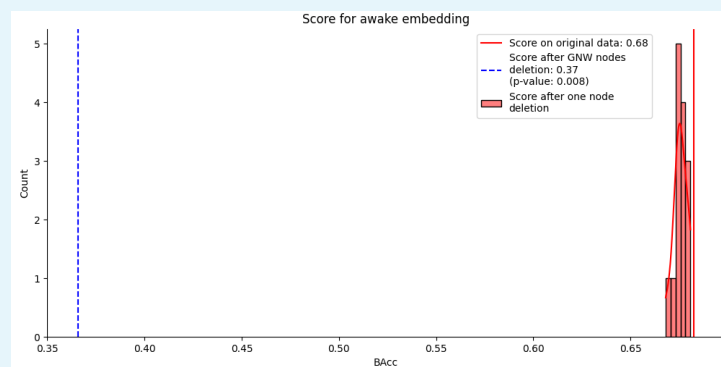
Appendix 10

Additional virtual ablation experiments

The proposed ablation analysis provides a unique way to non-invasively select target connections/regions. First, the GNW key regions are expanded to include primary sensory regions, primary somatosensory cortex S1, primary auditory cortex A1, and visual area V1. When the awake state is altered, the addition of these target regions in the ablation study further increases the statistical significance of the results (Fig. 1). However, perhaps only a few important regions drive the prediction performance. Thus, in the proposed analysis, instead of removing all GNW-related connections, only one GNW-related connection is removed. By using all combinations, the histogram of prediction performance is computed (Fig. 2). In conclusion, removing one GNW-related connection is insufficient to produce a significant shift in awareness.



Appendix 10—figure 1: Ablation study considering GNW and sensory related connections.



Appendix 10—figure 2: Ablation study considering GNW-related connections one by one. The resulting histogram of prediction performance is shown in red.

Appendix 11

Résumé en français

Le diagnostic des patients dans le coma est souvent difficile. Les examens cérébraux renseignent les médecins sur l'étendue des lésions cérébrales mais ne permettent pas de déterminer avec précision l'état de conscience du patient. De plus, aucune approche thérapeutique ne permet une restauration systématique de la conscience. Des études pionnières menées sur des patients et des Primates Non Humains (PNH) ont montré que la Stimulation Cérébrale Profonde (SCP) des noyaux intralaminaires du thalamus pouvait restaurer ou améliorer la conscience lorsqu'elle est altérée. Cependant, les conséquences corticales associées à la SCP restent largement inconnues et imprévisibles. Les techniques d'imagerie fonctionnelle, telles que l'Imagerie par Résonance Magnétique fonctionnelle de repos (IRMf de repos), peuvent aider à identifier des signatures de la conscience. L'activité cérébrale au repos, organisée en réseaux, peut être modélisée à l'aide de la connectivité fonctionnelle. Cette thèse vise à disséquer, à l'aide du modèle PNH, les effets sur la connectivité fonctionnelle d'une modulation de la conscience induite par des agents anesthésiques ou de la SCP à l'échelle du cerveau entier. Cela nécessite le développement de modèles interprétables et prédictifs des effets d'une telle modulation sur la fonction cérébrale globale. Nous travaillons principalement sur deux jeux de données précédemment acquis :

- jeu de données ANESTHÉSIE : IRMf acquise chez des PNHs dans des conditions d'éveil et d'anesthésie avec différents anesthésiants [14, 245].
- jeu de données SCP : IRMf acquise chez des PNHs dans des conditions d'éveil et d'anesthésie avec différents anesthésiants, avec ou sans SCP [234].

Nous émettons l'hypothèse que l'identification de variables latentes dans les signaux IRMf de repos peut nous informer sur la modulation des états de conscience.

Dans le Chapitre 3, tout d'abord, nous cherchons à identifier une signature spatiale, moyennée temporellement, de la conscience à la fois dans l'état éveillé et sous anesthésie. Nous y présentons les résultats de trois contributions. Nous utilisons une méthode à variables latentes (le MHA) qui décompose les signaux IRMf de repos en réseaux fonctionnels associés à l'accès conscient. Dans les articles Grigis, Gomez et al (MICCAI 2022) [89] et Grigis, Gomez, et al, (preprint, 2024) [90], nous appliquons ce modèle sur le jeu de données ANESTHÉSIE. Le réseau cérébral comprenant les

cortex fronto-parieto-cingulaire apparaît comme central dans la distinction des niveaux de conscience. Le modèle fait donc ressortir de manière automatique, un résultat en accord avec une grande théorie de la conscience, la théorie du Global Neuronal Workspace (GNW). Le modèle MHA permet d'aider au décodage cérébral, mettant en valeur une signature de la conscience et de la perte de conscience induite par l'anesthésie. Les résultats ne sont pas affectés par les variations entre différents agents anesthésiques. Nous pouvons donc en déduire que nous obtenons une signature globale de la conscience, distincte des effets potentiels associés à un anesthésiant spécifique. Afin d'étudier la restauration de la conscience, nous étendons cette analyse aux PNH éveillés ou réveillés par DBS du thalamus central. Pour cela, dans l'article Gomez, et al. (ISBI 2023) [84], nous appliquons le MHA sur le jeu de données SCP. Le modèle suggère de manière automatique que le cortex antérieur (préfronto-cingulaire) et le cortex postérieur (pariéto-cingulaire) sont tous deux impliqués dans la conscience, un sujet qui fait débat au sein de la communauté scientifique. L'analyse confirme également l'influence positive de la SCP du thalamus central sur la restauration des signatures de la conscience. En outre, le modèle identifie un réseau qui capture les effets de la stimulation. Le modèle démêle efficacement les différentes sources de variabilité du signal, offrant des indications précieuses pour l'évaluation de l'impact cortical de la stimulation cérébrale et l'identification des régions susceptibles de subir des dommages collatéraux. Ces trois contributions nous permettent d'affirmer que les modèles de variables latentes peuvent informer de manière non supervisée sur les réseaux corticaux spécifiquement liés au traitement de l'information consciente.

Suite à cette analyse moyennée temporellement, reconnaissant l'importance de la dynamique temporelle dans l'analyse de la conscience, nous proposons de remettre en question les méthodes conventionnelles de connectivité fonctionnelle dynamique. Pour identifier les schémas récurrents dominants (c'est-à-dire les différents états du cerveau) à partir de la connectivité fonctionnelle, une technique d'apprentissage automatique non supervisée (K-Means) a été proposée précédemment. Dans le cadre de cette thèse, nous développons de nouveaux outils d'analyse en tirant parti des avancées des techniques d'apprentissage profond auto-supervisé.

Dans le Chapitre 4, nous utilisons un modèle d'apprentissage profond auto-supervisé contrastif pour prédire les schémas cérébraux caractéristiques de différents états de conscience (Grigis, Gomez et al (preprint, 2022) [91]). Ses performances s'alignent étroitement sur celles d'un modèle linéaire SVC soulignant la simplicité de la tâche de classification en aval, guidée par les résultats du K-means. Toutefois, l'espace latent acquis démontre la capacité du modèle à traiter des tâches plus difficiles en acquérant des représentations complexes. L'intégration d'une stratégie d'apprentissage contrastif auto-supervisé permet d'atténuer la circularité associée aux pseudo-étiquettes du

K-means. Notamment, les prédictions du modèle divergent de celles du K-means lors des transitions entre les schémas cérébraux. Au-delà de son rôle d'outil prédictif, le réseau proposé présente la capacité de modéliser les oscillations dynamiques du cerveau lorsque celui-ci passe d'un état à l'autre. Il est alors possible de générer une carte des connexions prédominantes impliquées dans les changements d'états. En délimitant les connexions les plus influentes dans la prédiction de schémas cérébraux spécifiques, il devient possible de discerner les connexions qui sont essentielles à la discrimination entre les différents niveaux d'éveil, offrant ainsi des informations précieuses sur les schémas cérébraux. Cette contribution confirme que les variables latentes encodent des informations sur la dynamique et les transitions entre les états.

Enfin, pour mieux comprendre la dynamique des états de conscience, dans le Chapitre 5, nous nous écartons du cadre conventionnel de classification en sous-groupes et introduisons une méthode de réduction de dimensions. Cette approche vise à condenser les états de conscience en un nombre limité de variables interprétables et explicables. Dans les contributions Gomez et al. (MICCAI 2022) [85] et Gomez et al. (Preprint 2024) [86], nous proposons un Auto-Encoder Variationnel (VAE) pour trouver une organisation dynamique d'états cérébraux, visualiser leur organisation et leurs transitions. Un modèle génératif de VAE a déjà été utilisé pour capturer divers états statiques de la conscience dans un espace latent de faible dimension (différentes phases du sommeil par exemple). Nous démontrons ici qu'un tel modèle, à deux variables latentes, peut caractériser efficacement la dynamique cérébrale à travers les matrices de connectivité fonctionnelle dynamique, surpassant l'approche catégorielle traditionnelle du K-means. En deux dimensions, les performances du VAE dépassent celles d'autres modèles génératifs linéaires (par exemple, la PCA probabiliste) et non linéaires (par exemple, le VAE sparse). Toutefois, cette tendance n'est plus observée dans des dimensions plus élevées (par exemple, en 15 dimensions). En résumé, nous montrons qu'un modèle VAE 2D fait apparaître un espace de caractéristiques stratifié suivant les schémas cérébraux, et qu'il permet de reconstruire de manière cohérente de nouveaux schémas cérébraux, en tirant parti de l'aspect génératif du modèle. De plus, les simulations réalisées (étude du champs récepteur et étude d'ablation) offrent une première approche pour modéliser des transitions entre schémas cérébraux. Elles peuvent aussi servir de méthodes non invasives pour sélectionner les connexions/régions cibles impliquées dans le passage vers l'éveil conscient.

Ces contributions présentent un travail original, recoupant des résultats associés à des théories de la conscience de manière non supervisée. Toutefois, les données utilisées sont limitées et les résultats préliminaires obtenus devront être reproduits sur des cohortes plus importantes pour obtenir des conclusions plus ambitieuses. Les méthodes d'apprentissage profond ont été

exploitées dans le cadre d'une approche computationnelle de la modélisation du cerveau, mais nous avons finalement privilégié des approches plus simples afin de mieux répondre à certaines questions et de ne pas surdimensionner nos modèles. Les boîtes noires ont sans aucun doute le potentiel de rendre le fonctionnement du cerveau plus transparent, mais elles nécessiteront une quantité accrue de données pour élucider le mystère de la conscience.

Bibliography

- [1] Abubakar Abid and James Zou. Contrastive Variational Autoencoder Enhances Salient Features, February 2019. arXiv:1902.04601 [cs, stat].
- [2] Sophie Achard, Chantal Delon-Martin, Petra E. Vértes, Félix Renard, Maleka Schenck, Francis Schneider, Christian Heinrich, Stéphane Kremer, and Edward T. Bullmore. Hubs of brain functional networks are radically reorganized in comatose patients. *Proceedings of the National Academy of Sciences*, 109(50):20608–20613, December 2012. Publisher: Proceedings of the National Academy of Sciences.
- [3] Aidas Aglinskas, Joshua K. Hartshorne, and Stefano Anzellotti. Contrastive machine learning reveals the structure of neuroanatomical variation within autism. *Science*, 376(6597):1070–1074, June 2022. Publisher: American Association for the Advancement of Science.
- [4] Christine Ahrends and Diego Vidaurre. Dynamic Functional Connectivity, January 2023.
- [5] Michael T. Alkire, Christopher D. Asher, Amanda M. Franciscus, and Emily L. Hahn. Thalamic Microinfusion of Antibody to a Voltage-gated Potassium Channel Restores Consciousness during Anesthesia. *Anesthesiology*, 110(4):766–773, April 2009.
- [6] Michael T. Alkire, Jayme R. McReynolds, Emily L. Hahn, and Akash N. Trivedi. Thalamic Microinjection of Nicotine Reverses Sevoflurane-induced Loss of Righting Reflex in the Rat. *Anesthesiology*, 107(2):264–272, August 2007.
- [7] Elena A. Allen, Eswar Damaraju, Sergey M. Plis, Erik B. Erhardt, Tom Eichele, and Vince D. Calhoun. Tracking whole-brain connectivity dynamics in the resting state. *Cerebral Cortex (New York, N.Y.: 1991)*, 24(3):663–676, March 2014.
- [8] Luigi Antelmi, Nicholas Ayache, Philippe Robert, and Marco Lorenzi. Sparse Multi-Channel Variational Autoencoder for the Joint Analysis of Heterogeneous Data. In *International Conference on Machine Learning*, pages 302–311. PMLR, May 2019.
- [9] Bernard J. Baars. *A Cognitive Theory of Consciousness*. Cambridge University Press, New York, 1988.

- [10] Jonathan L. Baker, Jae-Wook Ryou, Xuefeng F. Wei, Christopher R. Butson, Nicholas D. Schiff, and Keith P. Purpura. Robust modulation of arousal regulation, performance, and frontostriatal activity through central thalamic deep brain stimulation in healthy nonhuman primates. *Journal of Neurophysiology*, 116(5):2383–2404, November 2016. Publisher: American Physiological Society.
- [11] Rembrandt Bakker, Thomas Wachtler, and Markus Diesmann. CoCoMac 2.0 and the future of tract-tracing databases. *Frontiers in Neuroinformatics*, 0, 2012. Publisher: Frontiers.
- [12] David Balduzzi and Giulio Tononi. Integrated Information in Discrete Dynamical Systems: Motivation and Theoretical Framework. *PLOS Computational Biology*, 4(6):e1000091, 2008. Publisher: Public Library of Science.
- [13] Berardino Barile, Aldo Marzullo, Claudio Stamile, Françoise Durand-Dubief, and Dominique Sappey-Marinier. Data augmentation using generative adversarial neural networks on brain structural connectivity in multiple sclerosis. *Computer Methods and Programs in Biomedicine*, 206:106113, July 2021.
- [14] Pablo Barttfeld, Lynn Uhrig, Jacobo D Sitt, Mariano Sigman, Béchir Jarraya, and Stanislas Dehaene. Signature of consciousness in the dynamics of resting-state brain activity. *Proc. Natl. Acad. Sci.*, page 19, 2015.
- [15] André M Bastos, Jacob A Donoghue, Scott L Brincat, Meredith Mahnke, Jorge Yanar, Josefina Correa, Ayan S Waite, Mikael Lundqvist, Jefferson Roy, Emery N Brown, and Earl K Miller. Neural effects of propofol-induced unconsciousness and its reversal using thalamic stimulation. *eLife*, 10:e60824, April 2021. Publisher: eLife Sciences Publications, Ltd.
- [16] Annabelle M. Belcher, Cecil C. Yen, Haley Stepp, Hong Gu, Hanbing Lu, Yihong Yang, Afonso C. Silva, and Elliot A. Stein. Large-Scale Brain Networks in the Awake, Truly Resting Marmoset Monkey. *Journal of Neuroscience*, 33(42):16796–16804, October 2013. Publisher: Society for Neuroscience Section: Articles.
- [17] A.L. Benabid, P. Pollak, A. Louveau, S. Henry, and J. de Rougemont. Combined (Thalamotomy and Stimulation) Stereotactic Surgery of the VIM Thalamic Nucleus for Bilateral Parkinson Disease. *Applied Neurophysiology*, 50(1-6):344–346, January 1988.
- [18] James L Bernat. Chronic disorders of consciousness. *The Lancet*, 367(9517):1181–1192, April 2006.
- [19] Denis Le Bihan. *Cerveau de cristal*. Odile Jacob, 2012. Cairndomain: www.cairn-sciences.info.

- [20] B. Biswal, F. Z. Yetkin, V. M. Haughton, and J. S. Hyde. Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magnetic Resonance in Medicine*, 34(4):537–541, October 1995.
- [21] Ned Block. On a confusion about a function of consciousness. *Behavioral and Brain Sciences*, 18(2):227–247, June 1995. Publisher: Cambridge University Press.
- [22] Melanie Boly, Anil K. Seth, Melanie Wilke, Paul Ingmundson, Bernard Baars, Steven Laureys, David B. Edelman, and Naotsugu Tsuchiya. Consciousness in humans and non-human animals: recent advances and future directions. *Frontiers in Psychology*, 4:625, October 2013.
- [23] Luke W. Boorman, Samuel S. Harris, Osman Shabir, Llywelyn Lee, Beth Eyre, Clare Howarth, and Jason Berwick. Bidirectional alterations in brain temperature profoundly modulate spatiotemporal neurovascular responses in-vivo. *Communications Biology*, 6(1):1–12, February 2023. Publisher: Nature Publishing Group.
- [24] Pierre Bourdillon, Bertrand Hermann, Jacobo D. Sitt, and Lionel Naccache. Electromagnetic Brain Stimulation in Patients With Disorders of Consciousness. *Frontiers in Neuroscience*, 13, 2019.
- [25] Pierre Boveroux, Audrey Vanhaudenhuyse, Marie-Aurélie Bruno, Quentin Noirhomme, Séverine Lauwick, André Luxen, Christian Degueldre, Alain Plenevaux, Caroline Schnakers, Christophe Phillips, Jean-François Brichant, Vincent Bonhomme, Pierre Maquet, Michael D. Greicius, Steven Laureys, and Mélanie Boly. Breakdown of within- and between-network Resting State Functional Magnetic Resonance Imaging Connectivity during Propofol-induced Loss of Consciousness. *Anesthesiology*, 113(5):1038–1053, November 2010.
- [26] K. Brodmann. Vergleichende Lokalisationslehre der Grosshirnrinde in ihren Prinzipien dargestellt auf Grund des Zellenbaues / [K. Brodmann]., 1909.
- [27] Tyler J. Bruinsma, Vidur V. Sarma, Yoonbae Oh, Dong Pyo Jang, Su-Youne Chang, Greg A. Worrell, Val J. Lowe, Hang Joon Jo, and Hoon-Ki Min. The Relationship Between Dopamine Neurotransmitter Dynamics and the Blood-Oxygen-Level-Dependent (BOLD) Signal: A Review of Pharmacological Functional Magnetic Resonance Imaging. *Frontiers in Neuroscience*, 12:238, April 2018.
- [28] Christopher P. Burgess, Irina Higgins, Arka Pal, Loic Matthey, Nick Watters, Guillaume Desjardins, and Alexander Lerchner. Understanding disentangling in beta-VAE. *arXiv:1804.03599 [cs, stat]*, April 2018. arXiv: 1804.03599.

- [29] Joana Cabral, Diego Vidaurre, Paulo Marques, Ricardo Magalhães, Pedro Silva Moreira, José Miguel Soares, Gustavo Deco, Nuno Sousa, and Morten L. Kringelbach. Cognitive performance in healthy older adults relates to spontaneous switching between states of functional connectivity during rest. *Scientific Reports*, 7(1):5135, July 2017. Number: 1 Publisher: Nature Publishing Group.
- [30] Evan Calabrese, Alexandra Badea, Christopher L. Coe, Gabriele R. Lubach, Yundi Shi, Martin A. Styner, and G. Allan Johnson. A diffusion tensor MRI atlas of the postmortem rhesus macaque brain. *NeuroImage*, 117:408–416, August 2015.
- [31] T. Caliński and J. Harabasz. A dendrite method for cluster analysis. *Communications in Statistics*, 3(1):1–27, 1974.
- [32] Tianqing Cao, Shenghong He, Luchen Wang, Xiaoke Chai, Qiheng He, Dongsheng Liu, Dong Wang, Nan Wang, Jiangong He, Shouyang Wang, Yi Yang, Jizong Zhao, and Huiling Tan. Clinical neuromodulatory effects of deep brain stimulation in disorder of consciousness: A literature review. *CNS Neuroscience & Therapeutics*, n/a(n/a), December 2023. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/cns.14559>.
- [33] Lucrezia Carboni, Michel Dojat, and Sophie Achard. Nodal-statistics-based equivalence relation for graph collections. *Physical Review E*, 107(1):014302, January 2023.
- [34] Mathilde Caron, Piotr Bojanowski, Armand Joulin, and Matthijs Douze. Deep Clustering for Unsupervised Learning of Visual Features. In *ECCV*, pages 139–156, 2018.
- [35] Mathilde Caron, Ishan Misra, Julien Mairal, Priya Goyal, Piotr Bojanowski, and Armand Joulin. Unsupervised Learning of Visual Features by Contrasting Cluster Assignments. *NeurIPS*, 33:9912–9924, 2020.
- [36] Adenauer G. Casali, Olivia Gosseries, Mario Rosanova, Mélanie Boly, Simone Sarasso, Karina R. Casali, Silvia Casarotto, Marie-Aurélié Bruno, Steven Laureys, Giulio Tononi, and Marcello Massimini. A theoretically based index of consciousness independent of sensory processing and behavior. *Science Translational Medicine*, 5(198):198ra105, August 2013.
- [37] Pablo Castro, Andrea Luppi, Enzo Tagliazucchi, Yonatan S-Perl, Lorina Naci, Adrian M Owen, Jacobo D Sitt, Alain Destexhe, and Rodrigo Cofré. Dynamical structure-function correlations provide robust and generalizable signatures of consciousness in humans, December 2023.

- [38] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A Simple Framework for Contrastive Learning of Visual Representations. In *International Conference on Machine Learning*, pages 1597–1607. PMLR, November 2020. ISSN: 2640-3498.
- [39] Herbert M. Chinyanga. Temperature regulation and anesthesia. *Pharmacology & Therapeutics*, 26(2):147–161, January 1984.
- [40] Darko Chudy, Vedran Deletis, Veronika Paradžik, Ivan Dubroja, Petar Marčinković, Darko Orešković, Hana Chudy, and Marina Raguž. Deep brain stimulation in disorders of consciousness: 10 years of a single center experience. *Scientific Reports*, 13(1):19491, November 2023. Number: 1 Publisher: Nature Publishing Group.
- [41] Jan Claassen, Kevin Doyle, Adu Matory, Caroline Couch, Kelly M. Burger, Angela Velazquez, Joshua U. Okonkwo, Jean-Rémi King, Soojin Park, Sachin Agarwal, David Roh, Murad Megjhani, Andrey Eliseyev, E. Sander Connolly, and Benjamin Rohaut. Detection of Brain Activation in Unresponsive Patients with Acute Brain Injury. *The New England Journal of Medicine*, 380(26):2497–2505, June 2019.
- [42] Andy Clark. A case where access implies qualia? *Analysis*, 60(1):30–38, January 2000.
- [43] cnrtl. CONSCIENCE : Définition de CONSCIENCE, 2023.
- [44] Luca Cocchi, Leonardo L. Gollo, Andrew Zalesky, and Michael Breakspear. Criticality in the brain: A synthesis of neurobiology, models and cognition. *Progress in Neurobiology*, 158:132–152, November 2017.
- [45] Harvey R. Colten, Bruce M. Altevogt, and Institute of Medicine (US) Committee on Sleep Medicine and Research. Sleep Physiology. In *Sleep Disorders and Sleep Deprivation: An Unmet Public Health Problem*. National Academies Press (US), 2006.
- [46] Etienne Combrisson and Karim Jerbi. Exceeding chance level by chance: The caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy. *Journal of Neuroscience Methods*, 250:126–136, 2015.
- [47] Cogitate Consortium, Oscar Ferrante, Urszula Gorska-Klimowska, Simon Henin, Rony Hirschhorn, Aya Khalaf, Alex Lepauvre, Ling Liu, David Richter, Yamil Vidal, Niccolò Bonacchi, Tanya Brown, Praveen Sripad, Marcelo Armendariz, Katarina Bendtz, Tara Ghafari, Dorottya Hetenyi, Jay Jeschke, Csaba Kozma, David R. Mazumder, Stephanie Montenegro, Alia Seedat, Abdelrahman Sharafeldin, Shujun Yang, Sylvain Baillet, David J. Chalmers,

- Radoslaw M. Cichy, Francis Fallon, Theofanis I. Panagiotaropoulos, Hal Blumenfeld, Floris P. de Lange, Sasha Devore, Ole Jensen, Gabriel Kreiman, Huan Luo, Melanie Boly, Stanislas Dehaene, Christof Koch, Giulio Tononi, Michael Pitts, Liad Mudrik, and Lucia Melloni. An adversarial collaboration to critically evaluate theories of consciousness, June 2023. Pages: 2023.06.23.546249 Section: New Results.
- [48] John P. Cunningham and Byron M. Yu. Dimensionality reduction for large-scale neural recordings. *Nature Neuroscience*, 17(11):1500–1509, November 2014. Publisher: Nature Publishing Group.
- [49] Kamalaker Dadi, Mehdi Rahim, Alexandre Abraham, Darya Chyzhyk, Michael Milham, Bertrand Thirion, and Gaël Varoquaux. Benchmarking functional connectome-based predictive models for resting-state fMRI. *NeuroImage*, 192:115–134, May 2019.
- [50] Debadatta Dash, Vinayak Abrol, Anil Kumar Sao, and Bharat Biswal. The model order limit: Deep sparse factorization for resting brain. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pages 1244–1247, Washington, DC, April 2018. IEEE.
- [51] Debadatta Dash, Anil Kumar Sao, Jun Wang, and Bharat Biswal. How many fmri scans are necessary and sufficient for resting brain connectivity analysis? In *2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pages 494–498, November 2018.
- [52] Gustavo Deco, Joana Cabral, Mark W Woolrich, Angus B A Stevner, Tim J van Hartevelt, and Morten L Kringelbach. Single or multiple frequency generators in on-going brain activity: A mechanistic whole-brain model of empirical MEG data. *Neuroimage*, 152:538–550, May 2017.
- [53] Gustavo Deco, Josephine Cruzat, Joana Cabral, Enzo Tagliazucchi, Helmut Laufs, Nikos K. Logothetis, and Morten L. Kringelbach. Awakening: Predicting external stimulation to force transitions between different brain states. *Proceedings of the National Academy of Sciences*, 116(36):18088–18097, September 2019.
- [54] Gustavo Deco, Josephine Cruzat, and Morten L. Kringelbach. Brain songs framework used for discovering the relevant timescale of the human brain. *Nature Communications*, 10:583, February 2019.
- [55] Gustavo Deco, Viktor K. Jirsa, Peter A. Robinson, Michael Breakspear, and Karl Friston. The Dynamic Brain: From Spiking Neurons to Neural Masses and Cortical Fields. *PLoS Computational Biology*, 4(8):e1000092, August 2008.

- [56] Gustavo Deco, Morten L Kringelbach, Viktor K Jirsa, and Petra Ritter. The dynamics of resting fluctuations in the brain: metastability and its dynamical cortical core. *Sci. Rep.*, 7(1):3095, June 2017.
- [57] Thomas Deffieux, Charlie Demene, Mathieu Pernot, and Mickael Tanter. Functional ultrasound neuroimaging: a review of the preclinical and clinical state of the art. *Current Opinion in Neurobiology*, 50:128–135, June 2018.
- [58] Stanislas Dehaene and Jean-Pierre Changeux. Experimental and theoretical approaches to conscious processing. *Neuron*, 70(2):200–227, April 2011.
- [59] A. Demertzi, S. Laureys, and M. Boly. Coma, Persistent Vegetative States, and Diminished Consciousness. *Encyclopedia of Consciousness*, page 147, 2009.
- [60] A. Demertzi, E. Tagliazucchi, S. Dehaene, G. Deco, P. Barttfeld, F. Raimondo, C. Martial, D. Fernández-Espejo, B. Rohaut, H. U. Voss, N. D. Schiff, A. M. Owen, S. Laureys, L. Naccache, and J. D. Sitt. Human consciousness is supported by dynamic complex patterns of brain signal coordination. *Science Advances*, 5(2):eaat7603, 2019.
- [61] Murat Demirtaş, Joshua B. Burt, Markus Helmer, Jie Lisa Ji, Brendan D. Adkinson, Matthew F. Glasser, David C. Van Essen, Stamatios N. Sotiropoulos, Alan Anticevic, and John D. Murray. Hierarchical Heterogeneity across Human Cortex Shapes Large-Scale Neural Dynamics. *Neuron*, 101(6):1181–1194.e13, March 2019.
- [62] Daniel C. Dennett. *The Path Not Taken*, 1995. Issue: 2 Number: 2 Pages: 252-253 Volume: 18.
- [63] René Descartes. *L'Homme de René Descartes et un Traitté de la formation du foetus du mesme auteur . Avec les remarques de Louys de La Forge, ... sur le Traitté de l'homme de René Descartes et sur les figures par luy inventées*. C. Angot, Paris, 1664. Country: FR fig. in-4. Publié par Claude Clerselier, et suivi de la version de la préface que M. Schuyl a mise au devant de la version latine qu'il a faite du traité de l'homme.
- [64] Lee R. Dice. Measures of the Amount of Ecologic Association Between Species. *Ecology*, 26(3):297–302, 1945. Publisher: Ecological Society of America.
- [65] Brian L. Edlow, Emi Takahashi, Ona Wu, Thomas Benner, Guangping Dai, Lihong Bu, Patricia Ellen Grant, David M. Greer, Steven M. Greenberg, Hannah C. Kinney, and Rebecca D. Folkerth. Neuroanatomic Connectivity of the Human Ascending Arousal System Critical to Consciousness and Its Disorders. *Journal of Neuropathology & Experimental Neurology*, 71(6):531–546, June 2012.

- [66] Denis A Engemann, Federico Raimondo, Jean-Rémi King, Benjamin Rohaut, Gilles Louppe, Frédéric Faugeras, Jitka Annen, Helena Cassol, Olivia Gosseries, Diego Fernandez-Slezak, Steven Laureys, Lionel Naccache, Stanislas Dehaene, and Jacobo D Sitt. Robust EEG-based cross-site and cross-protocol classification of states of consciousness. *Brain*, 141(11):3179–3192, November 2018.
- [67] John Eraifej, Joana Cabral, Henrique M. Fernandes, Joshua Kahan, Shenghong He, Laura Mancini, John Thornton, Mark White, Tarek Yousry, Ludvic Zrinzo, Harith Akram, Patricia Limousin, Tom Foltynie, Tipu Z. Aziz, Gustavo Deco, Morten Kringelbach, and Alexander L. Green. Modulation of limbic resting-state networks by subthalamic nucleus deep brain stimulation. *Network Neuroscience*, 7(2):478–495, June 2023.
- [68] Kathinka Evers. Neurotechnological assessment of consciousness disorders: five ethical imperatives. *Dialogues in Clinical Neuroscience*, 18(2):155–162, June 2016.
- [69] Xue Fan and Henry Markram. A Brief History of Simulation Neuroscience. *Frontiers in Neuroinformatics*, 13, 2019.
- [70] William Feindel. The physiologist and the neurosurgeon: the enduring influence of Charles Sherrington on the career of Wilder Penfield. *Brain: A Journal of Neurology*, 130(Pt 11):2758–2765, November 2007.
- [71] David Ferrier. The Functions of the Brain. *Journal of Mental Science*, 22(100):598–603, 1876. Publisher: Cambridge University Press.
- [72] Stephen Fleming, Chris Frith, Mel Goodale, Hakwan Lau, Joseph E. LeDoux, Alan L. F. Lee, Matthias Michel, Adrian Owen, Megan A. K. Peters, and Heleen A. Slagter. The Integrated Information Theory of Consciousness as Pseudoscience, September 2023.
- [73] Michael D. Fox, Abraham Z. Snyder, Justin L. Vincent, Maurizio Corbetta, David C. Van Essen, and Marcus E. Raichle. The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proceedings of the National Academy of Sciences of the United States of America*, 102(27):9673–9678, July 2005.
- [74] Frank Freyer, James A Roberts, Robert Becker, Peter A Robinson, Petra Ritter, and Michael Breakspear. Biophysical mechanisms of multistability in resting-state cortical rhythms. *J. Neurosci.*, 31(17):6353–6361, April 2011.
- [75] Frank Freyer, James A Roberts, Petra Ritter, and Michael Breakspear. A canonical model of multistability and scale-invariance in biological systems. *PLoS Comput. Biol.*, 8(8):e1002634, August 2012.

- [76] Garrett Friedman, Katherine W. Turk, and Andrew E. Budson. The Current of Consciousness: Neural Correlates and Clinical Aspects. *Current Neurology and Neuroscience Reports*, 23(7):345–352, July 2023.
- [77] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics (Oxford, England)*, 9(3):432–441, July 2008.
- [78] Chris D. Frith and Geraint Rees. A Brief History of the Scientific Approach to the Study of Consciousness. In *The Blackwell Companion to Consciousness*, pages 1–16. John Wiley & Sons, Ltd, 2017. Section: 1 _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781119132363.ch1>.
- [79] Siyuan Gao, Gal Mishne, and Dustin Scheinost. Nonlinear manifold learning in functional magnetic resonance imaging uncovers a low-dimensional space of brain dynamics. *Human Brain Mapping*, 42(14):4510–4524, October 2021.
- [80] Clément M. Garin, Nachiket A. Nadkarni, Brigitte Landeau, Gaël Chételat, Jean-Luc Picq, Salma Bougacha, and Marc Dhenain. Resting state functional atlas and cerebral networks in mouse lemur primates at 11.7 Tesla. *NeuroImage*, 226:117589, February 2021.
- [81] Joseph T. Giacino, Joy Hirsch, Nicholas Schiff, and Steven Laureys. Functional Neuroimaging Applications for Assessment and Rehabilitation Planning in Patients With Disorders of Consciousness. *Archives of Physical Medicine and Rehabilitation*, 87(12, Supplement):67–76, December 2006.
- [82] Enrico Glerean, Juha Salmi, Juha M. Lahnakoski, Iiro P. Jääskeläinen, and Mikko Sams. Functional Magnetic Resonance Imaging Phase Synchronization as a Measure of Dynamic Functional Connectivity. *Brain Connectivity*, 2(2):91–101, April 2012. Publisher: Mary Ann Liebert, Inc., publishers.
- [83] Gary H. Glover. Overview of Functional Magnetic Resonance Imaging. *Neurosurgery clinics of North America*, 22(2):133–139, April 2011.
- [84] C. Gomez, J. Tasserie, L. Uhrig, B. Jarraya, and A. Grigis. Exploration of the Neural Correlates of Consciousness Using Linear Latent Model. In *2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI)*, pages 1–5, April 2023. ISSN: 1945-8452.
- [85] Chloé Gomez, A. Grigis, L. Uhrig, and B. Jarraya. A generative model of brain states dynamic in non-human primates based on variational auto-encoder, October 2022.

- [86] Chloé Gomez, Lynn Uhrig, Vincent Frouin, Edouard Duchesnay, Béchir Jarraya, and Antoine Grigis. Deep learning models reveal the link between dynamic brain connectivity patterns and states of consciousness, March 2024.
- [87] O Gosseries, A Demertzi, Q Noirhomme, J Tshibanda, R Hustinx, P Maquet, E Salmon, G Moonen, A Luxen, S Laureys, and X De Tiège. Que mesure la neuro-imagerie fonctionnelle :. *Rev Med Liege*, 2008.
- [88] Richard H. Granger and Robert A. Hearn. Models of thalamocortical system. *Scholarpedia*, 2(11):1796, November 2007.
- [89] Antoine Grigis, Chloé Gomez, Vincent Frouin, Lynn Uhrig, and Béchir Jarraya. Interpretable Signature of Consciousness in Resting-State Functional Network Brain Activity. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part I*, pages 261–270, Berlin, Heidelberg, September 2022. Springer-Verlag.
- [90] Antoine Grigis, Chloé Gomez, Vincent Frouin, Lynn Uhrig, and Béchir Jarraya. Revisiting the standard for modeling functional brain network activity: application to consciousness, March 2024.
- [91] Antoine Grigis, Chloé Gomez, Jordy Tasserie, Corentin Ambroise, Vincent Frouin, Béchir Jarraya, and Lynn Uhrig. Predicting Cortical Signatures of Consciousness using Dynamic Functional Connectivity Graph-Convolutional Neural Networks, May 2022. Pages: 2020.05.11.078535 Section: New Results.
- [92] Gerald Hahn, Gorka Zamora-López, Lynn Uhrig, Enzo Tagliazucchi, Helmut Laufs, Dante Mantini, Morten L Kringelbach, Bechir Jarraya, and Gustavo Deco. Signature of consciousness in brain-wide synchronization patterns of monkey and human fmri signals. *NeuroImage*, 226:117470, 2021.
- [93] Yuval Noah Harari. *Sapiens: Une brève histoire de l’humanité*. Albin Michel, September 2015. Google-Books-ID: M89yCgAAQBAJ.
- [94] R. V. L. Hartley. Transmission of Information. *Bell System Technical Journal*, 7(3):535–563, 1928. [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/j.1538-7305.1928.tb01236.x](https://onlinelibrary.wiley.com/doi/pdf/10.1002/j.1538-7305.1928.tb01236.x).
- [95] R. Hassler, G. D. Ore, A. Bricolo, G. Dieckmann, and G. Dolce. EEG and clinical arousal induced by bilateral long-term stimulation of pallidal systems in traumatic vigil coma. *Electroencephalography and Clinical Neurophysiology*, 27(7):689–690, September 1969.

- [96] Lizette Heine, Andrea Soddu, Francisco Gomez, Audrey Vanhauzenhuysse, Luaba Tshibanda, Marie Thonnard, Vanessa Charland-Verville, Murielle Kirsch, Steven Laureys, and Athena Demertzi. Resting State Networks and Consciousness. *Frontiers in Psychology*, 3, 2012.
- [97] Stewart Heitmann and Michael Breakspear. Putting the “dynamic” back into dynamic functional connectivity. *Network Neuroscience*, 2(2):150–174, June 2018.
- [98] Irina Higgins, Loic Matthey, Arka Pal, Christopher Burgess, Xavier Glorot, Matthew Botvinick, Shakir Mohamed, and Alexander Lerchner. beta-VAE: learning basic visual concepts with a constrained variational framework. In *International Conference on Learning Representations 2017*, page 22, 2017.
- [99] R. Hindriks, M. H. Adhikari, Y. Murayama, M. Ganzetti, D. Mantini, N. K. Logothetis, and G. Deco. Can sliding-window correlations reveal dynamic functional connectivity in resting-state fMRI? *NeuroImage*, 127:242–256, February 2016.
- [100] Denis Hoa. *L’IRM pas à pas*. Lulu, Grabels, September 2007.
- [101] Yuki Hori, David J Schaeffer, Kyle M Gilbert, Lauren K Hayrynen, Justine C Cléry, Joseph S Gati, Ravi S Menon, and Stefan Everling. Altered Resting-State Functional Connectivity Between Awake and Isoflurane Anesthetized Marmosets. *Cerebral Cortex*, page bhaa168, June 2020.
- [102] Andreas Horn, Gregor Wenzel, Friederike Irmén, Julius Huebl, Ningfei Li, Wolf-Julian Neumann, Patricia Krause, Georg Bohner, Michael Scheel, and Andrea A Kühn. Deep brain stimulation induced normalization of the human functional connectome in Parkinson’s disease. *Brain*, 142(10):3129–3143, October 2019.
- [103] Anthony G. Hudetz, Xiping Liu, and Siveshigan Pillay. Dynamic Repertoire of Intrinsic Brain States Is Reduced in Propofol-Induced Unconsciousness | Brain Connectivity. *J Brain Connectivity*, 5(1):10–22, 2015.
- [104] Scott Huettel, Allen Song, and Gregory McCarthy. *Functional Magnetic Resonance Imaging, Second Edition*. Sinauer Associates Inc., U.S., December 2008.
- [105] R. Matthew Hutchison, L. Stan Leung, Seyed M. Mirsattari, Joseph S. Gati, Ravi S. Menon, and Stefan Everling. Resting-state networks in the macaque at 7T. *NeuroImage*, 56(3):1546–1555, June 2011.
- [106] R. Matthew Hutchison, Thilo Womelsdorf, Elena A. Allen, Peter A. Bandettini, Vince D. Calhoun, Maurizio Corbetta, Stefania Della Penna, Jeff H.

- Duyn, Gary H. Glover, Javier Gonzalez-Castillo, Daniel A. Handwerker, Shella Keilholz, Vesa Kiviniemi, David A. Leopold, Francesco de Pasquale, Olaf Sporns, Martin Walter, and Catie Chang. Dynamic functional connectivity: promise, issues, and interpretations. *NeuroImage*, 80:360–378, October 2013.
- [107] R. Matthew Hutchison, Thilo Womelsdorf, Joseph S. Gati, Stefan Everling, and Ravi S. Menon. Resting-state networks show dynamic functional connectivity in awake humans and anesthetized macaques. *Human Brain Mapping*, 34(9):2154–2177, 2013. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/hbm.22058>.
- [108] A. Hyvärinen and E. Oja. Independent component analysis: algorithms and applications. *Neural Networks*, 13(4):411–430, June 2000.
- [109] William James. *The principles of psychology, Vol I*. The principles of psychology, Vol I. Henry Holt and Co, New York, NY, US, 1890. Pages: xii, 697.
- [110] Sung Ho Jang, Jong Sun Park, Dong Gu Shin, Seong Ho Kim, and Min Son Kim. Relationship between consciousness and injury of ascending reticular activating system in patients with hypoxic ischaemic brain injury. *Journal of Neurology, Neurosurgery & Psychiatry*, 90(4):493–494, April 2019. Publisher: BMJ Publishing Group Ltd Section: PostScript.
- [111] Mainak Jas, Denis Engemann, Federico Raimondo, Yousra Bekhti, and Alexandre Gramfort. Automated rejection and repair of bad trials in MEG/EEG. In *2016 International Workshop on Pattern Recognition in Neuroimaging (PRNI)*, pages 1–4, June 2016.
- [112] E. R. John, L. S. Pritchep, W. Kox, P. Valdés-Sosa, J. Bosch-Bayard, E. Aubert, M. Tom, F. diMichele, and L. D. Gugino. Invariant Reversible QEEG Effects of Anesthetics. *Consciousness and Cognition*, 10(2):165–183, June 2001.
- [113] E. G. Jones. The Neuron Doctrine 1891. *Journal of the History of the Neurosciences*, 3(1):3–20, January 1994.
- [114] Kathleen Kalmar and Joseph T. Giacino. The JFK Coma Recovery Scale–Revised. *Neuropsychological Rehabilitation*, 15(3-4):454–460, 2005.
- [115] Amrit Kashyap and Shella Keilholz. Dynamic properties of simulated brain network models and empirical resting-state data. *Network Neuroscience (Cambridge, Mass.)*, 3(2):405–426, 2019.

- [116] Kastler, Bruno. *Comprendre l'IRM: manuel d'auto-apprentissage*. Imagerie médicale diagnostic. Elsevier Masson, Issy-les-Moulineaux, 8e édition entièrement révisée édition, 2018.
- [117] Jeremy Kawahara, Colin J. Brown, Steven P. Miller, Brian G. Booth, Vann Chau, Ruth E. Grunau, Jill G. Zwicker, and Ghassan Hamarneh. Brain-netcnn: Convolutional neural networks for brain networks; towards predicting neurodevelopment. *NeuroImage*, 146:1038–1049, 2017.
- [118] Jung-Hoon Kim, Yizhen Zhang, Kuan Han, Zheyu Wen, Minkyu Choi, and Zhongming Liu. Representation learning of resting state fMRI with variational autoencoder. *NeuroImage*, 241:118423, November 2021.
- [119] Diederik P. Kingma and Max Welling. Auto-Encoding Variational Bayes. *arXiv:1312.6114 [cs, stat]*, May 2014. arXiv: 1312.6114.
- [120] H. Klüver and P. C. Bucy. "Psychic blindness" and other symptoms following bilateral temporal lobectomy in Rhesus monkeys. *American Journal of Physiology*, 119:352–353, 1937.
- [121] H. Klüver and P. C. Bucy. Preliminary analysis of functions of the temporal lobes in monkeys. 1939. *The Journal of Neuropsychiatry and Clinical Neurosciences*, 9(4):606–620, 1939.
- [122] Bettina Knie, M. Tanya Mitra, Kartik Logishetty, and K. Ray Chaudhuri. Excessive daytime sleepiness in patients with Parkinson's disease. *CNS drugs*, 25(3):203–212, March 2011.
- [123] Trupti M. Kodinariya and Prashant R. Makwana. Review on determining number of cluster in k-means clustering. *Applied Mathematics and Information Sciences*, 10, 2013.
- [124] Sotiris Kotsiantis. Supervised machine learning: A review of classification techniques. *Informatica*, 31, 10 2007.
- [125] Sid Kouider and Emmanuel Dupoux. Partial Awareness Creates the "Illusion" of Subliminal Semantic Priming. *Psychological Science*, 15(2):75–81, February 2004.
- [126] Elliot Krames, P. Hunter Peckham, and Ali R. Rezai. *Neuromodulation*. Elsevier Academic Press, Amsterdam Boston San Diego, CA, 1st ed edition, 2009.
- [127] Uriah Kriegel. Consciousness: Phenomenal consciousness, access consciousness, and scientific practice. In *Philosophy of Psychology and Cognitive Science*, pages 195–217. North-Holland, January 2007.

- [128] Morten L Kringelbach, Anthony R McIntosh, Petra Ritter, Viktor K Jirsa, and Gustavo Deco. The rediscovery of slowness: Exploring the timing of cognition. *Trends Cogn. Sci.*, 19(10):616–628, October 2015.
- [129] Aaron Kucyi. Just a thought: How mind-wandering is represented in dynamic brain connectivity. *NeuroImage*, 180(Pt B):505–514, October 2018.
- [130] Éditions Larousse. conscience latin conscientia de scire savoir - LAROUSSE, 2023.
- [131] Steven Laureys. The neural correlate of (un)awareness: lessons from the vegetative state. *Trends in Cognitive Sciences*, 2005.
- [132] Steven Laureys, Joseph T. Giacino, Nicholas D. Schiff, Manuel Schabus, and Adrian M. Owen. How should functional imaging of patients with disorders of consciousness contribute to their clinical rehabilitation needs? *Current opinion in neurology*, 19(6):520–527, December 2006.
- [133] Steven Laureys and Giulio Tononi, editors. *The neurology of consciousness: cognitive neuroscience and neuropathology*. Elsevier/Academic Press, Amsterdam ; Boston, 1st ed edition, 2009. OCLC: ocn299125477.
- [134] C. Lecrux and E. Hamel. Neuronal networks and mediators of cortical neurovascular coupling responses in normal and altered brain states. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1705):20150350, October 2016.
- [135] Joseph E. LeDoux, Matthias Michel, and Hakwan Lau. A little history goes a long way toward understanding why we study consciousness the way we do today. *Proceedings of the National Academy of Sciences*, 117(13):6976–6984, March 2020. Publisher: Proceedings of the National Academy of Sciences.
- [136] D. Lehmann, W. K. Strik, B. Henggeler, T. Koenig, and M. Koukkou. Brain electric microstates and momentary conscious mind states as building blocks of spontaneous thinking: I. Visual imagery and abstract thoughts. *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology*, 29(1):1–11, June 1998.
- [137] Francisca P. Leite, Doris Tsao, Wim Vanduffel, Denis Fize, Yuka Sasaki, Larry L. Wald, Anders M. Dale, Ken K. Kwong, Guy A. Orban, Bruce R. Rosen, Roger B. H. Tootell, and Joseph B. Mandeville. Repeated fMRI Using Iron Oxide Contrast Agent in Awake, Behaving Macaques at 3 Tesla. *NeuroImage*, 16(2):283–294, June 2002.

- [138] Nora Leonardi, William R. Shirer, Michael D. Greicius, and Dimitri Van De Ville. Disentangling dynamic networks: Separated and joint expressions of functional connectivity patterns in time. *Human Brain Mapping*, 35(12):5984–5995, 2014. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/hbm.22599>.
- [139] Chao Li, Yiran Wei, and Xi Chen. BrainNetGAN: Data augmentation of brain connectivity using generative adversarial network for dementia classification. *arXiv:2103.08494 [cs, eess]*, March 2021. arXiv: 2103.08494.
- [140] Ssu-Ju Li, Yu-Chun Lo, Hsin-Yi Lai, Sheng-Huang Lin, Hui-Ching Lin, Ting-Chun Lin, Ching-Wen Chang, Ting-Chieh Chen, Christine Chin-Jung Hsieh, Shih-Hung Yang, Feng-Mao Chiu, Chao-Hung Kuo, and You-Yin Chen. Uncovering the Modulatory Interactions of Brain Networks in Cognition with Central Thalamic Deep Brain Stimulation Using Functional Magnetic Resonance Imaging. *Neuroscience*, 440:65–84, August 2020.
- [141] Yin Siang Liaw and George J. Augustine. The claustrum and consciousness: An update. *International Journal of Clinical and Health Psychology*, 23(4), October 2023. Publisher: Elsevier.
- [142] Shuyu Lin, Ronald Clark, Robert Birke, Sandro Schonborn, Niki Trigoni, and Stephen Roberts. Anomaly Detection for Time Series Using VAE-LSTM Hybrid Model. In *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4322–4326, Barcelona, Spain, May 2020. IEEE.
- [143] Maria I. Lioudyno, Alexandra M. Birch, Brian S. Tanaka, Yuri Sokolov, Alan L. Goldin, K. George Chandy, James E. Hall, and Michael T. Alkire. Shaker-Related Potassium Channels in the Central Medial Nucleus of the Thalamus Are Important Molecular Targets for Arousal Suppression by Volatile General Anesthetics. *Journal of Neuroscience*, 33(41):16310–16322, October 2013. Publisher: Society for Neuroscience Section: Articles.
- [144] Ran Liu, Cem Subakan, Aishwarya H. Balwani, Jennifer Whitesell, Julie Harris, Sanmi Koyejo, and Eva Dyer. A generative modeling approach for interpreting population-level variability in brain structure. *bioRxiv*, page 2020.06.04.134635, June 2020. Publisher: Cold Spring Harbor Laboratory Section: New Results.
- [145] Thomas T. Liu, Alican Nalci, and Maryam Falahpour. The global signal in fMRI: Nuisance or Information? *NeuroImage*, 150:213–229, April 2017.
- [146] S. Lloyd. Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28(2):129–137, March 1982. Conference Name: IEEE Transactions on Information Theory.

- [147] N. K. Logothetis, J. Pauls, M. Augath, T. Trinath, and A. Oeltermann. Neurophysiological investigation of the basis of the fMRI signal. *Nature*, 412(6843):150–157, July 2001.
- [148] Elodie Lopes, Ricardo Rego, Manuel Rito, Clara Chamadoira, Duarte Dias, and João Paulo Cunha. Estimation of ANT-DBS Electrodes on Target Positioning Based on a New Percept™ PC LFP Signal Analysis. *Sensors*, 22:6601, September 2022.
- [149] Robin Louiset, Edouard Duchesnay, Antoine Grigis, Benoit Dufumier, and Pietro Gori. SepVAE: a contrastive VAE to separate pathological patterns from healthy ones, July 2023. arXiv:2307.06206 [cs, stat].
- [150] Andrea I. Luppi, Joshua Cain, Lennart R. B. Spindler, Urszula J. Górska, Daniel Toker, Andrew E. Hudson, Emery N. Brown, Michael N. Diringer, Robert D. Stevens, Marcello Massimini, Martin M. Monti, Emmanuel A. Stamatakis, Melanie Boly, and the Curing Coma Campaign and Its Contributing Collaborators. Mechanisms Underlying Disorders of Consciousness: Bridging Gaps to Move Toward an Integrated Translational Science. *Neurocritical Care*, 35(1):37–54, July 2021.
- [151] Andrea I. Luppi, Lynn Uhrig, Jordy Tasserie, Camilo M. Signorelli, Emmanuel A. Stamatakis, Alain Destexhe, Bechir Jarraya, and Rodrigo Cofre. Local orchestration of distributed functional patterns supporting loss and restoration of consciousness in the primate brain. *Nature Communications*, 15(1):2171, March 2024. Publisher: Nature Publishing Group.
- [152] Andrea I. Luppi, Jakub Vohryzek, Morten L. Kringelbach, Pedro A. M. Mediano, Michael M. Craig, Ram Adapa, Robin L. Carhart-Harris, Leor Roseman, Ioannis Pappas, Alexander R. D. Peattie, Anne E. Manktelow, Barbara J. Sahakian, Paola Finoia, Guy B. Williams, Judith Allanson, John D. Pickard, David K. Menon, Selen Atasoy, and Emmanuel A. Stamatakis. Distributed harmonic patterns of structure-function dependence orchestrate human consciousness. *Communications Biology*, 6(1):1–19, January 2023. Publisher: Nature Publishing Group.
- [153] Daniel J. Lurie, Daniel Kessler, Danielle S. Bassett, Richard F. Betzel, Michael Breakspear, Shella Kheilholz, Aaron Kucyi, Raphaël Liégeois, Martin A. Lindquist, Anthony Randal McIntosh, Russell A. Poldrack, James M. Shine, William Hedley Thompson, Natalia Z. Bielczyk, Linda Douw, Dominik Kraft, Robyn L. Miller, Muthuraman Muthuraman, Lorenzo Pasquini, Adeel Razi, Diego Vidaurre, Hua Xie, and Vince D. Calhoun. Questions and controversies in the study of time-varying functional connectivity in resting fMRI. *Network Neuroscience*, 4(1):30–69, February 2020.

- [154] H. Lv, Z. Wang, E. Tong, L.M. Williams, G. Zaharchuk, M. Zeineh, A.N. Goldstein-Piekarski, T.M. Ball, C. Liao, and M. Wintermark. Resting-State Functional MRI: Everything That Nonexperts Have Always Wanted to Know. *AJNR: American Journal of Neuroradiology*, 39(8):1390–1399, August 2018.
- [155] Alex A. MacDonald, Lorina Naci, Penny A. MacDonald, and Adrian M. Owen. Anesthesia and neuroimaging: investigating the neural correlates of unconsciousness. *Trends in Cognitive Sciences*, 19(2):100–107, February 2015. Publisher: Elsevier.
- [156] Lorenzo Magrassi, Giorgio Maggioni, Caterina Pistarini, Carol Di Perri, Stefano Bastianello, Antonio G. Zippo, Giorgio A. Iotti, Gabriele E. M. Biella, and Roberto Imberti. 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, October 2016.
- [157] Dante Mantini, Maurizio Corbetta, Gian Luca Romani, Guy A. Orban, and Wim Vanduffel. Evolutionarily Novel Functional Networks in the Human Brain? *The Journal of Neuroscience*, 33(8):3259–3275, February 2013.
- [158] George A. Mashour, Pieter Roelfsema, Jean-Pierre Changeux, and Stanislas Dehaene. Conscious Processing and the Global Neuronal Workspace Hypothesis. *Neuron*, 105(5):776–798, March 2020.
- [159] Leland McInnes, John Healy, and James Melville. Umap: Uniform manifold approximation and projection for dimension reduction, 2020.
- [160] T. McLardy, F. Ervin, V. Mark, W. Scoville, and W. Sweet. Attempted inset-electrodes-arousal from traumatic coma: neuropathological findings. *Transactions of the American Neurological Association*, 93:25–30, 1968.
- [161] Arthur Mensch, Gael Varoquaux, and Bertrand Thirion. Compressed online dictionary learning for fast resting-state fMRI decomposition. In *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, pages 1282–1285, Prague, Czech Republic, April 2016. IEEE.
- [162] Michael P. Milham, Lei Ai, Bonhwang Koo, Ting Xu, Céline Amiez, Fabien Balezeau, Mark G. Baxter, Erwin L. A. Blezer, Thomas Brochier, Aihua Chen, Paula L. Croxson, Christienne G. Damatac, Stanislas Dehaene, Stefan Everling, Damian A. Fair, Lazar Fleysher, Winrich Freiwald, Sean Froudish-Walsh, Timothy D. Griffiths, Carole Guedj, Fadila Hadj-Bouziane, Suliann Ben Hamed, Noam Harel, Bassem Hiba, Bechir Jarraya, Benjamin Jung, Sabine Kastner, P. Christiaan Klink, Sze Chai Kwok, Kevin N. Laland, David A. Leopold, Patrik Lindenfors, Rogier B. Mars, Ravi S. Menon, Adam Messinger, Martine Meunier, Kelvin Mok, John H. Morrison, Jennifer Nacef, Jamie Nagy, Michael Ortiz Rios, Christopher I. Petkov, Mark

- Pinsk, Colline Poirier, Emmanuel Procyk, Reza Rajimehr, Simon M. Reader, Pieter R. Roelfsema, David A. Rudko, Matthew F. S. Rushworth, Brian E. Russ, Jerome Sallet, Michael Christoph Schmid, Caspar M. Schwiedrzik, Jakob Seidlitz, Julien Sein, Amir Shmuel, Elinor L. Sullivan, Leslie Ungerleider, Alexander Thiele, Orlin S. Todorov, Doris Tsao, Zheng Wang, Charles R. E. Wilson, Essa Yacoub, Frank Q. Ye, Wilbert Zarco, Yong-di Zhou, Daniel S. Margulies, and Charles E. Schroeder. An Open Resource for Non-human Primate Imaging. *Neuron*, 100(1):61–74.e2, October 2018.
- [163] Robyn L. Miller, Maziar Yaesoubi, Jessica A. Turner, Daniel Mathalon, Adrian Preda, Godfrey Pearlson, Tulay Adali, and Vince D. Calhoun. Higher Dimensional Meta-State Analysis Reveals Reduced Resting fMRI Connectivity Dynamism in Schizophrenia Patients. *PLOS ONE*, 11(3):e0149849, March 2016. Publisher: Public Library of Science.
- [164] M. Mishkin and K. H. Pribram. Analysis of the effects of frontal lesions in monkey. I. Variations of delayed alternation. *Journal of Comparative and Physiological Psychology*, 48(6):492–495, December 1955.
- [165] Joyneel Misra, Srinivas Govinda Surampudi, Manasij Venkatesh, Chirag Limbachia, Joseph Jaja, and Luiz Pessoa. Learning brain dynamics for decoding and predicting individual differences. *PLOS Computational Biology*, 17(9):e1008943, September 2021. Publisher: Public Library of Science.
- [166] Fatemeh Mokhtari, Milad I. Akhlaghi, Sean L. Simpson, Guorong Wu, and Paul J. Laurienti. Sliding window correlation analysis: Modulating window shape for dynamic brain connectivity in resting state. *NeuroImage*, 189:655–666, April 2019.
- [167] Martin M. Monti, Audrey Vanhaudenhuyse, Martin R. Coleman, Melanie Boly, John D. Pickard, Luaba Tshibanda, Adrian M. Owen, and Steven Laureys. Willful modulation of brain activity in disorders of consciousness. *The New England Journal of Medicine*, 362(7):579–589, February 2010.
- [168] Ricardo P. Monti, Romy Lorenz, Peter Hellyer, Robert Leech, Christoforos Anagnostopoulos, and Giovanni Montana. Decoding Time-Varying Functional Connectivity Networks via Linear Graph Embedding Methods. *Frontiers in Computational Neuroscience*, 11, 2017.
- [169] Ricardo Pio Monti, Alex Gibberd, Sandipan Roy, Matthew Nunes, Romy Lorenz, Robert Leech, Takeshi Ogawa, Motoaki Kawanabe, and Aapo Hyvärinen. Interpretable brain age prediction using linear latent variable models of functional connectivity. *PLOS ONE*, 15(6):e0232296, June 2020.
- [170] Hervé Morin. Neurosciences : une joute mondiale sur les théories de la conscience. *Le Monde.fr*, October 2023.

- [171] Hermann Munk. OF THE VISUAL AREA OF THE CEREBRAL CORTEX, AND ITS RELATION TO EYE MOVEMENTS *. *Brain*, 13(1):45–70, January 1890.
- [172] Kevin Murphy and Michael D. Fox. Towards a consensus regarding global signal regression for resting state functional connectivity MRI. *NeuroImage*, 154:169–173, July 2017.
- [173] Thikra A. Mustafa and Mohammed A. Mohammed-Rasheed. Accumulation and cytotoxicity assessment of TAT-IONPs on cancerous mammalian cells. *Animal Biotechnology*, 32(1):100–105, February 2021. Publisher: Taylor & Francis _eprint: <https://doi.org/10.1080/10495398.2019.1658595>.
- [174] Lionel Naccache. Why and how access consciousness can account for phenomenal consciousness. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(1755):20170357, July 2018. Publisher: Royal Society.
- [175] Lorina Naci, Leah Sinai, and Adrian M. Owen. Detecting and interpreting conscious experiences in behaviorally non-responsive patients. *NeuroImage*, 145(Pt B):304–313, January 2017.
- [176] National Institute of Neurological Disorders and Stroke. Locked-In Syndrome, January 2023.
- [177] Markus Ojala and Gemma C. Garriga. Permutation Tests for Studying Classifier Performance. In *2009 Ninth IEEE International Conference on Data Mining*, pages 908–913, Miami Beach, FL, USA, December 2009. IEEE.
- [178] George Ojemann, Nick Ramsey, and Jeffrey Ojemann. Relation between functional magnetic resonance imaging (fMRI) and single neuron, local field potential (LFP) and electrocorticography (ECoG) activity in human cortex. *Frontiers in Human Neuroscience*, 7, 2013.
- [179] openAI. New AI classifier for indicating AI-written text, 2023.
- [180] Adrian M. Owen. Disorders of Consciousness. *Annals of the New York Academy of Sciences*, 1124(1):225–238, 2008. _eprint: <https://nyaspubs.onlinelibrary.wiley.com/doi/pdf/10.1196/annals.1440.013>.
- [181] Adrian M. Owen, Martin R. Coleman, Melanie Boly, Matthew H. Davis, Steven Laureys, and John D. Pickard. Detecting awareness in the vegetative state. *Science (New York, N.Y.)*, 313(5792):1402, September 2006.
- [182] Patricia Pais-Roldán, Brian L. Edlow, Yuanyuan Jiang, Johannes Stelzer, Ming Zou, and Xin Yu. Multimodal Assessment of Recovery from Coma in a Rat Model of Diffuse Brainstem Tegmentum Injury. *NeuroImage*, 189:615–630, April 2019.

- [183] B. Panizza. Osservazioni sul nervo ottico. *Gior. I. R. Ist Lomb. Sci. Lett. Arti.* 7, pages 237–252., 1855.
- [184] R. D. Pascual-Marqui, C. M. Michel, and D. Lehmann. Segmentation of brain electrical activity into microstates: model estimation and validation. *IEEE transactions on bio-medical engineering*, 42(7):658–665, July 1995.
- [185] George Paxinos, Xu-Feng Huang, and Arthur Toga. The Rhesus Monkey Brain in Stereotaxic Coordinates. *Faculty of Health and Behavioural Sciences - Papers (Archive)*, January 2000.
- [186] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. *Journal of machine learning research*, 12(Oct):2825–2830, 2011.
- [187] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12(85):2825–2830, 2011.
- [188] Fabian Pedregosa-Izquierdo. *Feature extraction and supervised learning on fMRI : from practice to theory*. PhD thesis, Université Pierre et Marie Curie - Paris VI, February 2015.
- [189] Vassilis Pelekanos, Robert M. Mok, Olivier Joly, Matthew Ainsworth, Diana Kyriazis, Maria G. Kelly, Andrew H. Bell, and Nikolaus Kriegeskorte. Rapid event-related, BOLD fMRI, non-human primates (NHP): choose two out of three. *Scientific Reports*, 10(1):7485, May 2020.
- [190] Wilder Penfield. *Mystery of the Mind: A Critical Study of Consciousness and the Human Brain*. Princeton Legacy Library. Princeton University Press, 1975.
- [191] Yonatan Sanz Perl, Hernán Bocaccio, Ignacio Pérez-Ipiña, Federico Zamberlán, Juan Piccinini, Helmut Laufs, Morten Kringelbach, Gustavo Deco, and Enzo Tagliazucchi. Generative Embeddings of Brain Collective Dynamics Using Variational Autoencoders. *Physical Review Letters*, 125(23):238101, December 2020.
- [192] Yonatan Sanz Perl, Carla Pallavicini, Ignacio Perez Ipiña, Morten Kringelbach, Gustavo Deco, Helmut Laufs, and Enzo Tagliazucchi. Data augmentation based on dynamical systems for the classification of brain states. *Chaos, Solitons & Fractals*, 139:110069, October 2020.

- [193] Yonatan Sanz Perl, Carla Pallavicini, Juan Piccinini, Athena Demertzi, Vincent Bonhomme, Charlotte Martial, Rajanikant Panda, Naji Alnagger, Jitka Annen, Olivia Gosseries, Agustin Ibañez, Helmut Laufs, Jacobo D. Sitt, Viktor K. Jirsa, Morten L. Kringelbach, Steven Laureys, Gustavo Deco, and Enzo Tagliazucchi. Low-dimensional organization of global brain states of reduced consciousness. *Cell Reports*, 42(5):112491, May 2023.
- [194] Fred Plum and Jerome B. Posner. *The Diagnosis of Stupor and Coma*. Oxford University Press, 1982. Google-Books-ID: Pbl4CH4NIQsC.
- [195] Russell A. Poldrack, Jeanette A. Mumford, and Thomas E. Nichols. *Handbook of functional MRI data analysis*. Cambridge University Press, Cambridge New York Melbourne Madrid, 2011. OCLC: 753167009.
- [196] Jonathan D Power, Kelly A Barnes, Abraham Z Snyder, Bradley L Schlaggar, and Steven E Petersen. Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *Neuroimage*, 59(3):2142–2154, February 2012.
- [197] Laurie Pycroft, John Stein, and Tipu Aziz. Deep brain stimulation: An overview of history, methods, and future developments. *Brain and Neuroscience Advances*, 2:239821281881601, December 2018.
- [198] N. Qiang, Q. Dong, Y. Sun, B. Ge, and T. Liu. Deep Variational Autoencoder for Modeling Functional Brain Networks and ADHD Identification. In *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, pages 554–557, April 2020. ISSN: 1945-8452.
- [199] Bo Rao, Dan Xu, Chaoyang Zhao, Shouchao Wang, Xuan Li, Wenbo Sun, Yadong Gang, Jian Fang, and Haibo Xu. Development of functional connectivity within and among the resting-state networks in anesthetized rhesus monkeys. *NeuroImage*, 242:118473, November 2021.
- [200] M Rappaport, K M Hall, K Hopkins, T Belleza, and D N Cope. Disability rating scale for severe head trauma: coma to community. *Archives of physical medicine and rehabilitation*, 63(3):118–123, March 1982.
- [201] J. E. Raymond, K. L. Shapiro, and K. M. Arnell. Temporary suppression of visual processing in an RSVP task: an attentional blink? *Journal of Experimental Psychology. Human Perception and Performance*, 18(3):849–860, August 1992.
- [202] Michelle J. Redinbaugh, Jessica M. Phillips, Niranjana A. Kambi, Sounak Mohanta, Samantha Andryk, Gaven L. Dooley, Mohsen Afrasiabi, Aeyal Raz, and Yuri B. Saalman. Thalamus Modulates Consciousness via Layer-Specific Control of Cortex. *Neuron*, 106(1):66–75.e12, April 2020.

- [203] Jason DM Rennie and Nathan Srebro. Loss functions for preference levels: Regression with discrete ordered labels. In *Proceedings of the IJCAI multi-disciplinary workshop on advances in preference handling*, volume 1. AAAI Press, Menlo Park, CA, 2005.
- [204] robert. conscience - Définitions, synonymes, conjugaison, exemples | Dico en ligne Le Robert, 2023.
- [205] James A. Roberts, Tjeerd W. Boonstra, and Michael Breakspear. The heavy tail of the human brain. *Current Opinion in Neurobiology*, 31:164–172, April 2015.
- [206] Peter Rousseeuw. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20:53–65, 1987.
- [207] Kadharbatcha S. Saleem and Nikos K. Logothetis. *A Combined MRI and Histology Atlas of the Rhesus Monkey Brain in Stereotaxic Coordinates*. Academic Press, September 2012.
- [208] Ilaria Sani, Brent C McPherson, Heiko Stemmann, Franco Pestilli, and Winrich A Freiwald. Functionally defined white matter of the macaque monkey brain reveals a dorso-ventral attention network. *eLife*, 8:e40520, January 2019. Publisher: eLife Sciences Publications, Ltd.
- [209] Antonis D. Savva, Michalis Kassinopoulos, Nikolaos Smyrnis, George K. Matsopoulos, and Georgios D. Mitsis. Effects of motion related outliers in dynamic functional connectivity using the sliding window method. *Journal of Neuroscience Methods*, 330:108519, January 2020.
- [210] N. D. Schiff, D. Rodriguez-Moreno, A. Kamal, K. H.S. Kim, J. T. Giacino, F. Plum, and J. Hirsch. fMRI reveals large-scale network activation in minimally conscious patients. *Neurology*, 64(3):514–523, February 2005. Publisher: Wolters Kluwer.
- [211] Nicholas D. Schiff. Multimodal Neuroimaging Approaches to Disorders of Consciousness. *The Journal of Head Trauma Rehabilitation*, 21(5):388, October 2006.
- [212] Nicholas D. Schiff. Central Thalamic Contributions to Arousal Regulation and Neurological Disorders of Consciousness. *Annals of the New York Academy of Sciences*, 1129(1):105–118, 2008. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1196/annals.1417.029>.
- [213] Nicholas D. Schiff. Recovery of consciousness after brain injury: a mesocircuit hypothesis. *Trends in Neurosciences*, 33(1):1–9, January 2010.

- [214] Nicholas D. Schiff. Central thalamic deep brain stimulation to support anterior forebrain mesocircuit function in the severely injured brain. *Journal of Neural Transmission*, 123(7):797–806, July 2016.
- [215] Steffen Schneider, Jin Hwa Lee, and Mackenzie Weygandt Mathis. Learnable latent embeddings for joint behavioural and neural analysis. *Nature*, 617(7960):360–368, May 2023. Number: 7960 Publisher: Nature Publishing Group.
- [216] Hartmut Schulz. Rethinking Sleep Analysis. *Journal of Clinical Sleep Medicine*, 04(02):99–103, April 2008. Publisher: American Academy of Sleep Medicine.
- [217] Lucas Seninge, Ioannis Anastopoulos, Hongxu Ding, and Joshua Stuart. VEGA is an interpretable generative model for inferring biological network activity in single-cell transcriptomics. *Nature Communications*, 12(1):5684, September 2021. Number: 1 Publisher: Nature Publishing Group.
- [218] National Health Service. Disorders of consciousness, May 2022. Section: conditions.
- [219] Anil K. Seth and Tim Bayne. Theories of consciousness. *Nature Reviews Neuroscience*, 23(7):439–452, July 2022. Number: 7 Publisher: Nature Publishing Group.
- [220] Kristen A. Severson, Soumya Ghosh, and Kenney Ng. Unsupervised Learning with Contrastive Latent Variable Models. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(01):4862–4869, July 2019. Number: 01.
- [221] Sadia Shakil, Chin-Hui Lee, and Shella Dawn Keilholz. Evaluation of sliding window correlation performance for characterizing dynamic functional connectivity and brain states. *NeuroImage*, 133:111–128, June 2016.
- [222] Claude Shannon and Warren Weaver. The Mathematical Theory of Communication. *Urbana: University of Illinois Press*, 1949.
- [223] Kimron L. Shapiro, Jane Raymond, and Karen Arnell. Attentional blink. *Scholarpedia*, 4(6):3320, June 2009.
- [224] Kelly Shen, Gleb Bezgin, R. Matthew Hutchison, Joseph S. Gati, Ravi S. Menon, Stefan Everling, and Anthony R. McIntosh. Information Processing Architecture of Functionally Defined Clusters in the Macaque Cortex. *Journal of Neuroscience*, 32(48):17465–17476, November 2012. Publisher: Society for Neuroscience Section: Articles.
- [225] Kelly Shen, Gleb Bezgin, Michael Schirner, Petra Ritter, Stefan Everling, and Anthony R. McIntosh. A macaque connectome for large-scale network simulations in TheVirtualBrain. *Scientific Data*, 6(1):123, July 2019.

- [226] Camilo Miguel Signorelli, Lynn Uhrig, Morten Kringelbach, Bechir Jarraya, and Gustavo Deco. Hierarchical disruption in the cortex of anesthetized monkeys as a new signature of consciousness loss. *NeuroImage*, 227:117618, December 2020.
- [227] Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. Deep inside convolutional networks: Visualising image classification models and saliency maps, 2014.
- [228] Jacobo Diego Sitt, Jean-Remi King, Imen El Karoui, Benjamin Rohaut, Frederic Faugeras, Alexandre Gramfort, Laurent Cohen, Mariano Sigman, Stanislas Dehaene, and Lionel Naccache. Large scale screening of neural signatures of consciousness in patients in a vegetative or minimally conscious state. *Brain: A Journal of Neurology*, 137(Pt 8):2258–2270, August 2014.
- [229] Perry Sprawls. *Magnetic Resonance Imaging*. Medical Physics Publishing, 2000.
- [230] Jost Tobias Springenberg, Alexey Dosovitskiy, Thomas Brox, and Martin Riedmiller. Striving for simplicity: The all convolutional net, 2015.
- [231] Edited by Leanne Stevens. *Introduction to Psychology & Neuroscience*. Dalhousie University Libraries Digital Editions, September 2020.
- [232] J. Tasserie. *Functional neuro-imaging study of Deep Brain Stimulation mechanisms for the restoration of consciousness using a Non-Human Primate model*. PhD thesis, Université Paris-saclay, December 2020.
- [233] Jordy Tasserie, Antoine Grigis, Lynn Uhrig, Morgan Dupont, Alexis Amadon, and Béchir Jarraya. Pypreclin: An automatic pipeline for macaque functional MRI preprocessing. *NeuroImage*, 207:116353, February 2020.
- [234] Jordy Tasserie, Lynn Uhrig, Jacobo D. Sitt, Dragana Manasova, Morgan Dupont, Stanislas Dehaene, and Béchir Jarraya. Deep brain stimulation of the thalamus restores signatures of consciousness in a nonhuman primate model. *Science Advances*, 8(11):eabl5547, March 2022.
- [235] G. Teasdale and B. Jennett. Assessment of coma and impaired consciousness. A practical scale. *Lancet (London, England)*, 2(7872):81–84, July 1974.
- [236] Aurore Thibaut, Nicholas Schiff, Joseph Giacino, Steven Laureys, and Olivia Gosseries. Therapeutic interventions in patients with prolonged disorders of consciousness. *The Lancet Neurology*, 18(6):600–614, June 2019. Publisher: Elsevier.

- [237] Michael E. Tipping and Christopher M. Bishop. Probabilistic Principal Component Analysis. *Journal of the Royal Statistical Society Series B*, 61(3):611–622, 1999. Publisher: Royal Statistical Society.
- [238] Giulio Tononi. An information integration theory of consciousness. *BMC Neuroscience*, 5(1):42, November 2004.
- [239] Giulio Tononi. Consciousness as integrated information: a provisional manifesto. *The Biological Bulletin*, 215(3):216–242, December 2008.
- [240] Giulio Tononi, Melanie Boly, Marcello Massimini, and Christof Koch. Integrated information theory: from consciousness to its physical substrate. *Nature Reviews. Neuroscience*, 17(7):450–461, July 2016.
- [241] Nicolas Tsapis. Agents de contraste pour l'imagerie médicale - Les exemples de l'IRM et de l'ultrasonographie. *médecine/sciences*, 33(1):18–24, January 2017. Number: 1 Publisher: Éditions EDK, Groupe EDP Sciences.
- [242] Julie Tseng and Jordan Poppenk. Brain meta-state transitions demarcate thoughts across task contexts exposing the mental noise of trait neuroticism. *Nature Communications*, 11(1):3480, July 2020. Number: 1 Publisher: Nature Publishing Group.
- [243] L. Uhrig, S. Dehaene, and B. Jarraya. A Hierarchy of Responses to Auditory Regularities in the Macaque Brain. *Journal of Neuroscience*, 34(4):1127–1132, January 2014.
- [244] Lynn Uhrig, David Janssen, Stanislas Dehaene, and Béchir Jarraya. Cerebral responses to local and global auditory novelty under general anesthesia. *NeuroImage*, 141:326–340, November 2016.
- [245] Lynn Uhrig, Jacobo D. Sitt, Amaury Jacob, Jordy Tasserie, Pablo Barttfeld, Morgan Dupont, Stanislas Dehaene, and Bechir Jarraya. Resting-state Dynamics as a Cortical Signature of Anesthesia in Monkeys. *Anesthesiology*, 129(5):942–958, November 2018.
- [246] Koene R.A. Van Dijk, Mert R. Sabuncu, and Randy L. Buckner. The Influence of Head Motion on Intrinsic Functional Connectivity MRI. *NeuroImage*, 59(1):431–438, January 2012.
- [247] Nees Jan van Eck and Ludo Waltman. Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, 84(2):523–538, August 2010.
- [248] Da Wang, Hui Li, Mengyang Xu, Binshi Bo, Mengchao Pei, Zhifeng Liang, and Garth J. Thompson. Differential Effect of Global Signal Regression Between Awake and Anesthetized Conditions in Mice. *Brain Connectivity*, December 2023. Publisher: Mary Ann Liebert, Inc., publishers.

- [249] Peng Wang, Ru Kong, Xiaolu Kong, Raphaël Liégeois, Csaba Orban, Gustavo Deco, Martijn P. van den Heuvel, and B.T. Thomas Yeo. Inversion of a large-scale circuit model reveals a cortical hierarchy in the dynamic resting human brain. *Science Advances*, 5(1):eaat7854, January 2019.
- [250] Elizabeth K. Warrington and L. Weiskrantz. New Method of Testing Long-term Retention with Special Reference to Amnesic Patients. *Nature*, 217(5132):972–974, March 1968. Number: 5132 Publisher: Nature Publishing Group.
- [251] John B. Watson. Psychology as the behaviorist views it. *Psychological Review*, 20(2):158–177, 1913. Place: US Publisher: Psychological Review Company.
- [252] L Weiskrantz and E. K. Warrington. Blindsight – residual vision following occipital lesions in man and monkey. *Brain Research*, pages 85: 1, 184–5, 1975.
- [253] Essa Yacoub, Mark D. Grier, Edward J. Auerbach, Russell L. Lagore, Noam Harel, Gregor Adriany, Anna Zilverstand, Benjamin Y. Hayden, Sarah R. Heilbronner, Kamil Uğurbil, and Jan Zimmermann. Ultra-high field (10.5 T) resting state fMRI in the macaque. *NeuroImage*, 223:117349, December 2020.
- [254] Andrew Zalesky, Alex Fornito, Luca Cocchi, Leonardo L. Gollo, and Michael Breakspear. Time-resolved resting-state brain networks. *Proceedings of the National Academy of Sciences of the United States of America*, 111(28):10341, July 2014. Publisher: National Academy of Sciences.
- [255] E. Zarahn, G. K. Aguirre, and M. D’Esposito. Empirical Analyses of BOLD fMRI Statistics. *NeuroImage*, 5(3):179–197, April 1997.
- [256] Jure Zbontar, Li Jing, Ishan Misra, Yann LeCun, and Stéphane Deny. Barlow twins: Self-supervised learning via redundancy reduction, 2021.
- [257] Xiaohang Zhan, Jiahao Xie, Ziwei Liu, Yew Soon Ong, and Chen Change Loy. Online deep clustering for unsupervised representation learning, 2020.
- [258] Ruyuan Zhang, Stephen A. Engel, and Kendrick Kay. Binocular Rivalry: A Window into Cortical Competition and Suppression. *Journal of the Indian Institute of Science*, 97(4):477–485, December 2017.
- [259] Qingyu Zhao, Ehsan Adeli, Nicolas Honnorat, Tuo Leng, and Kilian M. Pohl. Variational AutoEncoder For Regression: Application to Brain Aging Analysis. *arXiv:1904.05948 [cs, stat]*, July 2019. arXiv: 1904.05948.
- [260] Christelle Zielinski, Deirdre Bolger, and Valérie Chanoine. Le signal en EEG et MEG, February 2014.