Whole-brain models and consciousness: a novel framework for a mechanistic understanding of different states of consciousness

Tesis entregada a

LA UNIVERSIDAD DE VALPARAÍSO

en Cumplimiento Parcial de los requisitos para optar al grado de
Doctor en Ciencias con Mención en Biofísica y Biología Computacional

Facultad de Ciencias

Por:
RUBÉN HERZOG AMUNÁTEGUI
AGOSTO 2022

Dirigida por: Dr. Rodrigo Cofré Torres

Co-Dirigida por: Dr. Patricio Orio Álvarez

Universidad de Valparaíso
CHILE
Doctorado en Ciencias mención
Biofísica y Biología Computacional

Facultad de Ciencias

FACULTAD DE CIENCIAS
UNIVERSIDAD DE VALPARAÍSO

INFORME DE APROBACION
TESIS DE DOCTORADO

Se informa a la Facultad de Ciencias que la Tesis de Doctorado presentada por:

RUBÉN HERZOG AMUNÁTEGUI

Ha sido aprobada por la comisión de Evaluación de la tesis como requisito para optar al grado de Doctor en Ciencias con mención en Biofísica y Biología Computacional, en el examen de Defensa de Tesis rendido el día 22 del Mes de AGOSTO de 2022

Director/a de Tesis:

Dr/a. **RODRIGO COFRÉ TORRES**

Co-Director/a de Tesis:

Dr/a. **PATRICIO ORIO ÁLVAREZ**

Comisión de Evaluación de la Tesis

Dr/a.
Evaluador/a interno/a **WAEL EL-DEREDY**

Dr/a.
Evaluador/a in/externo/a **JORGE SILVA**

Dr/a.
Evaluador/a in/externo/a **NICOLÁS CROSSLEY**

DEDICATORIA

Esta tesis está dedicada a mi padre, Ignacio Roberto Herzog Sugar, a mi madre, María Jesús Amunátegui Ortíz y a mi hermana, Lea Herzog Amunátegui. Les agradezco su esfuerzo, confianza, apoyo, incentivos y formación.

AGRADECIMIENTOS

En esta sección quiero agradecer a quienes fueron fundamentales en el desarrollo de esta tesis. Agradezco a Samy Castro Novoa por su amistad, inspiración, compañerismo e interminables discusiones sobre variados tópicos, sin las cuales ésta tesis –y su autor– no serían lo que son. A Camila Estay Olmos, por ser parte clave en la decisión de dedicar mi tesis doctoral a la investigación de drogas psicodélicas. Le agradezco por su complicidad, su entrega y por las incontables noches retroalimentándonos respecto a nuestras decisiones científicas y vitales. Nuestro encuentro fue un punto de giro en mi vida y en mi trayectoria académica. A Fernanda Medina Badilla por su amistad, su entrega afectiva, su rigurosa agudeza filosófica, por haberme hospedado durante el periódo de defensa de tesis y por haber sido inspiración clave para entrar a un programa de doctorado. A Tamara Bernal Vilches por su amistad, apoyo y compañerismo científico, y por la fuerza que inspira. A mis actuales compañeros de casa en Bettina 63, Fernando Pineda y Francisco Ibarra, por su amistad y apoyo cotidiano. Les agradezco por compartir un cotidiano seguro, en confianza, abierto, relajado y libre de juicios. A Milagros Moncada Rojas, por su amor, apoyo, confianza y por su profunda sabiduría de hacer de lo complejo algo simple. A las amistades de la vida, que hacen parte del lugar del cual éste trabajo emerge: María Francisca Soto, Chagall Acevedo, Drago Ljubetic, Diego Aguayo, Matías Riquelme, Diego Pinto, Diego Álvarez, Diego Soto, Carolina Lara, Cristían Lizama y Jazmín Peña, Denisse Invierno, Miguel Correa, Victor Veloso, Franco, Fanny Raiz, Crispín y Nebu.

A Rodrigo Cofré Torres por su impecable guía, apoyo y sabiduría. Esta tesis relfeja también su excelente y aventurado desempeño como científico y como director. Al equipo de colaboradores, quienes participaron activamente de ésta tesis: Fernando Rosas, Pedro Mediano, Enzo Tagliazucchi, Yonatan Sanz, Juan Piccini, Andrea Luppi, Paul Lodder, Gustavo Deco y Robin Carhart-Harris. A Patricio Orio por haberme guiado en el proceso de entrar al doctorado, la disposición y la confianza durante mi paso por el doctorado. A la Fundación para Estudios de la Conciencia Humana (EcoH) por ser la familia epistémica y psiconautica con la cual proyectar la investigación y aplicación de diversos estados de conciencia para el bienestar y desarrollo humano.

A ANID por la Beca de Doctorado de 4 años para desarrollar la tesis. A Banco Santander por la beca para realizar una pasantía corta en el laboratorio de Enzo Tagliazucchi en la Universidad de Buenos Aires, Argentina.

RESUMEN

El alcance de la conciencia humana incluye estados que se alejan de lo que la mayoría de nosotros conocemos como vigilia ordinaria. Estos estados alterados de conciencia constituyen una excelente oportunidad para estudiar cómo cambios globales en la actividad cerebral se relacionan con diferentes variedades de experiencias subjetivas. Consideramos el problema de explicar cómo indicios globales de los estados alterados de conciencia emergen de la interacción entre la conectividad a gran escala y reglas dinámicas de bajo nivel del cerebro. Para este propósito, proponemos cerrar la brecha entre modelos generativos de la actividad del cerebro completo y los indicios globales propuestos por teorías de la conciencia. Entre los modelos actuales de cerebro completo, el modelo dynamic mean field (DMF, modelo dinámico de campo medio) es particularmente atractivo, ya que combina modelos realistas de neuronas únicas escalados a través de la aproximación de campo medio y datos multimodales de neuroimagen. A pesar de estas características favorables, una barrera importante para el uso masivo del modelo DMF es que sus implementaciones actuales son computacionalmente muy costosas – al punto que el modelo se vuelve inviable sin una infraestructura computacional de alto rendimiento. Es más, incluso con una infraestructura adecuada, las implementaciones actuales solo soportan simulaciones en parcelaciones del cerebro con menos de 100 regiones. Para remover estas barreras, introducimos una implementación del DMF amistosa para el usuario y computacionalmente eficiente, que llamamos *FastDMF*. Por medio de un conjunto de avances analíticos y numéricos –incluyendo un novedoso estimador del control inhibitorio y un algoritmo de optimización bayesiana– el *FastDMF* elude varios cuellos de botella computacionales de las implementaciones previas. En este trabajo usamos el *FastDMF* para proponer una explicación mecanicista del aumento en la entropía neural gatillado por drogas psicodélicas serotoninérgicas (agonistas del receptor 5HT2A), conocido como la

hipótesis del cerebro entrópico. Nuestros resultados reproducen el aumento global de entropía observado en experimentos *in vivo*, entregando la primera explicación basada en modelos para éste fenómeno. También encontramos que los cambios no son uniformes a lo largo del cerebro: la entropía aumentó en todas las regiones, pero el efecto mayor se localizó en regiones visuo-occipitales, tal como se ha observado en estudios de neuroimagen. Interesantemente, a nivel de cerebro completo, esta reconfiguración no se explica bien por la densidad de receptores 5HT2A, sino que se relaciona estrechamente a propiedades topológicas de la conectividad anatómica. Estos resultados ayudan a entender los mecanismos subyacentes al estado psicodélico desde reglas biofísicas de bajo nivel y, más generalmente, a entender la modulación farmacológica de los indicios globales de la conciencia.

ABSTRACT

The scope of human consciousness includes states departing from what most of us experience as ordinary wakefulness. These altered states of consciousness constitute a prime opportunity to study how global changes in brain activity relate to different varieties of subjective experience. We consider the problem of explaining how global signatures of altered consciousness arise from the interplay between large-scale connectivity and low-level dynamical rules. For this purpose, we propose to bridge the gap between bottom-up generative models of whole-brain activity and the top-down signatures proposed by theories of consciousness. Among the current whole-brain models, the dynamic mean field (DMF) model is a particularly attractive model, combining a biophysically realistic single-neuron model that is scaled up via a mean-field approach and multimodal imaging data. Despite these favourable features, an important barrier for a widespread usage of the DMF model is that current implementations are computationally expensive – to the extent that the model often becomes unfeasible when no high-performance computing infrastructure is available. Furthermore, even when such computing structure is available, current implementations can only support simulations on brain parcellations that consider less than 100 brain regions. To remove these barriers, here we introduce a user-friendly and computationally-efficient implementation of the DMF model, which we call *FastDMF*. By leveraging a suit of analytical and numerical advances – including a novel estimation of the feedback inhibition control parameter, and a Bayesian optimisation algorithm – the *FastDMF* circumvents various computational bottlenecks of previous implementations. Here, we use the FastDMF to propose a mechanistic explanation of the increase in neural entropy elicited by serotonergic psychedelics drugs (agonists of the 5-HT2A receptor), also know as the entropic brain hypothesis. Our results reproduce the overall entropy increase observed in previous experiments *in vivo*, providing the first model-based explana-

tion for this phenomenon. We also found that entropy changes were not uniform across the brain: entropy increased in all regions, but the larger effect were localised in visuo-occipital regions, as previously shown in neuroimaging studies. Interestingly, at the whole-brain level, this reconfiguration was not well explained by 5-HT2A receptor density, but related closely to the topological properties of the brain's anatomical connectivity. These results help us understand the mechanisms underlying the psychedelic state from the low-level biophysical rules and, more generally, the pharmacological modulation of top-down signatures of consciousness.

Contents

List of Figures

List of Tables

INTRODUCTION

3.1 Consciousness and the brain: linking biophysics with lived experience

Consciousness has been a puzzle beyond the scope of natural science for centuries; however, the significant progress seen during the last 30 years of research suggests that a rigorous scientific understanding of consciousness is possible (LeDoux et al., 2020; Seth, 2018; Overgaard, 2017). The dawn of the modern neuroscientific approach to consciousness can be traced back to Crick and Koch's proposal for identifying the neural correlates of consciousness (NCC) (Crick and Koch, 1990, 2003), understood as the minimal set of neural events associated with a certain subjective experience. The key intuition that fuels this proposal is that careful experimentation should suffice to reveal brain events that are systematically associated with conscious (as opposed to unconscious or subliminal) perception. Needless to say, the methodological challenges associated with this idea are vast – particularly concerning the determination of what constitutes conscious content (e.g., must content be explicitly reported, or are other less direct forms of inference equally valid (Tsuchiya et al., 2015; Cohen and Dennett, 2011))? Despite these problems, which are still actively debated, the program put forward by Crick and Koch succeeded to jump-start contemporary consciousness research. For recent reviews on the empirical search for NCC, see Ref. (Koch et al., 2016); for a theoretical examination of the concept of NCC, see Ref. (Chalmers, 2000); and for criticism to the concept of NCC, see Refs. (Noë and Thompson, 2004). While the quest for the NCC aims to provide answers to where and when consciousness occurs in the brain, subsequent theoretical efforts have attempted to discover systematic signatures within those NCC that could reflect key mechanisms underlying the emergence of consciousness. In other

words, these efforts try to answer how consciousness emerges from the processes that give rise to the NCC (Seth, 2007; Sergent and Naccache, 2012). Hence, theoretical models of consciousness strive to "compress" our empirical knowledge of the NCC, i.e., to provide rules that can predict when and where from how. The nature of those rules, in turn, determines the kind of explanation offered by a theoretical model of consciousness. Here, we consider two possible approaches: top-down and bottom-up (Stinson and Sullivan, 2018). On the one hand, top-down approaches start by identifying high-level signatures of consciousness, and then try to narrow down low-level biophysical mechanisms compatible with those signatures. On the other hand, bottom-up approaches build from dynamical rules of elementary units (such as neurons or groups of neurons (Deco et al., 2008)) and attempt to provide quantitative predictions by exploring the aggregated consequences of these rules across various temporal and spatial scales. We further subdivide explanations into those addressing conscious information access (e.g., perception in different sensory modalities) and those concerning consciousness as a temporally extended state, such as wakefulness, sleep, anaesthesia, and the altered states that can be elicited by pharmacological manipulation (Vaitl et al., 2005; Revonsuo et al., 2009; Overgaard and Overgaard, 2010; Tassi and Muzet, 2001; Ludwig, 1966; Tart, 1976; Bayne, 2007). Theories that rely heavily on a top-down perspective risk being under-determined in the reductive sense, i.e., they could be compatible with multiple and potentially divergent lower-level biological and physical mechanisms (Michel, 2019). While we do not know whether consciousness may be instantiated in other physical systems, we certainly do know that it is instantiated in the human brain, and therefore all theoretical models of consciousness should be consistent with the low-level biophysical details of the brain to be considered acceptable. In light of this potential under-determination, it is difficult to decide whether the different theories currently dominating the field are competing (in the sense of predicting mutually contradictory empirical findings) or convergent (in spite of being formulated from disparate perspectives). Without investigating theories of consciousness from the bottom-up, it could be simply too early for proposals of an experimentum crucis to decide between candidates (Reardon, 2019).

3.2 What is an altered state of consciousness? Examples and defining features

A basic distinction is commonly drawn between phenomenal and access consciousness (Block, 1995). The first represents the subjective experience of sensory perception, emotion, thoughts, etc.; in other words, what it feels like to perceive something, undergo a certain emotion, or engage in a certain thought process. The second represents the global availability of conscious content for cognitive functions, such as speech, reasoning, and decision-making, enabling the capacity to issue first-person reports.

The term "consciousness" is also used in reference to a third concept in which the definition is comparatively more elusive: that of temporally extended and qualitatively distinct modes or states of consciousness (Vaitl et al., 2005; Revonsuo et al., 2009; Overgaard and Overgaard, 2010; Tassi and Muzet, 2001; Ludwig, 1966; Tart, 1976; Bayne, 2007). This concept is perhaps best introduced by listing examples, such as our ordinary state of conscious wakefulness, the different phases of the wake-sleep cycle, dreaming during rapid eye movement (REM) sleep, sedation and general anaesthesia, post-comatose disorders, such as the unresponsive wakefulness syndrome, the acute effects of certain drugs (mainly serotonergic psychedelics and glutamatergic dissociatives), the state achieved in some contemplative traditions by means of meditation, hypnosis, and shamanic trance, among others. Following Ludwig (Ludwig, 1966) and Tart (Tart, 1972), we refer to these as "altered states of consciousness", adopting this term to emphasize their dissimilarity to ordinary conscious wakefulness. Altered states of consciousness have been studied for decades from different perspectives (Natsoulas, 1981), emphasizing the individual differences in conscious experiencing. Basic processes, such as sensory perception, reveal consistent and substantial inter-individual differences (Deutsch, 1986). Other inter-individual differences in experiences of imagery and thought in the waking state, dreaming, hypnosis, and other phenomena have been reported (Kunzendorf and Wallace, 2000). Furthermore, the same phenomenal event may be interpreted in different ways, evidencing cultural variations (Pasricha and Stevenson, 1986). For a complete account of altered states of consciousness following a multidisciplinary perspective, we refer the reader to Cardeña and Winkelman (2011).

Let us describe commonalities shared by altered states of consciousness, which point towards a potential general operational definition. First, altered states of con-

sciousness are temporally extended and typically (but not always) reversible. Second, they are not defined by the presence of specific subjective experiences, but instead by general and qualitative modifications to the contents of consciousness, including their experienced intensity (Revonsuo et al., 2009). Third, at least some states can be ordered along a hierarchy of levels, from states of "reduced" consciousness (e.g., general anaesthesia, sleep) to others considered "richer" (e.g., certain states achieved during meditation or induced by pharmacological means) (Carhart-Harris et al., 2014).

A proper definition of what constitutes an altered state of consciousness is, unfortunately, more elusive than suggested by the examination of these examples. If states of consciousness are transient, then what is their minimum accepted length? Do qualitative modifications of conscious content apply only to the sensory domain, or encompass other forms of subjective experience, as well? Does a *déjà-vu* (a brief episode of eerie familiarity with an unknown past event) qualify as an altered state of consciousness? What about an orgasm, or the state of pain caused by hitting one's finger with a hammer? Without doubt, these examples modify in one way or another the general contents of consciousness, but they are not commonly considered as altered states of consciousness.

The intuitive notion of "levels" of consciousness is also problematic (Bayne et al., 2016). We are familiar with the fact that some states appear to be "more conscious" than others; for instance, ordinary wakefulness would have a higher conscious level than deep sleep or an absence seizure. But in what sense is deep sleep more or less conscious than an absence seizure? Following this logic, how should dreaming, the acute effects of psychedelic drugs, and the state achieved by expert meditators be ordered along a hypothetical uni-dimensional hierarchy of levels of consciousness? It seems that altered states of consciousness can only be subject to partial ordering, with comparisons between certain pairs of states being questionable or outright meaningless.

These difficulties relate to two main problems. The first problem is granularity: how long is long enough to qualify as an altered state of consciousness? The second is compositeness: instead of a single level of intensity, multiple dimensions are likely required for an unambiguous characterization; however, it is unclear how many dimensions are needed and how they should be determined (Bayne et al., 2016; Bayne and Carter, 2018). A subsidiary issue related to the granularity problem is whether altered states of consciousness represent discrete states with sharply defined boundaries, or are more adequately understood as continuous transitions.

Several proposals have been put forward to circumvent these issues and define altered states of consciousness (Vaitl et al., 2005; Revonsuo et al., 2009; Overgaard and Overgaard, 2010; Tassi and Muzet, 2001; Ludwig, 1966; Tart, 1976; Bayne, 2007). Here, we adopt perforce a more pragmatic stance: we are interested in altered states of consciousness lasting enough to be investigated by modern neuroimaging techniques (>10 min). At the same time, we strive to show that whole-brain models can be sufficiently rich to transcend the unidimensional characterization of consciousness in terms of "levels".

For the purposes of this article, we divide altered states of consciousness into the following (neither exhaustive nor mutually exclusive) categories: natural or endogenous (e.g., the states within the sleep cycle), induced by pharmacological means (e.g., general anaesthesia, the psychedelic state), induced by other means (e.g., meditation, hypnosis), caused by pathological processes, either neurological or psychiatric (e.g., disorders of consciousness, epilepsy, psychotic episodes), and transitory versus permanent.

3.3 Top-down signatures of consciousness from bain signals

A major challenge in the study of altered states of consciousness has been to establish empirical signatures in brain signals that are characteristic of different states, thus allowing us to identify them "from the outside", i.e., not depending on self-report or behavioral tasks (Sergent and Naccache, 2012). Establishing and validating these signatures also carries significance from a clinical perspective, since they could lead to efficient and specific biomarkers for certain neuropsychiatric conditions (Sitt et al., 2014). Furthermore, when interpreted within a broader theory, some of these signatures may also provide new insights about the nature of the corresponding conscious states, advancing our fundamental understanding of consciousness itself.

In the following, we first provide a broad overview of general aspects of theories of consciousness and then illustrate what a signature of consciousness is by reviewing two well-known examples.

3.3.1 Functionalist and non-functionalist positions on the mind-brain problem

When we consider the most prominent contemporary theories of consciousness, we find that they mainly differ in what they take as valid empirical data to be explained by the theory. There are essentially two positions on this matter, which can be related to the influential division between functionalist and non-functionalist positions on the mind-brain problem. For a functionalist, the subjective quality of conscious experience is rejected as a valid target of scientific explanation. According to this view, most famously articulated by Daniel Dennett in *Consciousness Explained* (Dennett, 1993), only third-person objective measurements fall into the scope of a science of consciousness. This data is limited to observable behavior and neural activity recordings; for instance, whenever an experimental subject claims to be experiencing a certain shade of blue, the neuroscientist is not tasked with finding how a physical process in the brain can cause a subjective feeling of blue, but with determining the mechanisms leading the subject to declare such experience (Dennett, 2003). Non-functionalists, on the other hand, reject this position as a sophisticated form of behaviorism (Block, 1980). According to this view, introspection plays a crucial role in the scientific explanation of consciousness, because it reveals the very nature of the *explanandum* itself; any other kind of data represents, at best, an indirect approximation (Lutz et al., 2002; Shear and Varela, 1999; Chalmers, 1999). It is one of the defining features of consciousness, argue the defenders of this position, that it cannot be illusory (Frankish, 2016) since being conscious of something is precisely what bears that conscious experience into existence (Nida-Rümelin, 2016; Seager, 2017).

When translated into the domain of neuroscience, these positions inform the two most influential contemporary models of consciousness. The global neuronal workspace theory (GNW) (Baars, 2005; Dehaene and Naccache, 2001) links consciousness with the widespread and sustained propagation of activity in the cortex, serving the computational function of broadcasting information to be processed by specialized modules (Mashour et al., 2020). This theory was developed to explain the neural signatures of consciousness seen in cognitive neuroscience experiments—in other words, to explain third-person objective data. On the contrary, the integrated information theory (IIT) (Tononi, 2004; Balduzzi and Tononi, 2009; Oizumi et al., 2014) is based on certain first-person qualities of subjective experience, which are accessed by introspection and can be taken as “postulates”

or “axioms” for the theory (Oizumi et al., 2014). This theory strives to provide a quantitative characterization of consciousness, as well as to determine the neural correlates of conscious contents from first principles only (even though concrete predictions may be computationally intractable (Barrett and Mediano, 2019)). Both theories have been the target of intense criticism (Block, 2011; Tsuchiya et al., 2015; Aru et al., 2012; Lamme, 2006; Doerig et al., 2019; Tsuchiya et al., 2020), which can be taken as a sign that the scientific problem of consciousness remains unsolved.

While the GWT and the IIT are frequently pitted against each other, their predictions for human brains may still be mutually compatible (Seth et al., 2006; Tagliazucchi, 2017). For our purpose, what these two theories have in common is that they follow a top-down approach, in the sense that they both focus on abstract computational or information-theoretical principles, without necessarily specifying how these principles arise as a consequence of local dynamics within the underlying neural substrate. We argue that it is via detailed whole-brain modeling that the points of agreement and divergence between theories, and how they relate to the neurophysiology of the human brain, can and should be studied ahead of possible experiments.

3.3.2 Examples of signatures of consciousness

Since the conception of NCC, neuroscientists have turned to every available neuroimaging technology in the search for signatures of consciousness (Crick and Koch, 1990, 2003). Although many kinds of signatures have been explored (including some related to metabolic consumption (Laureys et al., 2004) or cortical connectivity (Laureys et al., 1999)), for the purposes of this thesis we will focus on signatures measurable with functional neuroimaging tools, like MEG, electroencephalography (EEG), and fMRI because they reflect the dynamics of neural activity. In the sequel, we illustrate the nature and application of signatures of consciousness by elaborating on two well-known examples.

The entropic brain hypothesis

One of the simplest, yet remarkably powerful, theoretical frameworks to furnish signatures of consciousness is Carhart-Harris’ entropic brain hypothesis (EBH) (Carhart-Harris et al., 2014; Carhart-Harris, 2018). Note that in this context we refer with entropy to the Shannon’s entropy (Shannon, 1948) —not to the one

used in thermodynamics—, which measures the degree of unpredictability associated with the observation of a variable. According to the EBH, the richness of conscious experience depends on the complexity of the underlying population-level neuronal activity, which determines the repertoire of states available for the brain to explore. Put simply, conscious states that involve richer experiences might require a more diverse set of brain configurations, which should leave a traceable footprint to be observed in the entropy (or related measure), of the corresponding brain signals. Following this rationale, the level of consciousness should be proportional (at least within reasonable range) to the statistical entropy of brain signals.

Integrated information theory

A strong limitation of standard brain entropy analyses is that they consider only the entropy of individual signals, without acknowledging the multivariate structure of brain dynamics. An attractive way of studying interdependencies between brain signals is with tools drawn from the integrated information theory (IIT) (Tononi, 2008). The IIT proposes an intimate relationship between consciousness and the ability of a physical system to be integrated in such a way that is "more that the sum of its parts", i.e., to display dynamical properties in the whole that are not observed in any of its parts.

The IIT builds on key information-theoretic ideas first presented in the seminal early work of Tononi, Sporns, and Edelman (Tononi et al., 1994) and has been subject of continuous development since (Tononi, 2004; Balduzzi and Tononi, 2009; Oizumi et al., 2014; Mediano et al., 2019). Following Mediano et al. (2019), we distinguish between empirical IIT and fundamentalist IIT as two separate branches of the theory. While fundamentalist IIT has been highly controversial and subject of extensive criticism (Mindt, 2017; Morch, 2019; Barrett and Mediano, 2019; Bayne, 2018), multiple efforts in empirical IIT have been made to overcome the computational challenges of the theory (Krohn and Ostwald, 2017; Kitazono et al., 2018; Toker and Sommer, 2019).

At the core of empirical applications of the IIT is a quantitative measure of integrated information, typically denoted by Φ. There is currently no agreed-upon Φ measure, although multiple proposals have been put forward (Mediano et al., 2019) and can be used to understand and compare the dynamical structure of systems of interest. Indeed, detailed procedures describing how to compute different versions of Φ are available (Mediano et al., 2019). Although the evidence supporting

the IIT as a fundamental theory of consciousness has been contested (Mediano, 2019), measures inspired by empirical IIT have proven useful in analyzing both empirical (Chang et al., 2012; Kim et al., 2018), as well as simulated (Toker and Sommer, 2019; Mediano et al., 2016), neural data. Altogether, the family of information-theoretic measures inspired by empirical IIT provides a valuable toolkit to study the multivariate dynamics of whole-brain models.

3.4 Bottom-up whole-brain models

While human neuroscience research has been increasingly dominated by imaging experiments, an important complement to this research is provided by computational neuroscience (Gerstner et al., 2012). In effect, neuroimaging data is usually insufficient to inform the underlying mechanisms at play behind neural phenomena unfolding at different spatial and temporal scales (Ramsey et al., 2010). In addition, since ethical considerations severely limit direct causal manipulation of human brain activity, most of the neuroimaging literature is limited to correlational studies.

The application of computational models to neuroimaging data with the purpose of making causal and mechanistic assertions has been proposed and developed in parallel with different objectives. For instance, deep neural networks can be used to model information-processing in the brain (Kriegeskorte and Douglas, 2018) by comparing their representational content via second-order isomorphisms (e.g., representational similarity analysis) (Kriegeskorte et al., 2008). These models can be used to investigate the plausibility of different computational architectures within cognitive neuroscience (Kriegeskorte and Kievit, 2013). Another example is dynamic causal modeling (DCM), which was developed to make model-based causal inferences from neuroimaging experiments (Friston et al., 2003). DCM is based on simulating brain signals under the assumption of different causal interactions and then performing model comparison and selection. Finally, whole-brain models are based on dynamical systems coupled by large-scale anatomical connectivity networks and are developed to reproduce the statistics of empirical brain signals at multiple scales (Schirner et al., 2018). We also distinguish whole-brain models from attempts to produce extremely detailed reproductions of large neural circuits (e.g., cortical columns) (Markram, 2006), mainly due to differences in model complexity.

Whole-brain models provide a practical, ethical, and inexpensive "digital scalpel", which allows researchers to explore the counterfactual consequences of modifying structural or dynamical aspects of the brain. More generally, whole-brain models build a bridge from local networked dynamics to the large-scale patterns of activity that are addressed by theoretical signatures of consciousness. As such, they represent a valuable tool to narrow the space of mechanistic explanations compatible with the observed neuroimaging data, including data acquired from subjects undergoing different altered states of consciousness.

In this section, we provide a brief introduction to whole-brain models to the unfamiliar reader, discussing their various types and the principles behind their tuning to empirical data.

3.4.1 What are whole-brain models?

Whole-brain models are sets of equations that describe the dynamics and interactions between neural populations in different brain regions. These models typically focus on the joint evolution of a set of key biophysical variables using systems of coupled differential equations (although discrete time step models can also be used, as will be discussed below). These equations can be built from knowledge concerning the biophysical mechanisms underlying different forms of brain activity, or as phenomenological models chosen by the kind of dynamics they produce. Then, local dynamics are combined by in vivo estimates of anatomical connectivity networks. In particular, fMRI, EEG, and MEG signals can be used to define the statistical observables, diffusion tensor imaging (DTI) can provide information about the structural connectivity between brain regions by means of whole-brain tractography, and positron emission tomography (PET) imaging can inform on metabolism and produce receptor density maps for a given neuromodulator.

Most whole-brain models are structured around three basic elements:

A. Brain parcellation: A brain parcellation determines the number of regions and the spatial resolution at which the brain dynamics take place. The parcellation may include cortical, sub-cortical, and cerebellar regions. Examples of well-known parcellations are the Hagmann parcellation (Hagmann et al., 2008), and the automated anatomical labeling (AAL) atlas (Tzourio-Mazoyer et al., 2002).

B. Anatomical connectivity matrix: This matrix defines the network of connections between brain regions. Most studies are based on the human connectome, obtained by estimating the number of white-matter fibers connecting brain areas from DTI data combined with probabilistic tractography (Sporns et al., 2005). For control purposes, randomized versions of the connectome (null hypothesis networks) may also be employed.

C. Local dynamics: The activity of each brain region is typically determined by the chosen local dynamics plus interaction terms with other regions. A variety of approaches have been proposed to model whole-brain dynamics, including cellular automata (Tagliazucchi et al., 2016a; Haimovici et al., 2013), the Ising spin model (Marinazzo et al., 2014; Abeyasinghe et al., 2020), autoregressive models (Messé et al., 2014), stochastic linear models (Saggio et al., 2016), non-linear oscillators (Cabral et al., 2014; Jobst et al., 2017), neural field theory (Robinson and Roy, 2015; Babaie Janvier and Robinson, 2018), neural mass models (Breakspear et al., 2003; Honey et al., 2009), and dynamic mean-field models (Deco et al., 2018b; Kringelbach et al., 2020). A detailed review of the different models that can be explored within this context can be found in Breakspear (2017); Deco et al. (2008).

The first two items are guided by available experimental data. In contrast, the choice of local dynamics is usually driven by the phenomena under study and the epistemological context at which the modeling effort takes place. Because of this hybrid nature, whole-brain models constructed following this process are sometimes called semi-empirical models. Whole-brain models can be constructed from in-house code, or more easily from platforms, such as The Virtual Brain (`https://www.thevirtualbrain.org/tvb/zwei`) (Ritter et al., 2013).

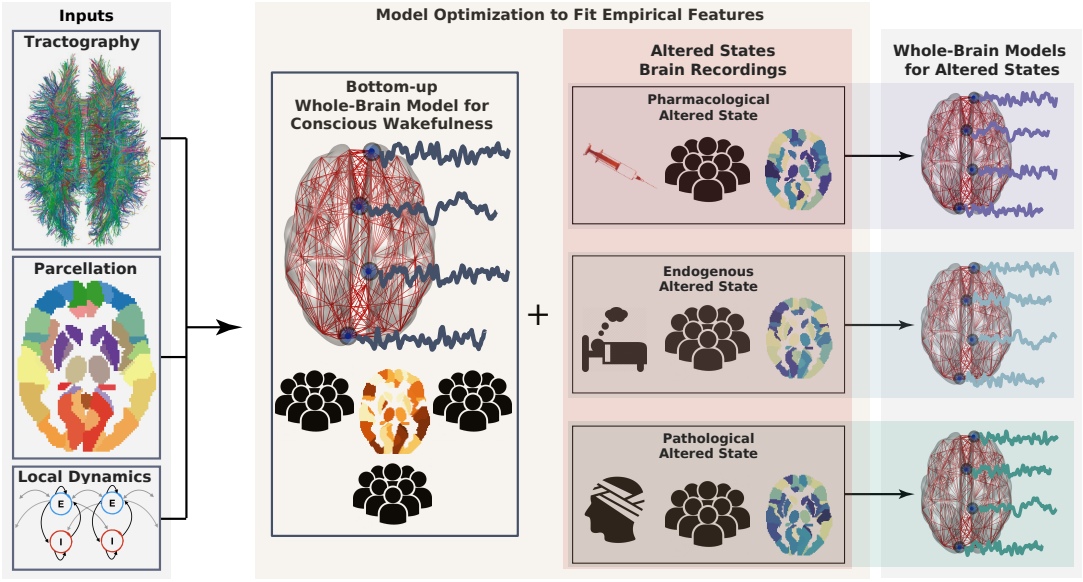

Figure 3.1: Workflow describing the construction of whole-brain models. First, model inputs are determined based on anatomical connectivity, a brain parcellation (representing a certain spatial resolution), and the local dynamics (left). Each region defined by the parcellation is endowed with a specific connectivity profile and local dynamics. Then, the model can be optimized to generate data as similar as possible to the brain activity observed during conscious wakefulness. Generally, this similarity is determined by certain statistical properties of the empirical brain signals, which constitute the target observable. The same or another observable is obtained from subjects during altered states of consciousness and used again as the target of an optimization algorithm to infer model parameters. Following a given working hypothesis, the model for wakeful consciousness can be perturbed in such a way that optimizes the similarity between the target observable for the altered state of consciousness and the data generated by the model. In this way, a whole-brain model for an altered state of consciousness can be used to test working hypotheses about its mechanistic underpinnings.

3.4.2 Models of local dynamics

Despite the diversity of models of local dynamics, such as the Wilson-Cowan (Destexhe and Sejnowski, 2009) or the Jansen-Ritt (Coronel-Oliveros et al., 2021), we chose to showcase two models that have been frequently used to assess mechanistic hypotheses behind both pharmacologically and physiologically-induced altered states of consciousness: the dynamic mean field model (Deco et al., 2014, 2018b) and the model comprised by Stuart-Landau non-linear coupled oscillators (Cabral et al., 2014; Jobst et al., 2017; Ipiña et al., 2020). These examples are chosen to rep-

resent a biologically realistic model (dynamic mean field) and a phenomenological model (Stuart-Landau oscillators); moreover, these models have been applied to different states of consciousness, making them pertinent in the context of the present discussion.

Dynamic mean-field (DMF) model

In this approach, the neuronal activity in a given brain region is represented by a set of differential equations describing the interaction between inhibitory and excitatory pools of neurons (Renart et al., 2004). The DMF presents one dynamical variable for each population, the synaptic current, and two auxiliary variables, the firing rate, and the synaptic gating, where the excitatory coupling is mediated by N-methyl-D-aspartate (NMDA) receptors and the inhibitory by gamma-aminobutyric acid (GABA)-A receptors. The interregional coupling is considered excitatory-to-excitatory only, and a feedback inhibition control in the excitatory current equation is included (Deco et al., 2014). There are three types of output variables: synaptic gating variable, currents, and firing rates, that is then included in a nonlinear hemodynamical model (Friston et al., 2000) to simulate the regional bold-oxygen-level-dependent (BOLD) signals.

The key idea behind the mean-field approximation is to reduce the high-dimensional randomly interacting elements to a system of elements treated as independent.

Thus, this approach represents the average activity of an homogeneous population of neurons by the activity of a single unit of this class, reducing in this way the dimensionality of the system. In spite of these approximations, the dynamic mean field model incorporates a detailed biophysical description of the local dynamics, which increases the interpretability of the model parameters. This model is core for this thesis (see 4.1 for corresponding equations).

Stuart-Landau non-Linear oscillator model

This approach builds on the idea that neural activity can exhibit—under suitable conditions—self-sustained oscillations at the population level (Cabral et al., 2014; Jobst et al., 2017; Ipiña et al., 2020; Tagliazucchi et al., 2016a; Deco et al., 2018a). In this model, — also known as the Hopf model — the dynamical behavior is repre-

sented by a non-linear oscillator with the addition of Gaussian noise at the proximity of a Hopf bifurcation (Marsden and McCracken, 2012). By changing a single model parameter (i.e., bifurcation parameter) across a critical value, the model gives rise to three qualitatively different asymptotic behaviors: harmonic oscillations, fixed point dynamics governed by noise, and intermittent complex oscillations when the bifurcation parameter is close to the bifurcation (i.e., at dynamical criticality). Correspondingly, the model is determined by two parameters: the bifurcation parameter of the Hopf bifurcation in the local dynamics, and the coupling strength factor that scales the anatomical connectivity matrix. In contrast to the DMF model, coupled Stuart-Landau non-linear oscillators constitute a phenomenological model, i.e., the model parameters do not map into any biophysically known variables. In this case, the model is attractive due to its conceptual simplicity, which is given by its capacity to produce three qualitatively different behaviors of interest by changing a single parameter.

3.4.3 How to fit whole-brain models to neuroimaging data?

Whole-brain models are tuned to reproduce specific features of brain activity. The way in which this is ensured is via optimization of the free parameters in the local dynamics plus the coupling strength. Parameter values are usually selected so that the model matches a certain statistical observable computed from the experimental data.

For example, the common choice when using the DMF whole-brain model is to use only one parameter to scale the strength of the connectivity matrix, usually known as the global coupling parameter. During model training, an exhaustive exploration of this parameter is conducted over a wide range of values. The parameter value is chosen to maximize the similarity between the observable computed from simulated and experimental data. For instance, the parameter can be chosen to minimize the Kolmogorov-Smirnov distance between the functional connectivity dynamics (FCD) distributions of the simulated and real data (Deco et al., 2014).

This kind of brute-force optimization is employed when the number of free parameters is low (i.e., two or three). However, it is also possible to separately optimize the parameters governing the local dynamics of each node, which dramatically increases the dimensionality of the search space, and thus requires more elaborated optimization techniques, such as genetic algorithms (Ipiña et al., 2020), or

Bayesian optimization (Wischnewski et al., 2022), among others. The advantage of considering a small set of global parameters resides in its simplicity and scalability, but unfortunately it misses the dynamical heterogeneity among brain regions. These heterogeneities can be modeled at the expense of increasing the parameter space. Essentially, the choice of model complexity (i.e., the number of free parameters) depends on the scientific question and its associated hypotheses.

Since adding more free parameters increases the computational cost of the optimization procedure, it becomes critical to choose parameters reflecting variables that are considered relevant, either from a general neurobiological perspective or in the specific context of the altered state under investigation. Depending on the latter, the parameters could be divided into groups that are allowed to change independently based on different criteria, including structural lesion maps, receptor densities, local gene expression profiles, and parcellations that reflect the neural substrate of certain cognitive functions, among others.

After choosing the parcellation, the equations governing the local dynamics and their interaction terms, the interregional coupling given by the structural connectivity matrix, and selecting a criterion to constrain the dimensionality of the parameter space. The last critical step is to define the observable which will be used to construct the target function for the optimization procedure. As mentioned above, one possibility is to optimize the model to reproduce the statistics of FCD. Perhaps a more straightforward option is to optimize the "static" functional connectivity matrix computed over the duration of the complete experiment, an approach followed by Refs. (Ipiña et al., 2020) and (Jobst et al., 2017), among others. Other observables related to the collective dynamics can be obtained from the synchrony and metastability, as defined in the context of the Kuramoto model (Jobst et al., 2017; Shanahan, 2010). In general, any meaningful computation summarizing the spatiotemporal structure of a neuroimaging dataset constitutes a valid observable, with the adequate choice depending on the scientific question and the altered state of consciousness under study.

Finally, some natural candidates for observables to be fitted by whole-brain models are precisely the high-level signatures of consciousness put forward by theoretical predictions, such as the different measures of information integration, complexity, and entropy that were reviewed in the previous section. The objective is to fit whole-brain models using these signatures as target functions and then assess the biological plausibility of the optimal model parameters, which allows for the testing of the consistency of these signatures from a bottom-up perspective. Alternatively,

signatures of consciousness can be computed from the model—initially fitted to other observables—and compared to the empirical results. This highlights the need to understand which kind of local dynamics allow the simultaneous reproduction of multiple observables derived from experimental data.

3.4.4 Whole-brain models applied to the study of consciousness

The available evidence suggests that states of consciousness are not determined by activity in individual brain areas, but emerge as a global property of the brain, which in turn is shaped by its large-scale structural and functional organization (Dehaene and Naccache, 2001; Dehaene and Changeux, 2011; Tononi and Edelman, 1998). According to this view, whole-brain models provide a fertile ground to explore how global signatures of different states of consciousness emerge from local dynamics. This promise is already being met, as shown by several recent articles (Cabral et al., 2014; Jobst et al., 2017; Ipiña et al., 2020; Tagliazucchi et al., 2016a; Deco et al., 2018a; Bocaccio et al., 2019).

For example, transitions from wakefulness into other states, such as the different stages of human sleep or the state induced by general anaesthetics, have been interpreted as phase transitions in neural mass models and in terms of the collective dynamics of coupled Stuart-Landau oscillators (Cabral et al., 2014; Jobst et al., 2017; Ipiña et al., 2020). Noise-driven systems at dynamical criticality result in dynamics compatible with neuroimaging recordings obtained during conscious wakefulness, and departures from these dynamics better reflect different states of unconsciousness (Tagliazucchi et al., 2016a; Deco et al., 2018a; Carhart-Harris et al., 2014; Bocaccio et al., 2019; Solovey et al., 2015; Alonso et al., 2014). These results are consistent with the hypothesis of statistical criticality (e.g., proximity to a second order phase transition) as a fundamental principle of brain organization (Chialvo, 2010). Even though parallels can be drawn between statistical and dynamical criticality, we limit our discussion to the former since the relationship between both concepts is complicated and beyond the scope of this thesis. Following the example of the perturbational complexity index (PCI) index (which is obtained by perturbing the cortex with Transcranial magnetic stimulation (TMS) and measuring the complexity of the elicited response) (Casali et al., 2013), whole-brain models can be systematically "perturbed" by incorporating changes into the dynamical equations. The in silico rehearsal of perturbations is useful to test hypotheses concerning which parts of the model are essential to produce different signatures of consciousness. A prominent

example of this perturbational analysis applied to whole-brain models can be found in a recent article (Deco et al., 2018a) where a whole-brain model based on coupled Stuart-Landau oscillators was fitted to empirical fMRI data acquired from subjects during deep sleep. The model was then modified by changing local bifurcation parameters with a greedy optimization algorithm, which unveiled the optimal perturbation profile to increase the similarity to a target brain state (in this case, conscious wakefulness). Another relevant example of this perturbational approach is found in Deco et al. (2018b), where a transition was shaped by the effects of neuromodulation. The authors investigated the transition from resting state activity acquired under a placebo condition towards the altered state of consciousness induced by the serotonin 2A receptor agonist lysergic acid diethylamide (LSD). A dynamical mean-field model was fitted to minimize the difference between FCD of the simulated activity and the empirical data of subjects in the placebo condition, which allowed to determine the optimal value of the global coupling parameter. Then, an empirical map of 5-HT_{2A} receptor density was used to modulate the synaptic gain, effectively simulating the heterogeneous effects of LSD across the whole brain. As a control, the authors showed that using maps for the density of other serotonin receptor subtypes decreased the goodness of fit, thus corroborating the well-known association between LSD and the 5-HT_{2A} receptor. This model is a key element of this thesis.

Another interesting possibility is to assess the consequences of stimulation protocols that are impossible to apply in vivo. An example is the Perturbative Integration Latency Index (PILI) (Deco et al., 2018a), which measures the latency of the return to baseline after a strong perturbation that generates dynamical changes detectable over long temporal scales (on the order of tens of seconds). This in silico perturbative approach allows to systematically investigate how the response of brain activity upon external perturbations is indicative of the state of consciousness, providing new mechanistic insights into the capacity of the human brain to integrate and segregate information over different time scales.

In Ref. (Ipiña et al., 2020), the authors used a model of coupled Stuart-Landau oscillators to model the regional changes in dynamical stability that occur during the wake-sleep cycle. Brain regions belonging to different resting state networks (RSN) (Damoiseaux et al., 2006) were considered as independent sources of variation for the local model parameters. Using a stochastic optimization algorithm, the authors represented the transition from wakefulness into deep sleep as a sequence of changes in the stability of brain activity within canonical RSN. A follow-up paper extended this analysis to other states of reduced consciousness (includ-

ing anaesthesia and patients suffering from disorders of consciousness) and investigated the possibility of inducing transitions to conscious wakefulness by means of simulated periodic stimulation at the resonant frequency of each node in the model (Sanz Perl et al., 2021).

3.5 Thesis overview

The theoretical background, computational advances, and results are part of a work developed by a interdisciplinary team of international collaboration. This collaborative effort produced three scientific articles, which represent the core to this thesis. The first one, *Whole-brain models to explore altered states of consciousness from the bottom up* (Cofré et al., 2020), sets the theoretical framework. This work was developed by Rodrigo Cofré, Rubén Herzog, Pedro A.M. Mediano, Juan Piccinini, Fernando E. Rosas, Yonatan Sanz Perl and Enzo Tagliazucchi, and it was published in 2020 Brain Sciences (https://doi.org/10.3390/brainsci10090626). Both Introduction and Discussion are highly based on that article. In the second one, *Neural mass modelling for the masses: Democratising access to whole-brain biophysical modelling with FastDMF*, we improved the dynamic mean field model such that thousand of regions can be simultaneously modelled in a usual desktop computer. This work was developed by Rubén Herzog, Pedro A.M. Mediano, Fernando E. Rosas, Andrea I. Luppi, Yonatan Sanz Perl, Enzo Tagliazucchi, Morten Kringelbach, Rodrigo Cofré and Gustavo Deco, and it was submitted to Neuroimage in April 2022 (https://doi.org/10.1101/2022.04.11.487903). This article corresponds to Chapter 5. Finally, the results of Chapter 6 are mainly based on *A mechanistic model of the neural entropy increase elicited by psychedelic drugs*, a work developed by Rubén Herzog, Pedro A. M. Mediano, Fernando E. Rosas, Robin Carhart-Harris, Yonatan Sanz Perl, Enzo Tagliazucchi and Rodrigo Cofre, published 2020 on Scientific Reports (https://doi.org/10.1038/s41598-020-74060-6).

3.5.1 Accelerating the application of biophysically grounded whole-brain models

In Chapter 5, we present the FastDMF, a user-friendly and computationally efficient implementation of the DMF model that aims to make it popular an accessible to the community. We made three important advances for the field of whole-brain mod-

elling: i) we developed a C++ implementation, significantly reducing the computational costs of the algorithm; ii) by combining analytical and numerical tools, we develop a novel estimation of the feedback inhibition control based on the structural connectivity, bypassing an important computational bottleneck; and iii) we coupled the FastDMF with a Bayesian optimization algorithm (BOA) significantly reducing the number of simulations required to fit the FastDMF to empirical neuroimaging data. Finally, our advances open the possibility of simulating, on a common laptop, thousands of brain regions in a biophysically grounded whole-brain model. The aforementioned advances are key to the results obtained in the case study developed in Chapter 6.

3.5.2 Case study: modeling neural entropy increases induced by psychedelics

In Chapter 6, to further highlight what we can learn from whole-brain models, we discuss an illustrative example of a bottom-up model that successfully matches a global signature of altered conscious (Herzog et al., 2020). After optimizing the model to fit the FCD of placebo and LSD conditions (Deco et al., 2018b), a significant entropy increase of brain signals was found in LSD versus placebo as a consequence of simulated 5-HT_{2A} receptor activation. Thus, the model was capable of identifying a low-level (i.e., molecular scale) mechanism leading to increased neural entropy, which is a robust signature of the psychedelic state (Carhart-Harris et al., 2014; Carhart-Harris, 2018).

METHODS

In this section we describe the computational and analytical methods associated to the dynamic mean field model. In addition, despite no empirical data was produced in this thesis, the empirical data sets (fMRI, DTI, PET) used are also described for completeness. All computational analysis were performed using Matlab, unless otherwise is stated.

4.1 Dynamic mean field model

The dynamic mean field (DMF) model introduced by Deco *et al.*,(Deco et al., 2014, 2018b) uses a set of coupled stochastic differential equations to model the dynamics of the activity at one or more interacting brain regions. In this model, each brain region is modelled as two coupled neuronal masses — one excitatory and one inhibitory — and considers excitatory and inhibitory synaptic currents mediated by NMDA and $GABA_A$ receptors, respectively. Different brain regions (usually defined by a given brain parcellation) are coupled via their excitatory populations exclusively, according to the structural connectivity matrix.

The key idea behind the mean-field approximation is to reduce a high-dimensional system of randomly interacting elements to a system of elements treated as independent. This approach represents the average activity of a homogeneous population of neurons by the activity of a single unit of this class.

The model establishes that the n-th brain area obeys the following equations:

$$I_n^{(\mathrm{E})} = W_\mathrm{E} I_0 + w_+ J_\mathrm{NMDA} S_n^{(\mathrm{E})} + G J_\mathrm{NMDA} \sum_p^N C_{np} S_p^{(\mathrm{E})} - J_n^\mathrm{FIC} S_n^{(\mathrm{I})} \,, \tag{4.1a}$$

$$I_n^{(\mathrm{I})} = W_\mathrm{I} I_0 + J_\mathrm{NMDA} S_n^{(\mathrm{E})} - S_n^{(\mathrm{I})} \,, \tag{4.1b}$$

$$r_n^{(\mathrm{E})} = F_\mathrm{E}\left(I_n^{(\mathrm{E})}\right) = \frac{g_\mathrm{nm}^{(\mathrm{E})}\, g_\mathrm{E}\left(I_n^{(\mathrm{E})} - I_{thr}^{(\mathrm{E})}\right)}{1 - \exp\left(-d_\mathrm{E}\, g_\mathrm{nm}^{(\mathrm{E})}\, g_\mathrm{E}\left(I_n^{(\mathrm{E})} - I_{thr}^{(\mathrm{E})}\right)\right)} \,, \tag{4.1c}$$

$$r_n^{(\mathrm{I})} = F_\mathrm{I}\left(I_n^{(\mathrm{I})}\right) = \frac{g_\mathrm{nm}^{(\mathrm{I})}\, g_\mathrm{I}\left(I_n^{(\mathrm{I})} - I_{thr}^{(\mathrm{I})}\right)}{1 - \exp\left(-d_\mathrm{I}\, g_\mathrm{nm}^{(\mathrm{I})}\, g_\mathrm{I}\left(I_n^{(\mathrm{I})} - I_{thr}^{(\mathrm{I})}\right)\right)} \,, \tag{4.1d}$$

$$\frac{dS_n^{(\mathrm{E})}(t)}{dt} = -\frac{S_n^{(\mathrm{E})}}{\tau_\mathrm{NMDA}} + \left(1 - S_n^{(\mathrm{E})}\right)\gamma r_n^{(\mathrm{E})} + \sigma \nu_n(t) \,, \tag{4.1e}$$

$$\frac{dS_n^{(\mathrm{I})}(t)}{dt} = -\frac{S_n^{(\mathrm{I})}}{\tau_\mathrm{GABA}} + r_n^{(\mathrm{I})} + \sigma \nu_n(t) \,, \tag{4.1f}$$

$$g_{\mathrm{nm}_n}^{(\mathrm{E,I})} = 1 + s^{(\mathrm{E,I})}\, d_n^\mathrm{rec} \,. \tag{4.1g}$$

Above, for each excitatory (E) and inhibitory (I) neural mass, the quantities I_n, r_n, and S_n represent its total input current, firing rate, and synaptic gating variable, respectively. The functions $F_\mathrm{E}(\cdot), F_\mathrm{I}(\cdot)$ determine the transfer functions (characterised by a *F-I curve*), representing the non-linear relationship between the input current and the output firing rate of excitatory and inhibitory neural populations. $g_{\mathrm{nm}_n}^{(\mathrm{E,I})}$ corresponds to the local neuromodulatory gain modulation associated with a given local receptor density d_n^rec, which is weighted by $s^{(\mathrm{E,I})}$. $s^{(\mathrm{E,I})}$ is a free parameter crucial for the fitting step in the case study of Chapter 6. Crucial for analytical derivations, J_n^FIC is the local feedback inhibitory control (FIC) of region n, and ν_n is uncorrelated Gaussian noise injected to region n (Wong and Wang, 2006). The remaining quantities involved in these equations are introduced in Table 4.1. The interested reader is referred to the original publication for further details.(Deco et al., 2014) The above stochastic differential equations were discretized and solved using the Euler-Maruyama integration method(Wu et al., 2013) and using the parameter values shown in Table 4.1. The first 10 seconds of simulation were discarded to ensure stability of the dynamical system.

Table 4.1: Dynamic Mean Field (DMF) model parameters

Parameter	*Symbol*	*Value*
External current	I_0	0.382 nA
Excitatory scaling factor for I_0	W_E	1
Inhibitory scaling factor for I_0	W_I	0.7
Local excitatory recurrence	w_+	1.4
Excitatory synaptic coupling	J_{NMDA}	0.15 nA
Threshold for $F(I_n^{(\mathrm{E})})$	$I_{thr}^{(\mathrm{E})}$	0.403 nA
Threshold for $F(I_n^{(\mathrm{I})})$	$I_{thr}^{(\mathrm{I})}$	0.288 nA
Gain factor of $F(I_n^{(\mathrm{E})})$	g_E	310 nC^{-1}
Gain factor of $F(I_n^{(\mathrm{I})})$	g_I	615 nC^{-1}
Shape of $F(I_n^{(\mathrm{E})})$ around $I_{thr}^{(\mathrm{E})}$	d_E	0.16 s
Shape of $F(I_n^{(\mathrm{I})})$ around $I_{thr}^{(\mathrm{I})}$	d_I	0.087 s
Excitatory kinetic parameter	γ	0.641
Amplitude of uncorrelated Gaussian noise υ_n	σ	0.01 nA
Time constant of NMDA	τ_{NMDA}	100 ms
Time constant of GABA	τ_{GABA}	10 ms

4.1.1 Balloon-Windkessel hemodynamic model of BOLD activity

To transform the simulated excitatory firing rate signals into BOLD activity, we used the Balloon-Windkessel following Stephan et al. (2007), which defines a dynamic relationship between firing rates and BOLD signals. This model defines a vasodilatory signal s_n for each region, subject to auto regulatory feedback γ_{BW} and decay κ, which influences the blood inflow f_n, inducing changes in blood volume υ_n, and deoxyhemoglobin content q_n following:

$$\frac{ds_n}{dt} = 0.5r_n^{(\mathrm{E})} + 3 - \kappa s_n - \gamma_{BW}(f_n - 1)\,, \tag{4.2a}$$

$$\frac{df_n}{dt} = s_n\,, \tag{4.2b}$$

$$\tau_{BW}\frac{dq_n}{dt} = \frac{f_n(1-\rho)^{1/f_n}}{\rho} - \frac{q_n \upsilon_n^{1/\alpha_{BW}}}{\upsilon_n}\,, \tag{4.2c}$$

$$B_n = V_0\left[k_1(1-q_n) + k_2(1-q_n/\upsilon_n) + k_3(1-\upsilon_n)\right]\,, \tag{4.2d}$$

where B_n is the simulated BOLD signal, ρ is the resting oxygen extraction fraction, α_{BW} represents the resistance of the veins, τ_{BW} is a time constant, and k_1, k_2, and k_3 are coefficients estimated from data. Accordingly, all biophysical parameters were taken from reference (Stephan et al., 2007).

4.1.2 A first-order analytical solution for the feedback inhibitory control

Here we provide an analytical derivation of how J_n^{FIC} depends on other model parameters. The derivation is based on two assumptions, whose range of validity is also discussed.

As a first step in the derivation, we write down the discrete-time form of $S_n^{(\mathrm{E})}(t)$ using Eq. (4.1e) according to the Euler-Maruyama method, which gives

$$S_n^{(\mathrm{E})}(t+1) = S_n^{(\mathrm{E})}(t) + \left(-\frac{S_n^{(\mathrm{E})}(t)}{\tau_{NMDA}} + (1 - S_n^{(\mathrm{E})}(t))\gamma r_n^{(\mathrm{E})}(t) \right) dt + \sigma \upsilon_n(t)\sqrt{dt}\,. \tag{4.3}$$

Then, we calculate the steady-state average of $S_n^{(\mathrm{E})}$ by computing the expected value to Eq. (4.3). Using the steady-state property, and introducing the shorthand notation $\mathbb{E}\left[S_n^{(\mathrm{E})}\right] := \mathbb{E}\left[S_n^{(\mathrm{E})}(t)\right] = \mathbb{E}\left[S_n^{(\mathrm{E})}(t+1)\right]$, a direct calculation shows that:

$$\mathbb{E}\left[S_n^{(\mathrm{E})}(t)\right] = \gamma\,\tau_{NMDA}\left(\mathbb{E}\left[r_n^{(\mathrm{E})}\right] - \mathbb{E}\left[S_n^{(\mathrm{E})}(t) r_n^{(\mathrm{E})}(t)\right]\right)\,. \tag{4.4}$$

To proceed further in the derivation one needs an expression for $\mathbb{E}\left[S_n^{(\mathrm{E})}(t) r_n^{(\mathrm{E})}(t)\right]$. It can be found via numerical simulations that the covariance between these two variables is lower than 0.1, so we assume that $\mathbb{E}\left[S_n^{(\mathrm{E})}(t) r_n^{(\mathrm{E})}(t)\right] = \mathbb{E}\left[S_n^{(\mathrm{E})}\right]\mathbb{E}\left[r_n^{(\mathrm{E})}\right]$ (**assumption 1**). Plugging this relationship into Eq. (4.4) yields

$$\mathbb{E}\left[S_n^{(\mathrm{E})}\right] = \frac{\gamma\,\tau_{NMDA}\mathbb{E}\left[r_n^{(\mathrm{E})}\right]}{1 + \gamma\,\tau_{NMDA}\mathbb{E}\left[r_n^{(\mathrm{E})}\right]}\,. \tag{4.5}$$

The FIC is assumed to successfully regulate the average excitatory firing rate if, for any G, they remain close to the firing rates obtained for the disconnected case ($G = 0$): $\mathbb{E}\left[r_n^{(\mathrm{E})}\right] = 3.4\,\mathrm{Hz}$ for all the nodes of the network. Importantly, this implies that $\mathbb{E}\left[S_n^{(\mathrm{E})}\right]$ is also constant across nodes, and using Eq. (4.5) one finds that $\mathbb{E}\left[S_n^{(\mathrm{E})}\right] = 0.179$. Numerical simulations for uncoupled regions give a steady-state average of $S_n^{(\mathrm{E})} = 0.179$, providing empirical confirmation of the accuracy of this approximation

for that scenario.

As a second step of the derivation, let's study the expected values of $I_n^{(\mathrm{I})}$, $r_n^{(\mathrm{E})}$, $r_n^{(\mathrm{I})}$, and $S_n^{(\mathrm{I})}$ using the equations provided in the previous section. The equations for $I_n^{(\mathrm{I})}$ and $S_n^{(\mathrm{I})}$ are linear, so computing the expected value is straightforward; in contrast, $\mathbb{E}\left[r_n^{(\mathrm{E})}\right] = \mathbb{E}\left[F(I_n^{(\mathrm{E})})\right]$ involves a non-linear function F. To move forward, we adopt a first-order Taylor expansion and assume that the approximation $\mathbb{E}\left[r_n^{(\mathrm{E})}\right] \approx F(\mathbb{E}\left[I_n^{(\mathrm{E})}\right])$ is accurate (**assumption 2**). By doing this, one obtains a system of four non-linear equations with four unknowns (the expected values of $I_n^{(\mathrm{I})}$, $I_n^{(\mathrm{E})}$, $S_n^{(\mathrm{I})}$ and $r_n^{(\mathrm{I})}$), which can be solved numerically.

Since the inputs to this system of equations — i.e. $\mathbb{E}\left[r^{(\mathrm{E})}\right]$ and $\mathbb{E}\left[S^{(\mathrm{E})}\right]$ — adopt the same values for all regions, the results yield the same values for all regions.

As a final step, we apply the expected value of Eq. (4.1a). Using the fact that $\mathbb{E}\left[S_n^{(\mathrm{E})}\right]$ does not depend on n, a direct calculation shows that

$$\mathbb{E}\left[I_n^{(\mathrm{E})}\right] = W_E I_0 + w_+ J_{NMDA}\mathbb{E}\left[S^{(\mathrm{E})}\right] + J_{NMDA}\mathbb{E}\left[S^{(\mathrm{E})}\right] G\beta_n - J_n^{\mathrm{FIC}}\mathbb{E}\left[S^{(\mathrm{I})}\right] , \tag{4.6}$$

where $\beta_n := \sum_p C_{np}$ is the strength of node n. To conclude, one can solve the above equation for J_n^{FIC} to obtain

$$J_n^{\mathrm{FIC}} = \alpha G\beta_n + c \, . \tag{4.7}$$

Above, $\alpha = \frac{J_{NMDA}\mathbb{E}\left[S^{(\mathrm{E})}\right]}{\mathbb{E}\left[S^{(\mathrm{I})}\right]}$ is a ratio between the expected fraction of open NMDA to GABA channels — which represents a global E/I balance parameter, and $c = \frac{1}{\mathbb{E}\left[S^{(\mathrm{I})}\right]}\left(W_E I_0 + w_+ J_{NMDA}\mathbb{E}\left[S^{(\mathrm{E})}\right] - \mathbb{E}\left[I^{(\mathrm{E})}\right]\right)$ is an offset parameter that corresponds to the J_n^{FIC} for uncoupled regions. To find values of α and c, one can plug $\mathbb{E}\left[r^{(\mathrm{E})}\right] = 3.4$ Hz, the aforementioned expected value obtained for uncoupled regions, and the usual model parameters, to obtain $\alpha = 0.67$ and $c = 0.97$. To match the value of $J_n^{\mathrm{FIC}} = 1$ for uncoupled regions (corresponding to $G = 0$), one can use approximately $c = 1$.

Please note that this first-order approximation is based on the expected values obtained for uncoupled regions, so one should expect that **assumption 2** may become less accurate as G increases, which reflects that the non-linear effects of $F(\cdot)$ play a more important role in determining the optimal value of J_n^{FIC} to attain stability.

4.1.3 Bayesian optimization

A Matlab implementation of Bayesian optimization(Shahriari et al., 2016; Ulmasov et al., 2016) with an expected-improvement acquisition function was used to optimize the DMF model parameters. The objective function was defined as the Kolmogorov-Smirnov (K-S) distance between the histogram of the pooled empirical FCD (FCD_{emp}) and the histogram of the simulated FCD (FCD_{dmf}). We simulated 500 seconds of BOLD signals sampled at TR=2, generating a number of time points comparable to the empirical BOLD signals. The optimization was run assuming a stochastic objective function, letting the algorithm to randomly select the initial conditions for each simulation. The algorithm stopped when the improvement in the objective function was less than 1%.

4.1.4 Estimating differential entropy of firing rates

One of the results of each simulation is a set of N (the number of simulated regions) time series representing the firing rate of each excitatory population ($r_n^{(E)}$), which we analysed to approximate the entropy each region. For a continuous random variable X with associated probability distribution function $p(x)$, and support $\mathcal{X}$, its differential entropy is defined as (Cover, 1999)

$$h(X) = -\int_{\mathcal{X}} p(x) \ln p(x) \, \mathrm{d}x \, . \tag{4.8}$$

Note that for discrete variables the integral is replaced by a sum, and the probability distribution function by a probability mass function. Estimating differential entropy from data is in general a hard problem, stemming (for the most part) from the difficulties in estimating a probability density from samples (Lizier, 2014).

We found that the probability distributions of firing rates $r_n^{(E)}$ were well approximated by a Gamma distribution (Figure 6.1A) for examples). To check the goodness of fit (GOF) of Gamma distribution to the simulated excitatory firing rates of each region, we generated 10^6 time points for each region using the AAL90 atlas and G=2 (Deco et al., 2018b). Then, a Gamma distribution was fitted and the Kolmogorov-Smirnov (K-S) distance between the firing rate distribution and 1000 random samples (same size) of the respective fitted Gamma distribution was computed. We summarise the Gamma GOF as the average K-S for each region under both condi-

tions. To assess the significance of average GOF values, we generated an acceptance interval computing the K-S distance between 100 independent samples (same size) sampled from exactly the same Gamma distribution (re-sampled K-S). This procedure was repeated for different set of Gamma distribution parameters. If the average K-S distance of a given region falls below the maximum re-sampled K-S distance, we consider this region to be well fitted by the Gamma distribution.

All regions fall within the acceptance interval (Fig 4.1), enabling us to use the Gamma distribution parameters to estimate each region's Shannon's differential entropy.

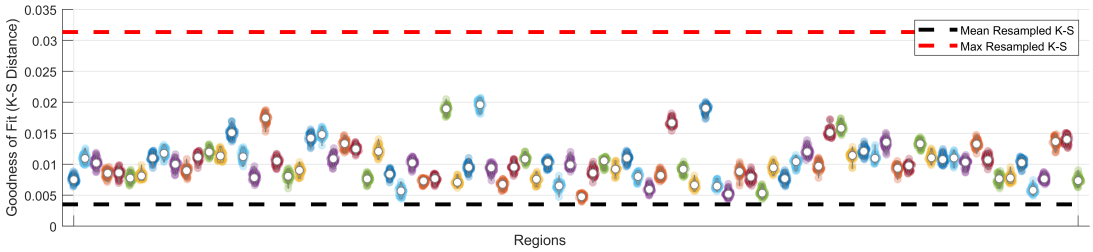

Figure 4.1: **Goodness of fit of Gamma distribution to firing rates distribution are within confidence intervals.** Distribution of K-S distance values are represented by violins, where white circle denotes average K-S. Black (red) dashed line represent the average (maximum) re-sampled K-S distance (see the text for details).

This observation greatly simplifies the estimation of the differential entropy of each region, which is reduced to estimating the shape and scale parameters of the gamma distribution, k and θ. Once these parameters are estimated (in our case by standard maximum likelihood estimation), the differential entropy of a Gamma distribution can be computed analytically in closed form as

$$h(X) = k + \ln\theta + \ln\Gamma(k) + (1-k)\psi(k), \tag{4.9}$$

where $\Gamma(\cdot)$ and $\psi(\cdot)$ are the standard Gamma and Digamma functions, respectively.

4.1.5 Linear models for local entropy change and three-way separation of AAL regions

To explain the changes in differential entropy (Δh_n) across regions and conditions (e.g. 5HT2A-R neuromodulation), we trained several linear mixed-effects models (McLean et al., 1991) using different sets of covariates designed to test specific hypotheses about the relation between entropy, receptor density, and network topol-

ogy. We aimed to split brain regions in 3 groups: S_1, S_2, and R_2, where the two former are regions whose change in entropy depends on the connectivity strength, and the later on receptor density.

First, we implemented a simple algorithm to find the group of regions where the change in entropy depends on region strength. To do so, we designed an optimisation procedure to find the optimal fit between Δh_n and connectivity strength, using an anatomical prior that included all occipital regions of the AAL90 atlas for a quadratic fit (i.e. the S_1 group). The rest of the regions were used for a linear fit between Δh_n and receptor density. Regions were assigned to the S_1 group if the residuals of the quadratic fit to strength were lower than the residuals for the linear fit to receptor density. An optimization algorithm randomly partitioned the remaining regions in two groups until a maximum in the R^2 of a linear mixed-effects model with 3 grouping variables (i.e. S_1, S_2, and R_2) is obtained. The model used a quadratic fit for strength-dependent regions, and a linear one for receptor-dependent regions. Then, the three-way separation of brain regions induces a linear-mixed effect model with 7 terms: a constant term and 2 fixed-effect terms for each group of regions: one for the connectivity strength and one for the receptor density. The model was fit by maximum likelihood, and explanatory power measured using R^2.

4.2 Data for Chapter 5

4.2.1 Participants

A total of 63 healthy subjects (36 females, mean ± SD, 23 ± 43.3 years) were selected from a data set previously described in a sleep-related study by Tagliazucchi and Laufs (Tagliazucchi and Laufs, 2014). Participants entered the scanner at 7 PM and were asked to relax, close their eyes, and not fight the sleep onset. We selected 13 subjects with contiguous resting state time series of at least 200 volumes to perform our analysis. The local ethics committee approves the experimental protocol (Goethe-Universität Frankfurt, Germany, protocol number: 305/07), and written informed consent was asked to all participants before the experiment. The study was conducted according to the Helsinki Declaration on ethical research.

4.2.2 MRI data acquisition

MRI images were acquired on a 3-T Siemens Trio scanner (Erlangen, Germany) and fMRI acquisition parameters were 1505 volumes of T2-weighted echo planar images, $TR/TE = 2080\ ms/30\ ms$, matrix 64×64, voxel size $3 \times 3 \times 3\ mm^3$, distance factor 50%; field of view 192 mm^2.

4.2.3 Brain parcellation to extract BOLD time series

AAL 90 To extract the time series of BOLD signals from each participant in a coarse parcellation, we used the AAL90 parcellation with 90 brain areas anatomically defined in Tzourio-Mazoyer et al., 2002. (Tzourio-Mazoyer et al., 2002)

Schaefer 1000 To extract the time series of BOLD signals from each participant, in a finer scale we used the Schaefer functional parcellation with 1000 brain areas, which was based on estimation from a large dataset (n = 1489).(Schaefer et al., 2018)

4.2.4 Structural connectivity

AAL 90 The structural connectome was obtained applying diffusion tensor imaging (DTI) to diffusion weighted imaging (DWI) recordings from 16 healthy right-handed participants (11 men and 5 women, mean age: 24.75 ± 2.54 years) recruited online at Aarhus University, Denmark, as used in previous studies. (Cabral et al., 2012; Deco et al., 2018b; Ipiña et al., 2020) Briefly, the construction of the structural connectivity matrix (SC) was performed following the following steps: the regions defined using the AAL template (Tzourio-Mazoyer et al., 2002) were warped from MNI space to the diffusion MRI native space using the FLIRT tool from the FSL toolbox (`www.fmrib.ox.ac.uk/fsl`, FMRIB, Oxford). Then, probabilistic tractography with default parameters of the FSL diffusion toolbox (Fdt) were used to estimate the connections between region. The local probability distribution of fibre direction at each voxel was estimated following Behrens et al. (2003) (Behrens et al., 2003) and the probtrackx tool in Fdt was used for the automatic estimation of crossing fibres within each voxel. Using a sampling of 5000 streamlines per voxel, the connection probability from a seed voxel i to another voxel j was defined as the proportion of fibres flowing through voxel i that reach voxel j. (Behrens et al., 2007) The fraction

of sampled fibres in all voxels in a region i that reach any voxel in region j in an AAL region i determines the connection probability P_{ij} between those regions. Due to the dependence of tractography on the seeding location, the probability P_{ij} is not necessarily equivalent to P_{ji}. However, these two probabilities are highly correlated across the brain for all participants (the least Pearson $r = 0.70$, $p < 10-50$), and thus the unidirectional connectivity probability P_{ij} between regions i and j was defined by averaging these two connectivity probabilities. This connectivity was considered as a measure of the structural connectivity resulting in a 90×90 symmetric weighted matrix C representing the network organization of each individual brain. A group averaged structural connectivity matrix was obtained by averaging across all 16 healthy participants.

Schaefer 1000 As described in Deco and Kringelbach (2020), we used the Human Connectome Project (HCP) database, which contains diffusion spectrum and T2-weighted neuroimaging data. Specifically, we estimated the structural connectivity using HCP dMRI dataset provided by the Special HCP dMRI, which uses excellent protocols taking 89 minutes for each of 32 HCP participants at the MGH centre. A detailed description of the acquisition parameters can be found in the HCP website.(Setsompop et al., 2013). The precise preprocessing is described in details in Horn and colleagues (Horn et al., 2017), In brief, the data were processed by using a q-sampling imaging algorithm implemented in DSI studio (http://dsi-studio.labsolver.org). A white-matter mask was computed by segmenting the T2-weighted images and co-registering the images to the b0 image of the diffusion data using SPM12. 200.000 fibres were sampled within the white-matter mask for each HCP participant. Fibres were transformed into MNI space using Lead-DBS. (Horn and Blankenburg, 2016) Finally, we used the standardized methods in Lead-DBS to extract the structural connectomes from the Schaefer-1000 parcellation. (Schaefer et al., 2018) A group averaged structural connectivity matrix was obtained by averaging across all 32 healthy participants.

4.3 Data for Chapter 6

4.3.1 Participants

All 15 participants (four women; mean age, 30.5 ± 8.0) were recruited via word of mouth and provided written informed consent to participate after study briefing and screening for physical and mental health. The study was approved by the National Research Ethics Service committee London-West London and was conducted in accordance with the revised declaration of Helsinki (2000), the International Committee on Harmonization of Good Clinical Practice guidelines, and National Health Service Research Governance Framework. Imperial College London sponsored the research, which was conducted under a Home Office license for research with schedule 1 drugs. The screening for physical health included electrocardiogram (ECG), routine blood tests, and urine test for recent drug use and pregnancy. A psychiatric interview was conducted and participants provided full disclosure of their drug use history. Key exclusion criteria included: < 21 years of age, personal history of diagnosed psychiatric illness, immediate family history of a psychotic disorder, an absence of previous experience with a classic psychedelic drug (e.g., LSD, mescaline, psilocybin/magic mushrooms or DMT/Ayahuasca), any psychedelic drug use within 6 weeks of the first scanning day, pregnancy, problematic alcohol use (i.e., > 40 units consumed per week), or a medically significant condition rendering the volunteer unsuitable for the study.

4.3.2 MRI data acquisition

The parcellated timeseries in the AAL90 parcellation were kindly provided by Gustavo Deco's, and further details are provided in Deco et al. (2018b). In brief, two BOLD-weighted fMRI data were acquired with a gradient echo planer imaging sequence, field-of-view = 220mm, 64 3 64 acquisition matrix, TR/TE = 2000/35ms, parallel acceleration factor = 2, 90 flip angle. Thirty-five interleaved oblique axial slices were acquired, each 3.4mm thick with zero slice gap (3.4mm isotropic voxels). The precise duration of each of the two BOLD scans was 7:20 min.

4.3.3 Structural connectivity

Structural connectivity data between the 90 AAL90 regions was obtained from 16 healthy subjects (5 female) using diffusion Magnetic Resonance Imaging (dMRI), registered to the MNI space and parcellated according to the AAL atlas on the subject's native space. Then, for each subject the histogram of fiber directions at each voxel was obtained, yielding an estimate of the number of fibers passing trough each voxel. The weight of the connection between regions i and j was defined as the proportion of fibers passing through region i that reach region j, which (since the directionality of connections cannot be determined using dMRI) yields a symmetric 90×90 structural connectivity (SC) matrix. Finally, the SC matrix of each subject was thresholded, pruning any connection lower than 0.1%, and SC matrices of all subjects were averaged to obtain a single SC matrix used in all simulations. This average matrix was kindly provided by Gustavo Deco *et al.*(Deco et al., 2018b)

4.3.4 PET receptor maps

For the density map of 5HT2A receptors, we used the Positron Emission Tomography (PET) data made public by Beliveau *et al.*,(Beliveau et al., 2017) which can be combined with the AAL90 parcellation to obtain an estimate of receptor density in every AAL90 region. Further details can be found in the original publications.(Deco et al., 2018b; Beliveau et al., 2017)

4.4 BOLD signals processing

The detailed BOLD preprocessing pipeline is provided in Deco et al. (2018b). For completeness, we describe critical preprocessing steps: data were corrected for head motions, finding that motion was similar in both conditions. The BOLD signals obtained for each region of the AAL90 was computed as the average of all the voxels defined within that region.

4.4.1 Filtering

Empirical and simulated BOLD signals (empirical or simulated) were filtered with a Butterworth (order 2) band-pass filter in the 0.01-0.1 frequency range.

4.4.2 Functional connectivity and functional connectivity dynamics

The functional connectivity (FC) matrix was obtained computing the Pearson correlation coefficient between all the pairs of simulated or recorded BOLD signals. The functional connectivity dynamics (FCD) was obtained by computing the FC(t), where t is given by consecutive sliding windows of length TR=30 and TR=28 of overlap. Then, we vectorized each of the FC(t) by taking the upper triangular, and finally computed the Pearson correlation coefficient between all these vectors to obtain the FCD matrix.

4.5 Null network models of the human connectome

Null network models of the human structural connectivity were used to evaluate the role of the local connectivity properties(Rubinov and Sporns, 2010) on the local changes in entropy induced by 5HT2A-R activation. To this end, we applied three different randomisation schemes to the structural connectivity in order to produce suitable surrogate networks. These surrogates were designed to preserve different network attributes of the original connectome: i) the overall density and strength (RAND), ii) the degree distribution (degree-preserving randomisation [DPR]); and iii) the strength distribution (strength-preserving randomisation [DSPR]) After randomisation, the DMF model was run with and without 5HT2A-R activation, and entropies were estimated. Every surrogate model was run 120 times, and the results averaged across runs. Null models were obtained using the Brain Connectivity Toolbox (Rubinov and Sporns, 2010).

NEURAL MASS MODELLING FOR THE MASSES: DEMOCRATISING ACCESS TO WHOLE-BRAIN BIOPHYSICAL MODELLING WITH FASTDMF

5.1 Introduction

Recent advances in non-invasive brain-imaging technology provide a fertile ground to investigate how the anatomical structure of the brain shapes complex neural dynamics, in healthy, pathological, and specific experimental conditions. As a matter of fact, the high versatility of the available imaging modalities has triggered a plethora of research efforts, which in turn have delivered important advances in human neuroscience (Deco and Kringelbach, 2020; Munn et al., 2021; Cofré et al., 2020).

One of these advances, as described in Chapter 3, include the dynamic mean field model (DMF), which by combining different neuroimaging modalities enables to simulate with biophysical detail the whole-brain activity in different conditions. Recently, the DMF has become one of the most popular biophyiscally-grounded whole-brain models, used to simulate BOLD signals measured during an ample range of different conditions and global brain states (Deco et al., 2018b; Demirtaş et al., 2019; Wang et al., 2019; Herzog et al., 2020; Kringelbach et al., 2020; Luppi et al., 2021; Gatica et al., 2021).

Despite these achievements, a few limitations prevent the DMF model from be-

ing more widely exploited by the scientific community. The main limitation is that current publicly available implementations of the DMF model are computationally expensive, with demanding time and memory requirements. This problem is exacerbated by two further aspects of the DMF: i) the need to calibrate the model's feedback inhibition control (FIC) parameters, which stabilise the firing rates of E pools, and have been shown to provide more realistic activity (Deco et al., 2014) and richer dynamics (Hellyer et al., 2016); and ii) the need to optimise the model's free parameters to fit statistical features of a given empirical data set. Both aspects involve running a large number of long simulations and thus inflate the overall computational costs, which often make the requirements of the DMF prohibitive for researchers with no access to high-performance computing infrastructure. This represents an undue obstacle to the ability of neuroscientists across the globe to contribute to brain modelling research: the choice of an appropriate computational model should depend only on its suitability to address specific research question, rather than being contingent on the availability of high-performance resources.

Moreover, the high computational burden of the current implementations of the DMF does not only restrict its usage, but also severely restricts the spatial resolution of what can be simulated. The existent implementations of the DMF are based on parcellations with less than 100 regions, –typically given by the Automatic Anatomical Labelling (AAL90)(Tzourio-Mazoyer et al., 2002) or the Desikan-Killiany(Desikan et al., 2006) parcellation– because, otherwise, it would be unfeasible to simulate brain activity based on fine-grained parcellations. However, recent advances in neuroimaging data analysis involve measuring brain activity at multiple spatial scales, from coarse to fine-grained, revealing new insights about underlying brain dynamics(Deco and Kringelbach, 2020) and its relevant operational scales(Kobeleva et al., 2021). Therefore, being restricted to simulating biophysically realistic brain activity at a coarse-grained spatial scale represents another barrier that hinders the ability of neuroscientists to leverage the full potential of DMF modelling.

To address both of these limitations, here we present *FastDMF*: a time- and memory-efficient implementation of the DMF model, which reduces its computational demands so it can be run and fit efficiently on any contemporary desktop computer, dramatically widening access to this kind of computational modelling approach. The FastDMF implementation is built by leveraging a number of key advances:

- It provides an improved implementation of the DMF model, which is significantly faster and reduces memory consumption by several orders of magnitude.
- It uses a novel connectome-dependent local inhibitory feedback mechanism, which replaces the standard FIC optimization problem and radically reduces the number of FIC calibration parameters.
- It leverages a Bayesian optimization algorithm, which substantially reduces the number of simulations required to fit the model to empirical data. This algorithm performs a smart sampling of the parameter space instead of the grid search used in previous approaches(Deco et al., 2018b; Luppi et al., 2021).

These advances make the FastDMF a computationally efficient, easily accessible, and biophysically grounded whole-brain model.

To showcase the advantages of the FastDMF, we first compare our implementation to the current one (published with Ref(Deco et al., 2018b)) in terms of computational costs. Then, we show analytical and numerical evidence supporting a connectome-dependent linear solution to the FIC optimization problem. Finally, using the Bayesian optimization algorithm, we fit an fMRI dat aset parcellated at two different spatial scales (N=90 and N=1000 regions). Using less than 350 optimization iterations, which can be performed in personal computers in matter of minutes, we achieve an excellent agreement between empirical and simulated data for both scales, showing the performance and broad applicability of the FastDMF to different scenarios. Furthermore, to facilitate the use of FastDMF by the community, we provide all codes that implement FastDMF. FastDMF is coded in C++ and is usable on both Matlab and Python, and can be found on `https://gitlab.com/concog/fastdmf`.

5.2 Results

5.2.1 Fast and efficient computational implementation

The first advancement to make the FastDMF widely usable is to provide an open-source implementation with two main advantages: faster execution time and less memory usage. This is achieved through three main improvements: i) The core of

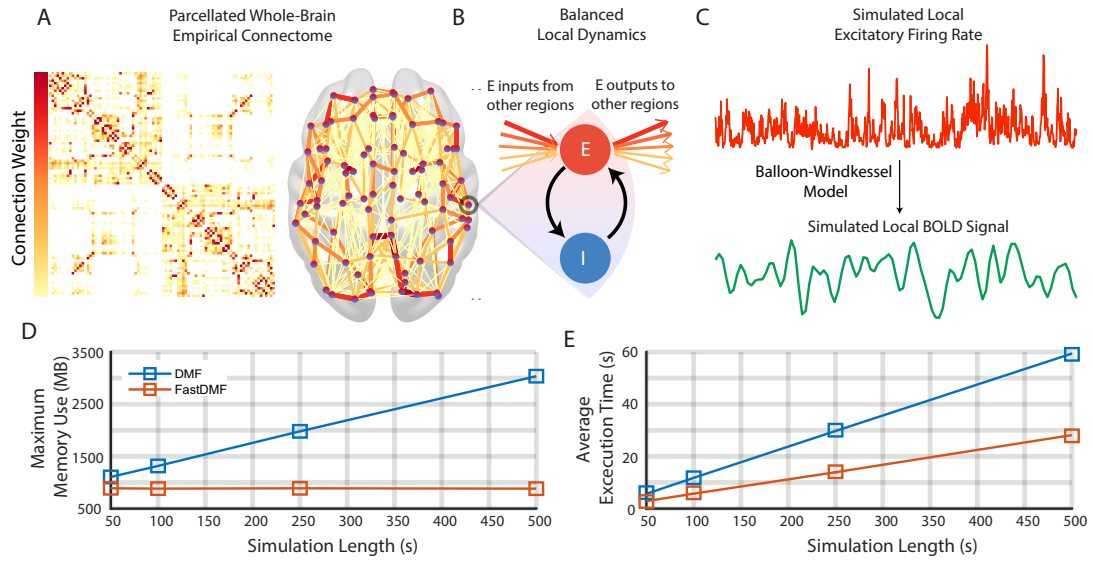

Figure 5.1: **Simulations using the fast dynamic mean field model (A)** Brain regions are defined by a parcellation (here, the AAL90) and their connections are empirically obtained from DTI (see Methods for details). In the schematic brain, circles are brain regions and connections between brain regions are colored according to their connection weight in the connectome. **(B)** The local dynamics of each brain region are simulated by recurrently connected pools of excitatory (E, red circle) and inhibitory (I, blue circle) neuronal populations. In turn, brain regions are connected through the E pool, such that the excitatory inputs from other regions are weighted according to their respective connection weight, and their sum is scaled by the Global Coupling parameter G. The connection from the I pool to the E one - J_n, the local inhibitory feedback - compensates for the excess of excitatory activity injected from other regions. **(C)** The simulated firing rate of each excitatory pool is used to generate BOLD-like signals for each brain region using the Balloon-Windkessel model. **(D)** Maximum memory used by the current implementation (DMF) and by ours (FastDMF) to simulate BOLD signals of different length. Both implementations used the connectome shown in **A**. **(E)** Same than **D**, but for the average execution time among 10 repetitions.

the FastDMF is written in C++ and takes advantage of the advanced linear algebra library Eigen (Guennebaud et al., 2010), which provides a fast and cross-platform toolkit for numerical operations. ii) The FastDMF and the Balloon-Windkessel hemodynamic (BW) function to compute the BOLD signals run in parallel, further reducing the overall execution time of the simulation. iii) The DMF and the BK numerical integrators access a shared memory via a simplified producer-consumer architecture, which, given the large difference in timescale between the BOLD and the firing rate signals, allows the FastDMF to radically reduce memory usage through

shared finite-size buffer.[1]

To benchmark our implementation, we ran simulations of varying lengths using the AAL90 structural connectome(Tzourio-Mazoyer et al., 2002) and measured both memory usage and execution time. As expected, the FastDMF was faster and more memory-efficient than the public Matlab implementation (Deco et al., 2018b). Thanks to the circular buffer, memory is effectively constant for all simulation times (Fig. 5.1D) and runtime per second of simulated activity is less than half of the Matlab DMF (Fig. 5.1E).

Finally, in addition to the underlying C++ implementation, the FastDMF library incorporates interfaces for both Matlab and Python, using the Mex(Mehrmann et al., 1999) and Boost Python(Abrahams and Grosse-Kunstleve, 2003) libraries respectively. These enable researchers to run and analyse simulations from a language of their choice, thus making the integration of the FastDMF in their existing pipelines easier.

5.2.2 Linear solution to the FIC optimization problem

The second main improvement was to provide a linear solution to the optimization of the feedback inhibition control(Deco et al., 2014) (FIC). In brief, the global coupling (G) parameter of the DMF scales the inputs that each region receives from the rest of the network, allowing to tune the model closer/farther to an optimal working point, where some desired statistical feature of the empirical neuroimaging data is reproduced, such as the functional connectivity (FC) or the functional connectivity dynamics (FCD). To compensate for the excitation that each pool receives from the other regions via the connectome, a FIC parameter is optimized through recursive adjustments to clamp the firing rate within a neurobiologically plausible range of 3-4 Hz for each local excitatory neural population, preventing the system from entering a hyper-excitation regime. The solution to this problem, based in an iterative process of increasing and decreasing the local inhibition until convergence, is explained in detail in Ref. (Deco et al., 2014), and it is implemented in the scripts accompanying the publication of Ref. (Deco et al., 2018b), and gives the optimal local

[1] In other words, the DMF equations are used to simulate firing rates, which are temporarily stored in a buffer, and then consumed by the BW integrator to generate the slow BOLD signals. Once a batch of firing rates has been consumed by the BW integrator, the DMF integrator is notified and allowed to write on the same memory address, and the cycle begins again. This mechanism makes the total memory usage of the simulation grow with the size of the resulting BOLD signals, as opposed to the firing rate signals, which are approximately 1000 times larger in memory.

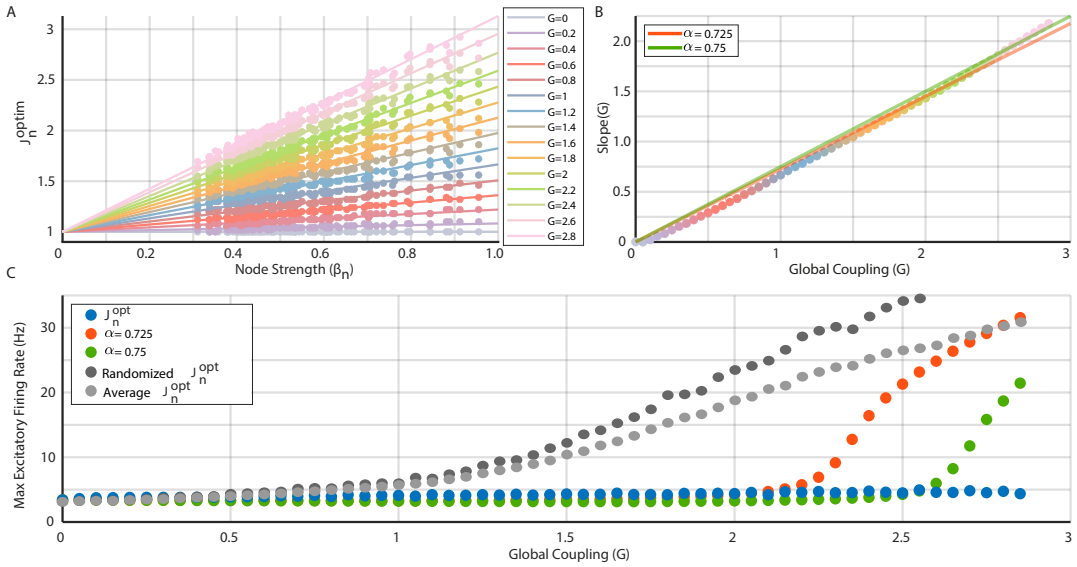

Figure 5.2: **Optimal local inhibitory feedback depends on global coupling and local anatomical connectivity strength** **(A)** Optimal local inhibitory feedback (J_n^{opt}) is well approximated by a linear function of region strength (β_n) for a wide range of G values (colors). **(B)** The slope of the J_n^{opt} vs β_n scales linearly with the global coupling, giving an optimal slope of 0.725(red solid line), however, using 0.75 (green solid line) better matches the high G values. **(C)** J_n^{opt} attains stability for all G values between [0,3], while the linear approximation using $\alpha = 0.725$ diverges to the high excitability regime close to G=2 (red dots). Using $\alpha = 0.75$ (green dots) attains stability in a larger range of G values, diverging at $G \sim 2.5$. To illustrate the topographical specificity of J_n^{opt}, we used a randomized version of J_n^{opt} (dark gray circles) and an homogeneous J_n equal to the average J_n^{opt} (light gray circles), both fails to control the global dynamics for increasing values of G.

inhibitory feedback $N-$dimensional vector (J^{opt}) for a given G, where N is the number of regions. However, for a standard setup these scripts took more than 2 months on a 32-core cluster to converge. Even worse, running the model using different connectivity matrices would require to separately optimize the FIC for each one of them, severely hindering the endeavour of investigating how structure shapes function through biophysical DMF modelling.

Despite the lack of prior information regarding the expected J^{opt} in the DMF models, results from other type of whole-brain models suggest that J^{opt} is correlated with local connectivity measures of the structural connectivity (Hellyer et al., 2016). Accordingly, here we show that a first-order analytical solution for a stationary state where the expected values of firing rates are fixed to 3.4 Hz (average excitatory firing rate for G=0) predicts that J^{opt} is region-specific, and its magnitude is proportional to G and to the local connectivity strength $\beta_n := \sum_p C_{np}$ is the strength of node n

computed from the anatomical connectivity matrix C. (see Chapter 4, subsection 4.1.2).

We write this solution as

$$J_n^{opt} = \alpha\, G\, \beta_n + 1\,, \tag{5.1}$$

where α is a scaling factor that represents the global excitation-to-inhibition (E/I) ratio (see Chapter 4, subsection 4.1.2). According to this approximation, for a fixed G, α is the only parameter to estimate in order to solve the feedback inhibitory control problem.

As a previous well-known and publicly available implementation of the DMF model Ref. (Deco et al., 2018b) used the AAL90 parcellation, we numerically tested our first-order approximation using this parcellation. After the optimization converged, for each G, we plotted J_n^{opt} vs the local connectivity strength (β_n), finding a linear relationship whose slope linearly increased with G (Fig. 5.2A,B). Then, to estimate α of Eq. 5.1, we found the scaling factor between the slope and G, with the constraint that the slope for $G = 0$ should be 0. We used least squares to find the optimal value of α (0.725). However, this value gives lower goodness of fit for higher G values, which are usually the values where the model better fits empirical data (Deco et al., 2014, 2018b). Accordingly, we used weighted least squares, giving 10 times more weight to G values larger than 2.1 (close to bifurcation with $\alpha = 0.725$), finding an optimal value for $\alpha = 0.75$. This approach, as expected, better matches the slope values for high G values and also extends the range where stability can be attained by the linear approximation (Fig. 5.2C, green dots).

Note that a second-order approximation can be also used for the relationship between J_n^{opt}, G, and β_n, but our numerical simulations show that the range where stability is attained by second-order approximations is reduced, in comparison to the first-order one.

Finally, as a control, we used two modified versions of J^{opt} to run the DMF model: i) a randomized (shuffled) version of J^{opt}, and ii) a homogeneous version, where J_n for all regions is equal to the average of J^{opt} (Fig. 5.2C). We found that both alternative versions of J^{opt} fail to attain stability for high values of $G > 0.5$), showing that J^{opt} is region-specific.

Our results alleviate the huge amount of simulations required to properly tune the FIC, reducing the number of free parameters of the optimization problem from

N to one. In addition, we provide a biophysical interpretation of FIC in terms of the local connectivity strength and the global E/I balance.

5.2.3 Reproducing the FCD of resting state fMRI data for two different spatial resolutions

The third major improvement was to implement a Bayesian optimization algorithm to find the optimal set of parameters that reproduces a statistical feature of the empirical data (or more than one). The current DMF implementation finds the optimal working point by using a sub-optimal strategy based on grid search. Bayesian optimization algorithms, on the other hand, estimate the objective function by sampling the parameter space efficiently, finding the minimum of a wide variety of functions (Shahriari et al., 2016; Ulmasov et al., 2016). Here, following previous applications of the DMF to fMRI data(Deco et al., 2018b), we chose as objective function the Kolmogorov-Smirnov (K-S) distance between the pooled empirical FCD distribution (FCD_{emp}, computed using the FCD from all subjects) and the FCD histogram obtained with the DMF (FCD_{dmf}). (Hansen et al., 2015) In addition, instead of first optimising FIC to clamp the firing rates, and then optimising G to reproduce the FCD, we jointly optimised G and α, expecting that the optimal working point also satisfies the firing rates constraints. We ran the optimization until convergence to K-S distances was comparable to previous studies (Deco et al., 2018b; Luppi et al., 2021).

To exhibit the advantages of the FastDMF, in addition to the AAL90 parcellation used up to this point for consistency with previous work, we also obtained a different estimate of the healthy human connectome, based on high-resolution diffusion data (Van Essen et al., 2013) parcellated according to most fine-grained scale of the Schaefer atlas, which comprises 1000 functionally defined cortical regions (Schaefer et al., 2018). We also used the same parcellation to obtain the functional MRI timeseries, obtaining data sets in two different spatial scales (see Chapter 4, section 4.2).

We emphasise that the two connectomes employed for the following analyses were obtained from different groups of healthy individuals, with different acquisition parameters, different reconstruction methods and softwares, different atlases (anatomically *versus* functionally defined), different resolution (differing by one order of magnitude) and only one of which includes subcortical regions (AAL90). In

other words, we chose to use two healthy connectomes that differ on virtually every relevant methodological dimension, to establish the general validity of our results.

Indeed, we found comparable values of goodness of fit for the two spatial scales (K-S ~ 0.15, (Fig. 5.3A,F)). Note that a difference in the optimal G value between the two spatial scales is to be expected: not just because two different connectomes were used, but also mathematically because of the larger number of regions in the Schaefer1000, which in consequence increases the average β_n, reducing the magnitude of the global coupling required to drive the system to a high-excitability regimen.

Crucially, for both connectomes α is within the values shown in Fig. 5.2, demonstrating the robustness and generalisability of our estimation of this parameter, even when the empirical healthy connectomes used for the simulation originate from different cohorts and were obtained through different reconstruction approaches at different resolutions.

For both parcellations, we required 350 iterations to converge. Note that without this approach we would have to simulate a subset of parameter combinations several times, taking the risk of using a too coarse binning of the parameters space (to save time), thus missing the minimum. In addition, the Bayesian optimization algorithm can take advantage of parallel computing, making the optimization procedure even faster.

We further evaluated the Bayesian optimization results by using the optimal parameters to run 96 simulations with different initial conditions and a simulation length comparable to resting state recordings (~ 10 mins). First, we checked the average excitatory firing rates, finding that for both spatial scales they remained in the desired range (3-4 Hz, Fig. 5.3B,G)). However, given the increased number of regions of the Schaefer1000 parcellation, the average firing rates showed increased variance, reflecting more susceptibility to the initial conditions. Second, we checked that the best simulation closely matched the empirical FCD distribution (Fig. 5.4A,F), reproducing the bimodality of the empirical FCD distribution. Note that the resting-state FC matrix is also well captured (Fig. 5.4B,G), even when it was not included as an objective function in the fitting procedure. Finally, we computed the K-S between the FCD of each subject (FCD_{emp_s}) and the pooled FCD (FCD_{emp}) to evaluate the goodness of fit respect to the variability of the empirical population. Accordingly, we computed the FCD between each simulation (FCD_{dmf_s}) and FCD_{emp}, finding a close match between both distributions for both spatial scales. As a final check, we

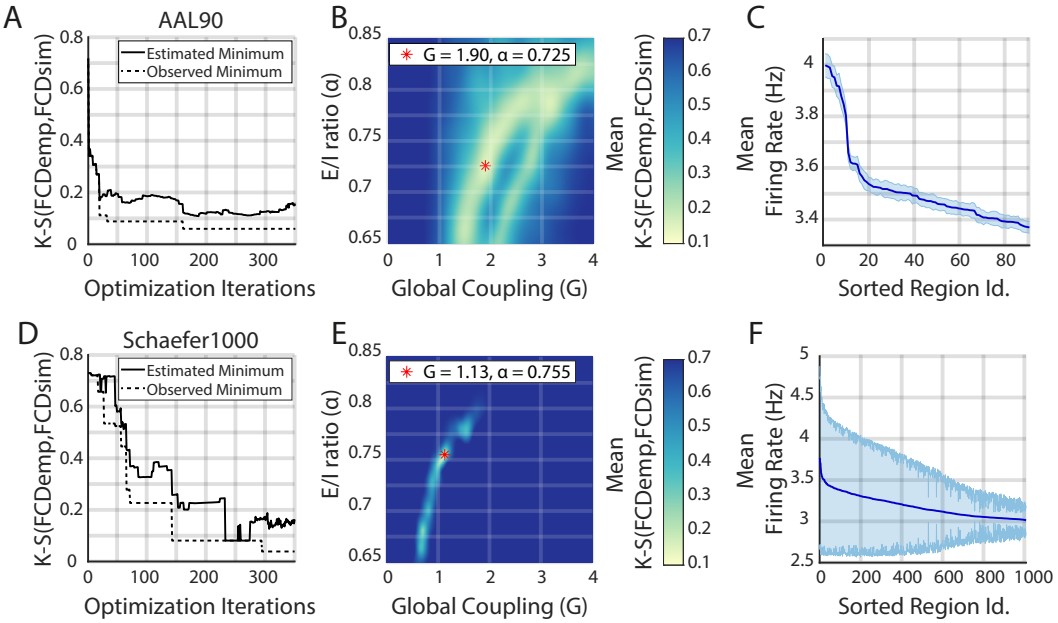

Figure 5.3: **Fitting the FastDMF model to empirical data with Bayesian Optimization**. **(A,D)** Convergence of the Bayesian optimization estimates of the minimum K-S distance (solid line), for the AAL90 and the Schaefer1000 parcellation, respectively. Dashed lines are the minimum observed K-S during the optimization process. **(B,E)** Bayesian optimization estimate of the mean K-S distance between the pooled empirical FCD (FCD_{emp}) and the simulated FCD (FCD_{sim}) as function of G and α. Red asterisk shows the optimal parameters. **(C,F)** Simulated brain regions sorted according to their average excitatory firing rate (solid line) with their respective standard deviation computed from 96 simulations (shaded area). All regions are within the expected range of firing rates.

also computed the mean squared error of the FC (MSE FC) as a measure of FC matrix similarity. This results shows that the optimal parameters generate simulations whose K-S values and FC matrices are close to those obtained from experimental subjects (Fig. 5.4C,D,H,I).

We highlight that this approach represents a considerable reduction in the number of simulations required to fit the model to empirical data, even more for the Schaefer1000 parcellation, which is prohibitive in current implementations. In addition, we remark that the optimal set of parameters that reproduces the empirical FCD also satisfies the local stability condition (i.e. stable firing rates), and also captures important structures of the FC matrix, showing the degree of generalization of the FastDMF model.

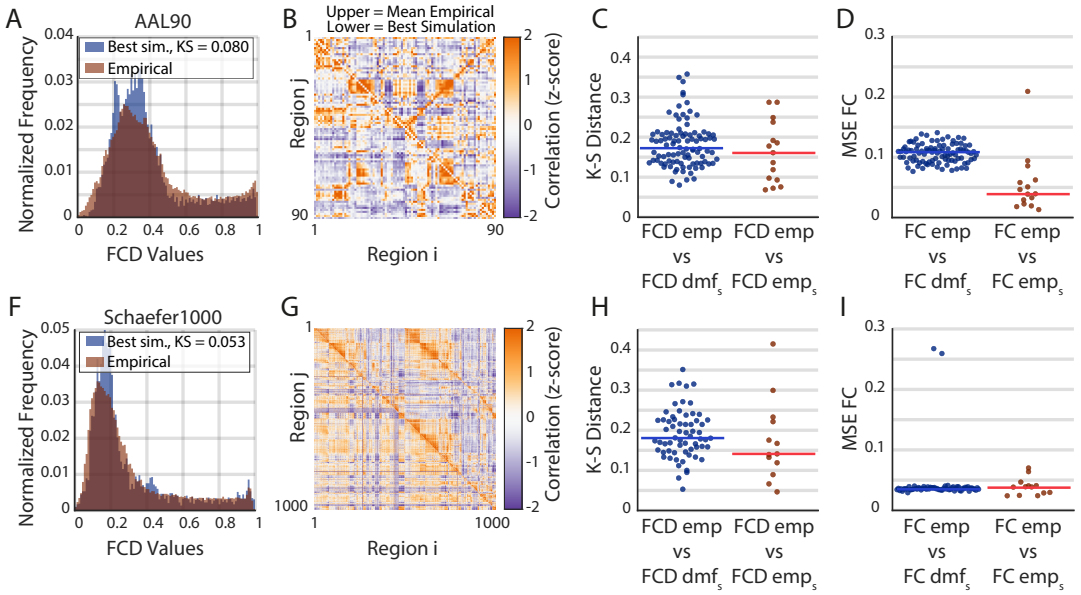

Figure 5.4: **Comparison between empirical and simulated FCD and FC for two spatial scales**. **(A,F)** FCD distribution for the pooled empirical data (red) and for the best simulation (blue) for the AAL90 and the Schaefer1000 parcellations, respectively. **(B,G)** FC matrix for the average empirical (upper triangular) and for the best simulation (lower triangular). **(C,H)** K-S distance between simulated, FCD_{dmf_s}, and pooled empirical FCD_{emp} (blue). The K-S distance between single subjects, FCD_{emp_s}, and pooled empirical FCD_{emp}, is shown in red. Horizontal line is the median. **(D,I)** Same than (**C,H**), but for mean squared error of the FC matrix. In this case, FC_{emp} is the average FC between all subjects.

5.3 Discussion

We introduced the FastDMF, an accessible, efficient, and biophysically grounded whole-brain model. Thanks to our combination of optimised implementation, a novel method of estimating the feedback inhibitory control, and efficient parameter search, we were able to fit empirical fMRI data using two parcellations representing drastically different spatial scales: the AAL90 (Tzourio-Mazoyer et al., 2002) and the Schaefer1000 (Schaefer et al., 2018) parcellations, highlighting the flexibility of our approach and its robustness to a variety of methodological choices. These examples show that our method provides accurate results for a wide range of G parameter values, thus outperforming existing tools for computing FIC. In terms of applications, the FastDMF opens the way for biophysically grounded whole-brain models of a range of novel applications such as turbulence analysis, connectome harmonics, or detailed perturbation protocols, which requires detailed and coarse-grained descriptions. This application could lead to promising novel insights about

brain dynamics and evoked responses. (Atasoy et al., 2016; Deco and Kringelbach, 2020; Deco et al., 2021; Escrichs et al., 2021).

Different alternatives have been proposed to model the FIC. A plasticity rule for a time-varying FIC has been used in a modified version of the DMF (Naskar et al., 2021), avoiding the optimization process, at the expense of increased model complexity. In fact, it has been suggested that the optimal FIC obtained by the iterative process Ref. (Deco et al., 2014) corresponds to the integrated form of the time-varying FIC (Hellyer et al., 2016). While our main purpose was to show the use of the FastDMF rather than explaining the possible mechanisms involved in the FIC or their potential functional implications, our model provides novel evidence in favor of a connection between the strength of the nodes in the connectome and FIC.

There is room for further improvement of our model. For example, a time-varying FIC could be implemented, following the approach proposed by Hellyer et al. (Hellyer et al., 2016). More general parameters allowing the incorporation of multiple receptor maps allowing to neuromodulate the activity of each brain region, principal gradient of gene expression or anatomical measures could be implemented in future developments(Kong et al., 2021). Additionally, new data modalities that can be derived from firing rates such as local field potentials or EEG-like could be future avenues of development.

The study of whole brain activity is an exciting research avenue that needs to be better developed to understand the causal interplay between brain structure and function in different brain health conditions, addressing the specific relevance of factors such as neuronal dynamics, neurotransmitter receptor density, and anatomical connectivity, among others. This, in turn, might accelerate brain research, identifying the biophysical mechanistic principles that relate different levels of brain organization, opening in this way the road for the development of new treatments to prevent and cure brain disease (Deco and Kringelbach, 2014; Cofré et al., 2020). In line with this perspective, the FastDMF provides a general and accessible tool of simulation and analysis to be applied in multiple neuroimaging scenarios, that can be called from Python and Matlab, which should favor its massive use by the scientific community.

A MECHANISTIC MODEL OF THE NEURAL ENTROPY INCREASE ELICITED BY PSYCHEDELIC DRUGS

6.1 Introduction

Psychedelic drugs provide a privileged opportunity to study the mind-brain relationship, and promise to revolutionise some of our current mental health treatments (Carhart-Harris et al., 2014; Scott and Carhart-Harris, 2019; Carhart-Harris et al., 2016a). However, while some aspects of the neurochemical action of psychedelics at the neuronal and sub-neuronal level are well known (Grasso et al., 2016; Wacker et al., 2017) our current understanding of their action at the whole-brain level is still very limited. A deeper understanding of the mechanisms that trigger the changes in conscious experience produced by psychedelics would greatly advance our knowledge of human consciousness, and medical development of psychedelics.

At a subjective level, serotonergic psychedelics (including LSD, dimethyltryptamine [DMT] and psilocybin, among others) potentially modulate mood, cognition, perception and self-awareness. At a neurophysiological level, recent research has identified (among many) two particularly prominent signatures of the psychedelic state: an overall dysregulation of neural population activity, most clearly seen as suppression of spectral power in the alpha (8-12 Hz) band (Carhart-Harris et al., 2012, 2016b; Timmermann et al., 2019); and an increase in the signal diversity of the neural activity, measured through the information-theoretic notion of entropy (Schartner et al., 2017). In particular, this acute entropy increase has

been linked to both short- and long-term effects of the psychedelic experience, including certain aspects of the reported subjective experience(Schartner et al., 2017) and subsequent personality changes (Lebedev et al., 2016).

Interestingly, the opposite effects have been reported for states of loss of consciousness, where a strong *decrease* in brain entropy has been repeatedly observed. This seems to be a core feature of loss of consciousness, generalising across states such as deep sleep,(Schartner et al., 2017) general anaesthesia (Zhang et al., 2001), and disorders of consciousness (Casali et al., 2013). Together, these facts are yielding converging evidence that entropy and related measures offer simple and powerful indices of consciousness.

Based on these findings, Carhart-Harris and colleagues have put forward the *Entropic Brain Hypothesis* (EBH) an entropy-based theory of conscious states (Carhart-Harris et al., 2014; Carhart-Harris, 2018). The EBH proposes the simple, yet powerful idea that the variety of states of consciousness can be indexed by the entropy of key parameters of brain activity; or, in other words, that the richness of subjective experience is directly related to the richness of on-going neural activity, where richness can be defined most simply as diversity. Note that even when different states can be indexed by their corresponding diversity, we are not claiming that one state is more conscious than other. Therefore, investigating the neurobiological origins of such changes in brain activity is a key step in the study of altered states of consciousness.

Multiple experiments in humans and animal models have established that the mind-altering effects of psychedelics depend on agonism specifically at the serotonin 2A receptor (5HT2A-R) (Barrett et al., 2018; Kraehenmann et al., 2017; Quednow et al., 2012). A recent simulation study involving whole-brain computational modelling confirmed that the topographic distribution of 5HT2A-R in the human brain is critical to reproduce the functional connectivity dynamics (FCD) of human fMRI data recorded under the acute effects of LSD (Deco et al., 2018b). Here we build upon this model to characterise the interplay between the entropy of brain signals, the distribution of 5HT2A receptors, and structural connectivity properties of the brain, with the overarching goal of explaining the sources of entropic effects, and thus altered consciousness, elicited by psychedelic drugs.

6.2 Results

We simulated whole-brain activity using the dynamic mean field (DMF) model by Deco *et al.* (Deco et al., 2014), with the addition that both excitatory and inhibitory neuromodulatory gains are modified, rather than only the excitatory one, as in Ref. (Deco et al., 2014). For details, see Chapter 4, subsection 4.1). Here, the DMF model combines a theoretical model of neural and synaptic dynamics with two empirical sources of information about the human brain: the human connectome, i.e. DTI-estimated connectivity between the 90 regions of the AAL90(Tzourio-Mazoyer et al., 2002) atlas; and average 5HT2A-R expression across the human brain obtained with positron emission tomography (PET) scans (Beliveau et al., 2017).

Each simulation of the DMF model generates 90 time series of excitatory and inhibitory firing rates, one of them for each region of the AAL90 atlas (Fig. 6.1A). These firing rates have a non-linear dependency on the local inputs, determined by a frequency-current (*F-I*) curve. Given that 5HT2A-R seems to modulate the sensitivity of neurons rather than driving them towards excitation or inhibition (Aghajanian and Marek, 1997; Licata et al., 1993; Grasso et al., 2016), we followed Deco et al. (Deco et al., 2018b) and modelled 5HT2A-R activation as a response-gain modulation (Abbott and Chance, 2005) of this F-I curve dependent on the receptor density at a given region (Fig. 6.1B).

We simulated the model in two conditions, with and without 5HT2A-R activation, which, in an analogy with neuroscientific terminology, we refer to as placebo (PLA) and 5HT2A-R conditions, respectively. In this way, we obtain 90 time series of excitatory firing rates for each condition, which we keep for further analysis.

Finally, using these time series, we estimate Shannon's differential entropy for each region under both conditions, yielding a topographical distribution of entropy values (Fig. 6.1C) that constitutes the main subject of analysis in this study. Further details about model specification and entropy estimation, as well as other methodological caveats, are presented in Chapter 4.

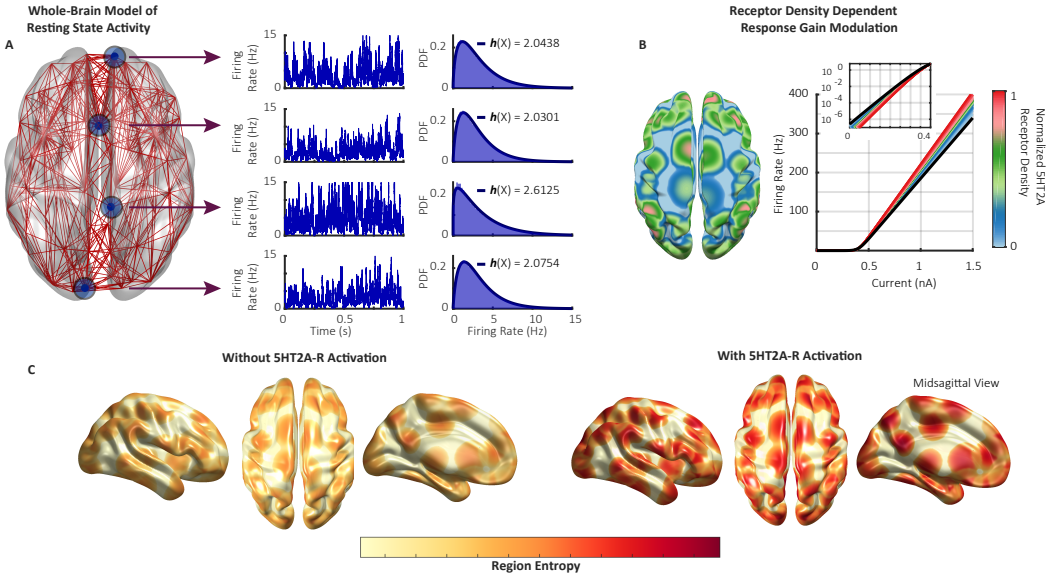

Figure 6.1: **Modelling the effect of 5HT2A-R activation on the whole-brain topographical distribution of entropy. (A)** Resting state activity is simulated using the Dynamic Mean-Field (DMF) model, in which each region's activity is represented by a time series of excitatory firing rates (constrained to 0-15 Hz for visualisation). The probability density function (PDF) and differential entropy ($\boldsymbol{h}$(X)) of each region is then estimated, obtaining a topographical distribution of entropy values. **(B)** 5HT2A-R agonism is modelled as a receptor-density-dependent response gain modulation. Black line is the frequency-current (F-I) curve of a population without 5HT2A-R agonism, and coloured curves show the resulting F-I curves of regions with increasing 5HT2A-R agonism. **(C)** 5HT2A-R activation changes the topographical distribution of entropy with respect to resting state activity, which constitutes the main subject of analysis in this study.

6.2.1 The DMF model with 5HT2A-R activation on excitatory and inhibitory populations reproduce empirical FCD of LSD

A Bayesian optimization algorithm (Chapter 4, subsection 4.1.3) was used to simultaneously optimize the global coupling parameter G, and α, a global parameter that controls the local inhibitory feedback parameter. The cost function was defined as the Kolmogorov-Smirnof distance (K-S) between the empirical and the simulated functional connectivity dynamics (FCD) distribution. Resting state BOLD signals of the placebo condition were used to fit the model associated to the non-drug condition. The data set and processing was exactly the same than the one used in Deco et al. (2018b) (see (Chapter 4, subsection 4.3 for details). The optimal average K-S (0.198 ± 0.068) was found at G=2.4 and α 0.75 (Figure 6.2A), where also the average firing rates fell within the biologically plausible (Deco et al., 2014) 3-4 Hz range (Fig-

ure 6.2B). Once the model was fit to the placebo data, the neuromodulatory gain for the excitatory and the inhibitory population were optimized such that the K-S of the FCD between the model and the data associated to LSD BOLD signals was minimized. There was several regions where the K-S were minimized (Figure 6.2D), so, for simplicity, we only used solutions were both neuromodulatory had the same magnitude, obtaining an average K-S of 0.120 ± 0.082. Note that using other possible minimum of the objective function shown in Figure 6.2D produces implausible firing rates and/or decreases in the entropy. Then, we compared the K-S values of the models when fit to their respective data sets (Figure 6.2E), finding an improvement of the goodness of fit of the model with neuromodulation to LSD data. Despite the increased response gain in both populations, the excitatory firing rates increased, but they were still within the 3-4 Hz range. (Figure 6.2F).

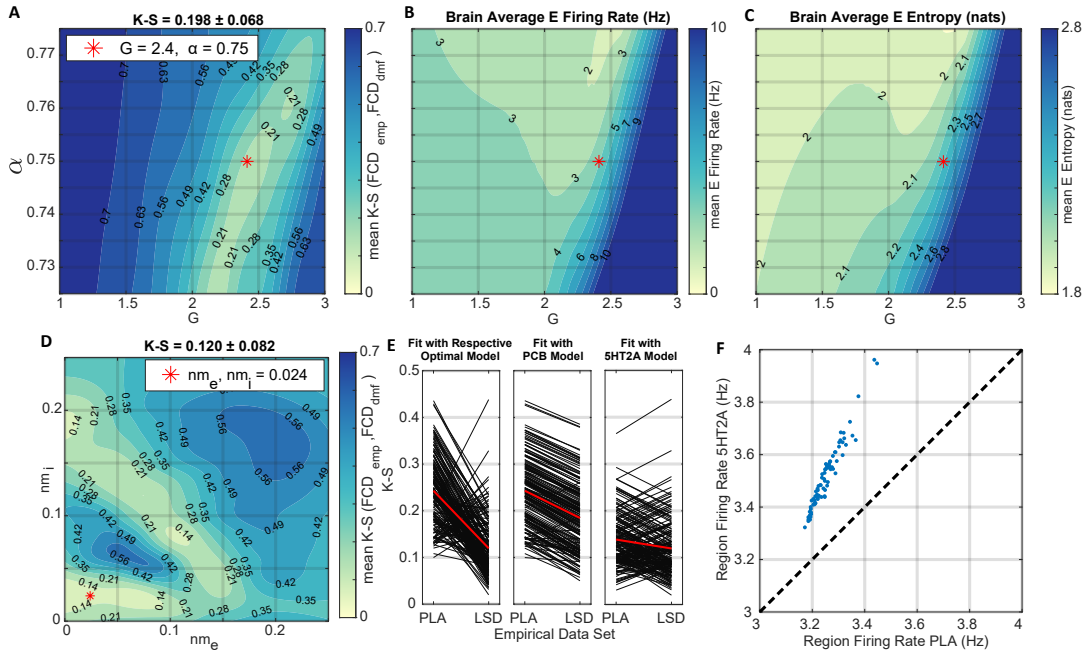

Figure 6.2: **Fitting DMF model to empirical BOLD data.** (**A**) Estimated average K-S distance between the empirical FCD obtained from placebo resting state condition and the FCD obtained with the model. G and α were optimized by Bayesian optimization. Red asterisks denotes the selected optima. The corresponding average excitatory firing rates, and entropies are shown in **B** and **C**, respectively. (**D**) Estimated average K-S distance between the empirical FCD obtained from LSD resting state condition and the FCD obtained with the model with neuromodulation (5HT2A). nm_e and nm_i were optimized by Bayesian optimization. (**E**) The model with neuromodulation improves the goodness of fit to empirical data. Black lines are simulations with the same initial condition, and red line is the average. (**F**) Scatter plots of the excitatory firing rates in both simulated conditions.

6.2.2 5HT2A-R activation causes a heterogeneous, linear increase in the entropy of simulated brain activity

We used the DMF model to test the main prediction of the EBH: that 5HT2A-R activation causes an increase in the overall entropy of neural signals (Fig. 6.3). In line with previous experimental findings with psychedelic drugs,(Schartner et al., 2017) the model shows a significant increase in the brain's entropy h due to 5HT2A-R activation, with an average entropy of $h^{\mathrm{PLA}} = 2.15\,\mathrm{nat}$ in the placebo condition, and $h^{\mathrm{5HT2A}} = 2.25\,\mathrm{nat}$ in the 5HT2A-R condition (dashed vertical lines in Fig. 6.3B, Wilcoxon signed-rank test $p < 10^{-6}$, Cohen's $d = 0.98 \pm 0.05$).

A closer look at the distribution of entropy changes, however, reveals a more nuanced picture, entropy increases for all regions, but the effect is larger for visuo-occipital regions, compared to the rest of the regions (Fig. 6.3D and Supp. Fig. 2). This suggests that, according to the model, 5HT2A-R agonism might trigger a complex reconfiguration of the topographically specific distribution of entropy, and not a mere homogeneous overall increase.

To investigate this reconfiguration, we analysed the local changes in entropy by plotting the entropy of the n^{th} region in both PLA and 5HT2A conditions, denoted by h_n^{5HT2A} and h_n^{PLA} respectively (Fig. 6.3D). Our results show that 5HT2A-R activation affects local entropy in a non-uniform and linear manner, especially in regions with high baseline resting state entropy. In particular, in such regions the effect of 5HT2A-R activation show a different tendency than the rest of the regions, such that the local entropy in the 5HT2A-R condition could not always be determined by the region's baseline entropy. This finding hints towards a more general theme: that local dynamical properties alone are not able to explain the local changes in activity induced by 5HT2A-R agonists like psychedelic drugs. We will explore this phenomenon in depth in the following sections.

Next, we studied the effect of 5HT2A-R agonism on local entropy by considering the relative change scores,

$$\Delta h_n = \frac{h_n^{\mathrm{5HT2A}} - h_n^{\mathrm{PLA}}}{h_n^{\mathrm{PLA}}} .$$

Based on recent *in vivo* experiments with serotonergic psychedelics we expect to find localized entropy increases on occipital, cingulate and parietal regions

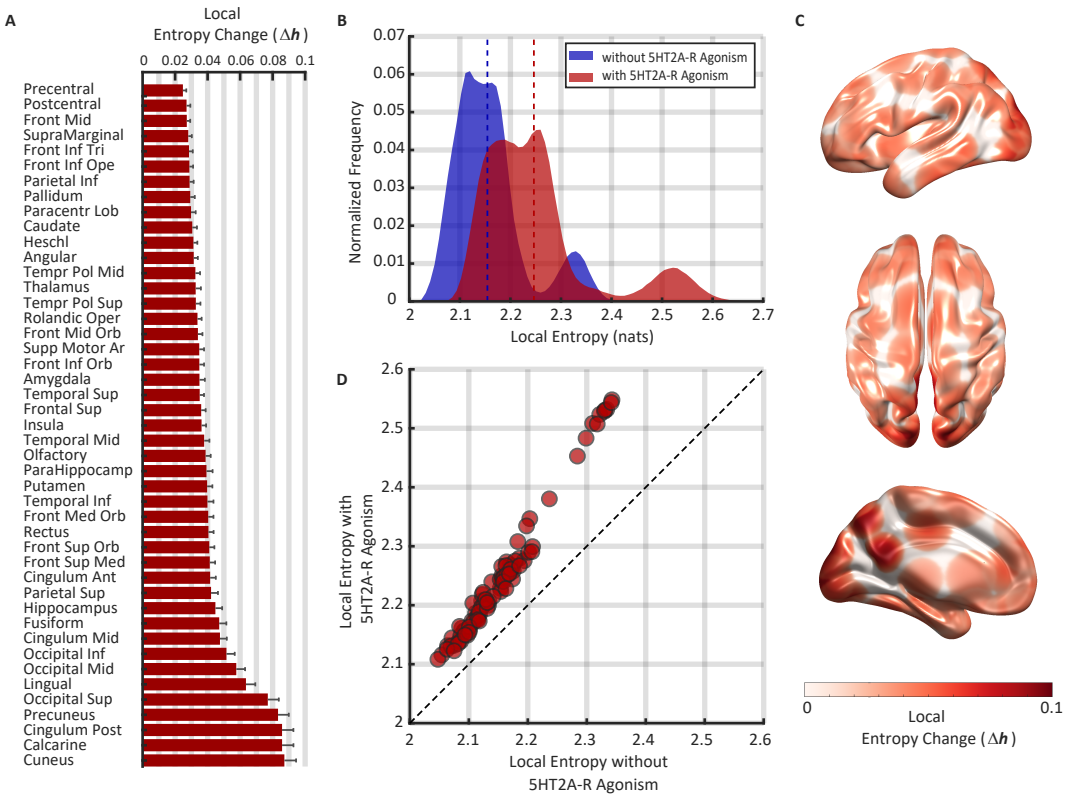

Figure 6.3: **Linear heterogeneous increase of entropy following 5HT2A-R activation. (A)** Effect of 5HT2A-R agonism on the local entropy each of region in the AAL atlas. See Supplementary Table 1 for abbreviations. Bars indicate the (bilateral) average relative change in local entropy, Δh_n, and error bars indicate 1 standard deviation across 1000 simulations. **(B)** Histograms of local entropy values for the condition with (red) and without (blue) 5HT2A-R activation. 5HT2A-R activation increased both the average and the spread of the local entropy distribution. **(C)** Topographical map of entropy changes. Brain regions are coloured according to their Δh_n values. **(D)** 5HT2A-R agonism changed the topographical distribution of entropy in linear manner. Each circle indicates the averages of each region across 1000 simulations.

(Lebedev et al., 2016; Schartner et al., 2017) as well as several changes on regions belonging to resting state networks (RSNs) (Carhart-Harris et al., 2016b; Tagliazucchi and Laufs, 2014; Lord et al., 2019). Accordingly, we found that the effect were localised specially in occipital areas, together with some cingulate and parietal regions, as expected. (Fig. 6.3A). Additionally, regions participating in the primary visual and default mode (DMN) RSNs – including occipital and cingulate areas, respectively – showed a marked tendency to increase their entropy, as well as regions participating in the frontoparietal executive network (FPN). Further discussion about the relation between these results and recent *in vivo* experiments with psychedelic drugs is presented in the Discussion section.

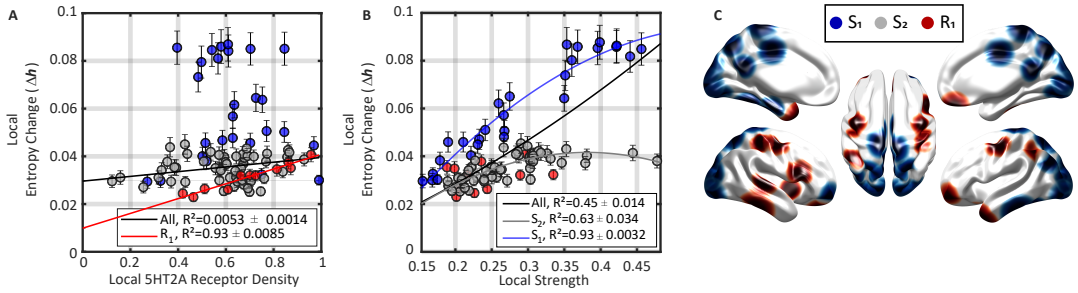

Figure 6.4: **Changes in local entropy are explained best by connectivity strength, then receptor density**. **(A)** Changes in entropy were overall independent from receptor density, although **(B)** they were well predicted by the connectivity strength of each region. We split into strength (blue and gray), and receptor dependent groups (red). The S_1 and S_2 groups showed no significant relationship with receptor density, while the R_1 group were highly correlated with it. **(C)** Topographical localisation of the three groups, following the same colour code. S_1 were mainly located in occipital, parietal and cingulate regions, while the R_1 ones were in temporal and frontal ones.

6.2.3 The topography of entropy changes is explained by local connectivity strength and 5HT2A-R density

Given the results above, and given the potential clinical and neuroscientific relevance of entropy reconfiguration in the psychedelic state, our next task is to elucidate what neurobiological factors underlie such phenomenon. Here we investigate which structural and dynamical features of the model are able to predict local entropy changes due to 5HT2A-R activation, and what we can learn about the large-scale action mechanism of psychedelic drugs and other 5HT2A-R agonists.

To explain the effects of 5HT2A-R activation, the first natural step is to factor in the density of 5HT2A-R at each region. Somewhat surprisingly, at a whole-brain level, receptor density was a very poor predictor of the entropy change due to 5HT2A-R activation (Fig. 6.4A, $R^2 = 0.0053 \pm 0.0014$). In contrast, we found that the connectivity strength (based on the DTI human connectome), defined as the sum of all the weighted links connecting a given region, exhibited a moderate correlation with local entropy changes (Fig. 6.4B, $R^2 0.45 \pm 0.14$).

By visual inspection, however, it is clear that the relationship between DTI-informed connectivity strength and Δh_n is complex and can be split in two separate tendencies. To investigate this phenomenon, we implemented a simple optimisation algorithm to find the set of regions with the strongest relationship be-

tween strength and Δh_n (see Chapter 4, subsection 4.1.5). This yielded a set of regions with a higher strength-Δh_n correlation (S_1, $R^2 = 0.93 \pm 0.0032$, as opposed to $R^2 = 0.45 \pm 0.14$ for the whole-brain), and induced a partition of all brain regions into three groups: those highly dependent on strength (S_1, 27 regions, blue in Fig. 6.4A)), those less dependent on strength (S_2, 42 regions , gray), and those highly dependent on receptors (R_1, 11 regions, red).

Building up on this partition, we studied the effect of receptor density on those regions where entropy change could not be predicted from connectivity strength – the R_1 group. We found a strong relationship between receptor density and entropy changes for the R_1 group, which resulted in positive correlation ($R^2 = 0.93 \pm 0.0085$). On the contrary, receptor density does not predict entropy changes for areas in the S_1 and S_2. This shows a complementary role of density and strength in S_1, S_2, and R_1 regions: entropy changes in R_1 regions strongly depend on the receptor density, but not connectivity strength; while changes in S_1, S_2 regions depend on connectivity strength, but not receptor density.

Overall, to assess the predictive power of connectivity strength and the receptor density on local entropy changes, we built a linear mixed model for Δh_n using connectivity strength, 5HT2A-R density, and the aforementioned three-way separation of brain regions as the predictor variables. Together these variables explain $94.5 \pm 0.002\%$ of the variance of Δh_n, confirming that they provide an accurate model for predicting the entropic effects of 5HT2A-R activation. This suggests that psychedelic drugs and other 5HT2A-R agonists do not have a simple, localised effect on brain activity, but instead amplify the fundamentally collective, emergent properties of the brain as a complex system of interacting elements.

6.2.4 The specific connectivity strength distribution explains relative changes in entropy

As a final analysis, we set out to investigate exactly which topological properties of the brain's structural connectivity explain the observed changes in entropy. With this exercise, we aim to answer two questions: whether any property simpler than connectivity strength *can* explain the results; and whether any property more complicated than connectivity strength *is needed* to do so.

To this end, we ran further simulations of the DMF model using suitable null network models of the human connectome (Fig. 6.5A), with the 5HT2A-R density

map held fixed. In particular, we used three null models designed to test increasingly strict null hypotheses that preserved different network attributes of the original connectome: i) the overall density and strength (RAND, Fig. 6.5B), ii) the degree distribution (degree-preserving randomisation [DPR], Fig. 6.5C); and iii) the strength distribution (strength-preserving randomisation, [DSPR], Fig. 6.5D). We simulated the DMF model in these surrogate networks, with and without neuromodulation, and computed the resulting entropy changes Δh_n.

Our first result is that random and degree-preserving surrogate networks are unable to reproduce the entropy changes observed in the human connectome: when compared against the Δh_n values obtained from the unperturbed network, neither of them produce values close to the original (Figs. 6.5E, 6.5F). This result asserts the findings in the previous section, showing that indeed node strength plays an important role in shaping the global pattern of entropy change associated with the action of psychedelics and other 5HT2A-R agonists. Simpler topological features, like the degree distribution, are not enough to explain such changes.

Perhaps more interestingly, the connectivity strength-preserving surrogate networks partially reproduced the entropy changes, suggesting that higher-order properties of the connectome are necessary (Fig. 6.5G). Together, these results show that, once the receptor distribution is fixed, the network strength distribution is *necessary* but not completely *sufficient* to explain the entropic effects of 5HT2A-R activation.

As previously suggested, other topological network features are known to mediate transitions of consciousness in other contexts,(Chennu et al., 2016) and investigating which network properties explain high-order dynamical signatures(Rosas et al., 2019) of psychedelics remains an exciting avenue of future work.

6.3 Discussion

In this study we investigated the brain entropy changes induced by serotonergic psychedelics by simulating whole-brain resting state activity with and without 5HT2A-R activation. In contrast to empirical studies, which usually only have access to coarse-grained fMRI or M/EEG data, our approach allows us to study firing rates of brain regions and hence to directly assess the effect of 5HT2A-R agonism at the level of neural population activity.

In agreement with the Entropic Brain Hypothesis,(Carhart-Harris, 2018) the

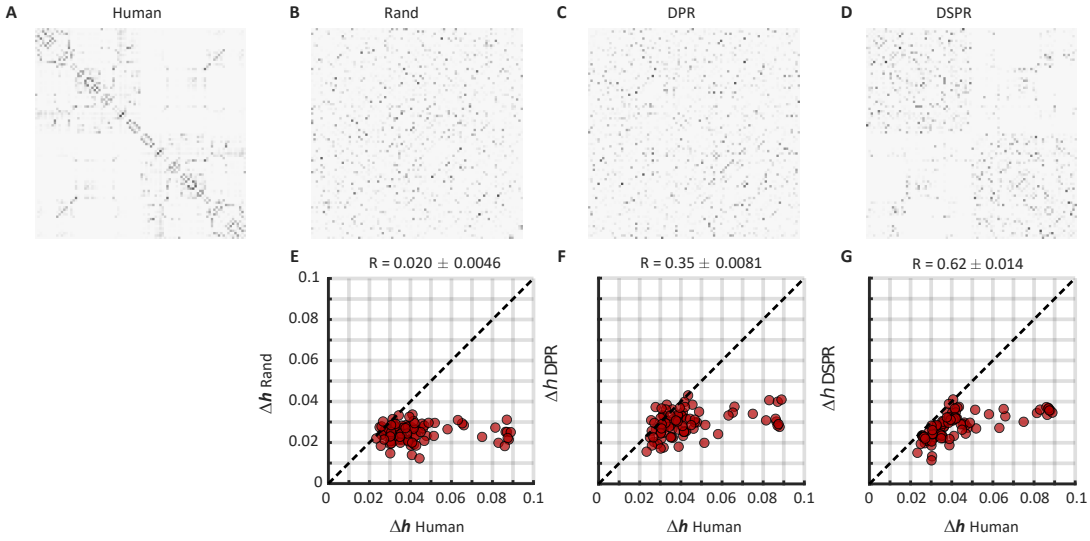

Figure 6.5: **Relative changes in entropy are partially reproduced by a strength-preserving null model of the connectome**. **(A)–(D)** Connectivity matrices used to control the role of local properties of the connectome on Δh_n. See main text for the description of the matrices and randomisation algorithm. **(E)–(G)**. Scatter plots of Δh_n for the human connectome against the three null models. DSPR yielded a high but not perfect correlation
showing that local network properties of human connectome are necessary but not completely sufficient to capture the effect of 5HT2A-R activation.

model shows a significant increase of global brain entropy, with a inhomogeneous and linear reconfiguration. The diversity of effects stresses the importance of extending the scope of the EBH from a simple level-based approach towards a multi-dimensional perspective, which might better characterise the richness of 5HT2A-R activation effect both on the brain and consciousness.

We propose a possible mechanism in the sense that the average increase in neural entropy induced by serotonergic psychedelic drugs can be explained (in statistical sense) in terms of quantities that provide a neurobiological interpretation about how this phenomena could occur in the brain. These results have important consequences for our understanding of consciousness, neurodynamics, and the psychedelic state. In the following, we highlight some of those implications, and propose possibilities for future work.

6.3.1 5HT2A-R-induced entropy changes are regionally heterogeneous

Our main result in this study is a mechanistic explanation of the main prediction of the EBH that serotonergic psychedelics increase the entropy of brain activity. However, one of the main takeaways is that this overall increase is tied to a spatially heterogeneous reconfiguration, rather than a globally consistent increase, in entropy.

Our simulation results show that 5HT2A-R activation triggers an entropy increase in sensory areas, as observed in occipital cortices, as well as the primary visual RSN. This agrees with the increased perceptual 'bandwidth' that characterises the psychedelic state,(Bayne and Carter, 2018) as higher entropy might be related to richer perceptual experience in a given moment of consciousness – potentially related to reduced gating. As a matter of fact, all the primary visual RSN regions are part of the S_1 group, suggesting that the effect of psychedelics on perceptual experience might be directly related to the connectivity of those regions. The localisation of the entropy increases may relate to domain-specific changes in consciousness, which could be interpreted as consistent with a recent dimensions of consciousness characterisation of the psychedelic experience.(Bayne et al., 2016)

6.3.2 Comparison to *in vivo* experiments with psychedelic drugs

Throughout the paper, we have focused on one particular signature of 5HT2A-R agonists on the brain – a global increase in average entropy. But much more is known about the effect of psychedelics on the brain, and studying these more nuanced effects is key to understanding the rich phenomenology of the psychedelic state. In this section we deepen this connection by providing a more detailed comparison between the behaviour of the model and experimental results with psychedelic drugs.

Our findings are in agreement with earlier studies where the effect of psychedelic drugs on the topographical distribution of entropy-related measures was correlated with subjective effects. For example, Schartner *et al.*(Schartner et al., 2017) studied the effect of LSD, psilocybin, and ketamine on the entropy rate of binarized MEG signals, and found localised increases in entropy rate on occipital and parietal regions. Similarly, Lebedev *et al.* (Lebedev et al., 2016) analysed the sample entropy of fMRI recordings of subjects under the effect of LSD, finding localized in-

creases on frontoparietal, medial occipital, posterior and dorsal cingulate regions. Many of those regions showed a consistent increase of their entropy with 5HT2A-R agonism in our study (Fig. 6.3A).

Moreover, many of those regions actually belong to the S_1 group (c.f. Fig. 6.4), which suggests that their entropy increase in experimental data might be directly related to the connectivity in those regions.

Together, these findings support the conclusion that the DMF model, once optimised, can reproduce not only FC and FCD (Deco et al., 2018b), but also some of the most salient localised entropy increases found on *in vivo* human studies up to date.

Also, on a more fundamental level, our findings suggest that psychedelics disrupt the functional organisation of the brain with an especially focal and pronounced action on highly anatomically connected brain regions. In fact, experimental and modelling evidence points towards a key role of connector hubs on psychedelic and psychopatological states (Muthukumaraswamy et al., 2013; Müller et al., 2018).

Perturbing the activity in these specific regions (via 5HT2A-R agonism) could have particularly profound implications for the regularity of brain activity and the quality of conscious experience. However, sub cortical regions may also play an important role on the modulation of brain activity, conscious experience, and the psychedelic state (Barrett et al., 2020).

On a separate line of inquiry, there is strong evidence associating the DMN to high-level cognitive functions, including the sense of space, time and recognition of (self) boundaries. Disruptions to the DMN have been linked to fundamental changes in experience, such as ego dissolution (Tagliazucchi et al., 2016b; Carhart-Harris, 2018). Our simulation results show that 5HT2A-R activation increases the entropy of all DMN regions, which is consistent with the reported decrease in the DMN network integrity (Muthukumaraswamy et al., 2013) induced by psychedelic drugs. In contrast, low-level motor functions such as motor regulation remain largely unaffected during the psychedelic state (Carhart-Harris and Friston, 2019), which is consistent with the modest entropy changes observed in the lingual and superior motor areas.

6.3.3 Current limitations and future research

The approach employed in this work presents certain limitations related to several aspects of the simulation and analysis. Acknowledging and understanding these limitations can help us extend and improve our approach, while introducing new questions in the field of psychedelic computational neuroscience.

To our knowledge, the DMF model is the only whole-brain model that implements neuromodulation and is capable of reproducing neuroimaging data in the placebo and psychedelic states. Nonetheless, it makes some important simplifications that are worth discussing. At the network level, the DTI-based connectome used here is known to be incomplete, thus improvements could be made to the model parameters of brain connectivity (Markov et al., 2014). At the dynamical level, the DMF model models neuronal populations as perfectly homogeneous within a given region, and it is known that finer-grained local structure of certain brain regions is likely to be key to explaining certain subjective effects of the psychedelic state (e.g. lattice structure in the visual cortex and geometric visual hallucinations (Ermentrout and Cowan, 1979)). Additionally, the version of the DMF model used here only considers 5HT2A-R agonism, while classic serotonergic psychedelic drugs also have high binding affinities for other receptors (e.g. in the case of LSD, the D1 and D2 dopamine receptors (Carhart-Harris and Nutt, 2017)).

These simplifications do not prevent the model from reproducing statistical features of brain signals under the placebo and 5HT2A-R conditions, but could result in an inability to reproduce finer aspects of the dynamics of the whole-brain activity in these conditions. Extending the model to reproduce other dynamical signatures of psychedelics (like alpha suppression (Timmermann et al., 2019) or reduced directed functional connectivity (Barnett et al., 2020)) constitutes a natural extension of this work with the recent example of Kringelbach *et al.*, (Kringelbach et al., 2020) where an extension of the DMF model can reproduce dynamical features of the brain in placebo and in the psychedelic state.

Finally, it is worth noting that all our analyses here are based on the univariate statistics of individual brain regions, not including any correlation or information flow between them. However, it is known that some high-level subjective effects of psychedelics (such as complex imagery(Carhart-Harris et al., 2016b) and ego dissolution(Tagliazucchi et al., 2016b)) are related to network, as opposed to single region, dynamics. Therefore, building a richer statistical description of the

brain's dynamics using recent information-theoretic tools (such as multivariate extensions of mutual information(Rosas et al., 2019) or Integrated Information Theory, IIT (Balduzzi and Tononi, 2008)) remains an exciting open problem. In fact, recent attempts of unifying IIT and the EBH in a single framework to understand the effects of psychedelic drugs (Swanson, 2018) are helping to bridge the gap between the univariate EBH and the multivariate IIT.

6.3.4 Final remarks

In this chapter we have provided the first mechanistic explanation of the neural entropy increase elicited by psychedelic drugs, using a whole-brain dynamical model with 5HT2A-R neuromodulation. Furthermore, we built a simple model able to predict a region's relative change in entropy from its local 5HT2A-R density and topological properties, showing that, somewhat paradoxically, at a whole-brain level receptor density is a poor predictor of 5HT2A-R activation effect.

Key to developing this predictive model was a three-way partition of brain regions according to the dependency they exhibited with connectivity strength, suggesting a differentiated action mechanism of 5HT2A-R agonists that depends on the local topology of brain regions. In summary, our results suggest that the local changes in entropy, as well as the global entropy increase, induced by 5HT2A-R activation can be explained from a region-specific interplay between structural connectivity and receptor density. Finally, controlled experiments with null network models confirm that receptor density and connectivity strength are necessary, but not sufficient, to explain the entropic effects of 5HT2A-R activation.

The spatially heterogeneous, complex nature of the observed effects of 5HT2A-R activation opens a challenging problem for understanding the clinical and scientific relevance of psychedelic drugs and their entropic effect. Furthermore, it stresses the necessity of moving beyond the current unidimensional approach to consciousness to a multi-dimensional one, that better captures the phenomenological and neurodynamical richness of psychedelic state.

DISCUSSION

7.1 A bottom-up framework for investigating consciousness

Consciousness research is in need of mechanistic accounts to explain why brain signals recorded during different states of consciousness can be consistently characterized by the presence of certain global signatures. Our motivation is not the replacement of the explanations of these signatures provided by theories, such as GNW or IIT. Instead, we aim to put forward a framework for their investigation from a bottom-up perspective. Eventually, we expect to converge on the high-level explanations furnished by some of these theories. Our inspiration is partially drawn from statistical thermodynamics, which provides a clear example of how the bottom-up and top-down perspectives can converge into a consistent picture of physical phenomena. Importantly, in this case, the resulting theory remained useful both as a set of phenomenological principles and computational rules (i.e., classical thermodynamics) but also as a framework to establish connections between those principles and the rules governing the microscopic properties of matter.

The principal idea behind our work is that whole-brain models can be used to test hypotheses concerning the mechanistic and causal underpinnings of different states of consciousness. We do not expect that whole-brain models are sufficiently advanced to identify those precise mechanisms; however, we propose that they can contribute to narrow the space of possible mechanistic explanations, therefore complementing current theories of consciousness from a bottom-up perspective.

Our framework rests upon the complementary nature of three key ingredients:

experimental data obtained through neuroimaging experiments, theoretical approaches to characterize signatures of consciousness, and bottom-up whole-brain computational models. The application of modern neuroimaging techniques to the study of signatures of consciousness has provided very effective tools to predict the brain activity patterns that are associated with different states of consciousness. However, as René Thom famously stated, "to predict is not to explain" (Thom, 1992). Hence, we now turn to the discussion of how models could bridge the gap between prediction and explanation.

The proposed framework and the key elements to model altered states of consciousness is summarized on Figure 7.1:

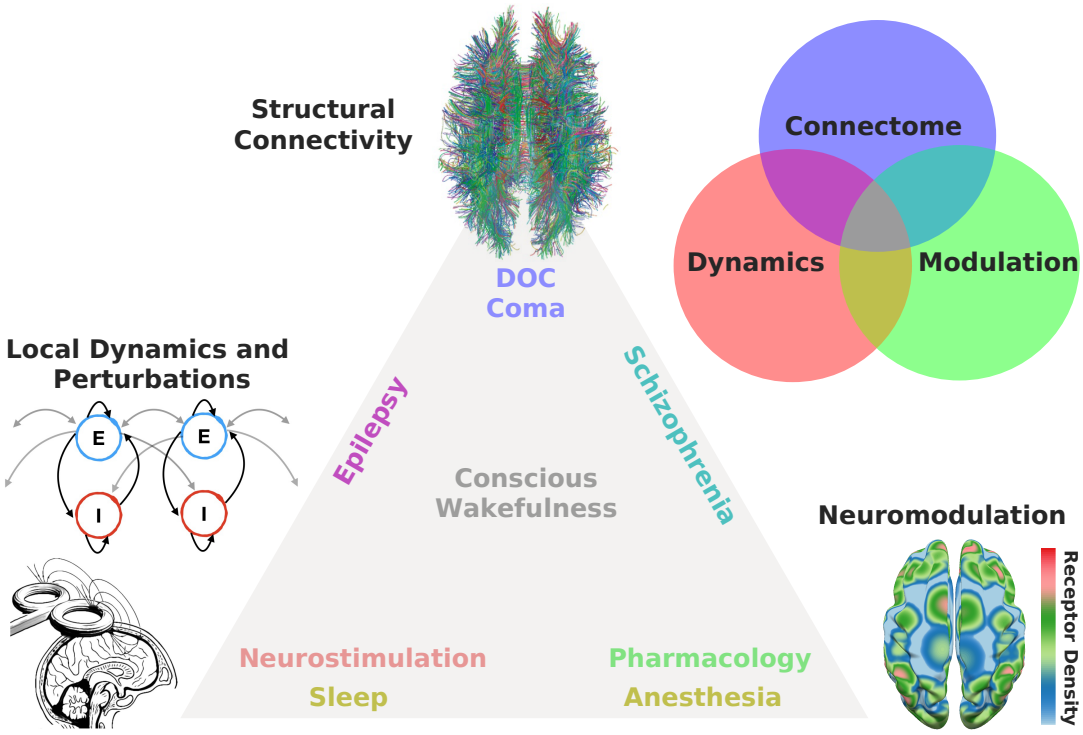

Figure 7.1: **Key variables for whole-brain models**. Representation of the three key variables that can be modified to construct whole-brain models of different altered states of consciousness. These variables correspond to local dynamics, anatomical connectivity, and priors related to neuromodulatory systems necessary to accommodate physiological, pathological, and pharmacologically-induced altered states of consciousness. Certain states may require the modification of multiple variables; for instance, focal seizures and propofol-induced anaesthesia are both associated with low complexity patterns of brain activity; yet, in the first case, these dynamics reflect structural abnormalities, while, in the second case, they reflect the activation of certain inhibitory pathways.

A. Connectome: Is the state of consciousness implicated with local or diffuse

structural abnormalities? This is frequently the case for neurological conditions, such as coma and post-comatose disorders of consciousness (e.g., unresponsive wakefulness syndrome, minimally conscious state) (Fernández-Espejo et al., 2011). In addition, subtler structural modifications can be implicated in certain psychiatric conditions presenting episodes of altered consciousness, such as different forms of schizophrenia (Kubicki et al., 2005). While several papers have investigated localized (e.g., stroke, tumors, focal epilepsy) structural damage from this perspective (Adhikari et al., 2017; Haimovici et al., 2016; Beuter et al., 2020; Váša et al., 2015; Hellyer et al., 2015; Sinha et al., 2017; Richardson, 2012; Aerts et al., 2020), the literature on whole-brain models applied to patients suffering from neurological impairments and from disorders of consciousness is very limited. The project of modeling pathological brain states perforce necessitates to incorporate individualized structural connectomes and lesion maps, thus moving towards simulation at the single patient level (Bansal et al., 2018; Jirsa et al., 2017).

B. Modulation: Is the state of consciousness a consequence of neuromodulatory changes, either endogenous or induced externally by means of pharmacological manipulation? Two typical examples are the altered states of consciousness induced by psychedelics/dissociatives, which are linked to agonism/antagonism at serotonin/glutamate receptors (Nichols, 2016). Certain psychiatric conditions are believed to arise as a consequence of neuromodulatory imbalances, e.g., dopaminergic imbalances are believed to play an important role in the pathophysiology of schizophrenia (Howes and Kapur, 2009). Most anaesthetic drugs reduce the complexity of the brain activity by targeting specific neuromodulatory sites, such as those activated by gamma-aminobutyric acid (GABA) (Peduto et al., 1991). Finally, sleep is a state of reduced consciousness triggered by activity in monoaminergic neurons with diffuse projections throughout the brain (Jouvet, 1972).

C. Dynamics: Is the altered state of consciousness captured by well-understood dynamical mechanisms? Does the model include parametrically controlled external perturbations? While changes in the local excitation/inhibition balance are ultimately caused by neurochemical processes, they are best understood in terms of their dynamical consequences. States such as epilepsy, deep sleep and general anaesthesia are believed to involve unbalanced excitation/inhibition (Gao et al., 2017). In some cases, dynamics may be sufficiently idiosyncratic to be captured by low dimensional phenomenological models,

as in the case of certain forms of epileptic activity (El Houssaini et al., 2020). Finally, local dynamics could be modified to simulate the effects of external neurostimulation (Deco et al., 2019; Sanz Perl et al., 2021).

Depending on the answers to these questions, the whole-brain model should incorporate changes to anatomical connectivity, local dynamics, or include empirical receptor density maps to add a new layer of neurobiological detail.

7.2 What can we learn?

The dynamics of whole-brain models can be perturbed arbitrarily. This is significant since it allows for the exploration of different mechanisms, leading to the observed empirical dynamics. In addition, they enable the exploration of how external stimulation can force transitions between states of consciousness, including the clinically relevant case of displacing whole-brain models from unconscious states towards wakefulness (Deco et al., 2019; Sanz Perl et al., 2021). Therapeutic alternatives to accelerate the recovery of disorder of consciousness (DOC) patients are scarce, and while some studies support the therapeutic role of external electrical stimulation (Hermann et al., 2020), very little is known about the optimal choice of stimulation sites and parameters. Whole-brain models could be useful for the optimization of stimulation protocols, as well as for assisting in clinical decision making. Localized stimulation and/or resection of neural tissue are surgical alternatives to treat certain severe forms of epilepsy, and whole-brain models have been explored with success to predict the outcome of these interventions (An et al., 2019). The same concept could apply to the development and in silico testing of new pharmaceuticals to treat psychiatric conditions, where whole-brain models could be used to reverse-engineer the optimal receptor affinity profiles required to restore statistical signatures of healthy brain dynamics. Finally, the combination of data produced by whole-brain models and machine learning classifiers could be useful for data augmentation in the context of automated diagnosis of rare neurological diseases (Perl et al., 2020b) and to generate input for deep learning architectures (e.g., variational autoencoders) capable of representing altered states of consciousness as trajectories within a low dimensionality latent space (Perl et al., 2020a).

7.2.1 Case study: modeling neural entropy increases induced by psychedelics

In this work we exemplify our framework by using the DMF model to explain a global signature of the psichedelic state, i.e. neural entropy increase. Using the DMF model optimized to fit the FCD of placebo and LSD conditions, a significant entropy increase of brain signals was found in LSD versus placebo as a consequence of simulated 5-HT_{2A} receptor activation. Thus, the model was capable of identifying a low-level (i.e., molecular scale) mechanism leading to increased neural entropy, which is a robust signature of the psychedelic state (Carhart-Harris et al., 2014; Carhart-Harris, 2018).

Since activation of the 5-HT_{2A} receptor is causally implicated with the conscious state induced by serotonergic psychedelics (Nichols, 2016; Kraehenmann et al., 2017; Barrett et al., 2018), the effect of the drug was modeled as a local change in the non-linearity of the inhibitory and excitatory populations. This change was proportional to the local density of 5-HT_{2A} receptors as determined by PET imaging. Brain entropy increases during the psychedelic state were the result of heterogeneous changes in the entropy of the regional firing rates. These changes in firing rate entropy depended both on the local anatomical connectivity and the 5-HT_{2A} receptor density.

Thus, starting from local dynamics describing the behavior of coupled excitatory and inhibitory pools of neurons, and introducing a perturbation which reflects serotonergic activation, the model provided a bottom-up confirmation of 5-HT_{2A} activation as the source of increased neural entropy during the psychedelic state. In the context of Figure 7.1, the model adopted changes in local dynamics (bottom left) informed by empirical maps of 5-HT_{2A} receptor density (bottom right).

7.3 Projections

A fruitful line of future work involves making more detailed comparisons with *in vivo* psychedelic neuroimaging data, and, potentially, subjective experience reports. For example, one natural option would be to use forward models of fMRI(Stephan et al., 2007) or M/EEG(Mosher et al., 1999) to bridge between the firing rates produced by the DMF model and other data modalities, to produce simulated data that

is more directly comparable with available empirical data.

Another exciting possibility is to explore model parameters to examine potential non-linearities and their implications for different relevant aspects of brain function. For example, there are reasons to believe that the dose-response relationship is non-linear for psychedelics and that over a certain threshold dosage (level of 5HT2A-R stimulation) new subjective and global brain function properties can appear (Madsen et al., 2019).

Most interestingly, a potentially very useful extension of this framework would be to include individual subject-level connectome and receptor data to build personalised models of response to psychedelic drugs, or other treatments. Such a framework could potentially make individualised predictions of the action of serotonergic psychedelics on specific individuals, aiding patient stratification and treatment customisation (Scott and Carhart-Harris, 2019). This would enable a much more comprehensive modelling framework, capable of correlating specific brain features with subjective experience reports.

CONCLUSIONS

In this work we developed, implemented and applied a bottom-up framework to study altered states of consciousness. First, we laid the theoretical fundamentals of the framework, proposing that current high-level theories of consciousness could be complemented by a mechanistic knowledge of how key signatures of consciousness, or its modification, are generated by low-level biophysical rules. Second, to ease the massification of such a framework, we improved the computationally efficiency of one of the most popular biophysically grounded whole-brain models, i.e. the dynamic meal field model. Our improved version, here called the FastDMF, involved a semi-analytic solution to the feedback inhibitory control optimization problem, bypassing a huge computational obstacle associated with the DMF. This enabled the reproduction of whole-brain activity with a resolution of thousands of regions. To further extend the possibilities of the DMF, we coupled it to a Bayesian optimization algorithm, paving the way to testing a wide diversity of hypothesis implemented as specific perturbations on the free parameters of the model. Third, we apply the FastDMF to the study of how a molecular perturbation, i.e. activation of the 5HT2A receptor, can change global signatures of the whole-brain activity – the increased neural entropy elicited by psychedelic drugs. By leveraging the theoretical and computational advances of the FastDMF, we fit the DMF to empirical fMRI data, revealing that both inhibitory and excitatory population are needed to jointly reproduce the changes in the FCD and in the neural entropy associated to the ingestion of LSD. In addition, we demonstrate that the entropy increases can be well explained by two local structural features: the local connectivity strength and the local receptor density. Finally, we made a synthesis of the framework in terms of the key elements to be perturbed in whole-brain models, which narrows the space of possible explanations of how different states of consciousness are intimately linked to low-level biophysical mechanisms. This way, we aimed to integrate in a single

framework subjective reports, high-level signatures of brain activity, and it corresponding underlying biophysical mechanisms.

Bibliography

Abbott, L. F. and Chance, F. S. (2005). Drivers and modulators from push-pull and balanced synaptic input. Progress in brain research, 149:147–155.

Abeyasinghe, P. M., Aiello, M., Nichols, E. S., Cavaliere, C., Fiorenza, S., Masotta, O., Borrelli, P., Owen, A. M., Estraneo, A., and Soddu, A. (2020). Consciousness and the dimensionality of doc patients via the generalized ising model. Journal of clinical medicine, 9(5):1342.

Abrahams, D. and Grosse-Kunstleve, R. W. (2003). Building hybrid systems with boost. python. C/C++ Users Journal, 21(LBNL-53142).

Adhikari, M. H., Hacker, C. D., Siegel, J. S., Griffa, A., Hagmann, P., Deco, G., and Corbetta, M. (2017). Decreased integration and information capacity in stroke measured by whole brain models of resting state activity. Brain, 140(4):1068–1085.

Aerts, H., Schirner, M., Dhollander, T., Jeurissen, B., Achten, E., Van Roost, D., Ritter, P., and Marinazzo, D. (2020). Modeling brain dynamics after tumor resection using the virtual brain. Neuroimage, 213:116738.

Aghajanian, G. and Marek, G. (1997). Serotonin induces excitatory postsynaptic potentials in apical dendrites of neocortical pyramidal cells. Neuropharmacology, 36(4-5):589–599.

Alonso, L. M., Proekt, A., Schwartz, T. H., Pryor, K. O., Cecchi, G. A., and Magnasco, M. O. (2014). Dynamical criticality during induction of anesthesia in human ecog recordings. Frontiers in neural circuits, 8:20.

An, S., Bartolomei, F., Guye, M., and Jirsa, V. (2019). Optimization of surgical intervention outside the epileptogenic zone in the virtual epileptic patient (vep). PLoS computational biology, 15(6):e1007051.

Aru, J., Bachmann, T., Singer, W., and Melloni, L. (2012). Distilling the neural correlates of consciousness. Neuroscience & Biobehavioral Reviews, 36(2):737–746.

Atasoy, S., Donnelly, I., and Pearson, J. (2016). Human brain networks function in connectome-specific harmonic waves. Nature communications, 7(1):1–10.

Baars, B. J. (2005). Global workspace theory of consciousness: toward a cognitive neuroscience of human experience. Progress in brain research, 150:45–53.

Babaie Janvier, T. and Robinson, P. A. (2018). Neural field theory of corticothalamic prediction with control systems analysis. Frontiers in Human Neuroscience, page 334.

Balduzzi, D. and Tononi, G. (2008). Integrated information in discrete dynamical systems: Motivation and theoretical framework. PLoS Computational Biology, 4(6):e1000091.

Balduzzi, D. and Tononi, G. (2009). Qualia: the geometry of integrated information. PLoS computational biology, 5(8):e1000462.

Bansal, K., Nakuci, J., and Muldoon, S. F. (2018). Personalized brain network models for assessing structure–function relationships. Current Opinion in Neurobiology, 52:42–47.

Barnett, L., Muthukumaraswamy, S. D., Carhart-Harris, R. L., and Seth, A. K. (2020). Decreased directed functional connectivity in the psychedelic state. NeuroImage, 209:116462.

Barrett, A. B. and Mediano, P. A. (2019). The phi measure of integrated information is not well-defined for general physical systems. Journal of Consciousness Studies, 26(1-2):11–20.

Barrett, F. S., Krimmel, S. R., Griffiths, R. R., Seminowicz, D. A., and Mathur, B. N. (2020). Psilocybin acutely alters the functional connectivity of the claustrum with brain networks that support perception, memory, and attention. Neuroimage, 218:116980.

Barrett, F. S., Preller, K. H., Herdener, M., Janata, P., and Vollenweider, F. X. (2018). Serotonin 2a receptor signaling underlies lsd-induced alteration of the neural response to dynamic changes in music. Cerebral Cortex, 28(11):3939–3950.

Bayne, T. (2007). Conscious states and conscious creatures: Explanation in the scientific study of consciousness. Philosophical Perspectives, 21:1–22.

Bayne, T. (2018). On the axiomatic foundations of the integrated information theory of consciousness. Neuroscience of consciousness, 2018(1):niy007.

Bayne, T. and Carter, O. (2018). Dimensions of consciousness and the psychedelic state. Neuroscience of consciousness, 2018(1):niy008.

Bayne, T., Hohwy, J., and Owen, A. M. (2016). Are there levels of consciousness? Trends in cognitive sciences, 20(6):405–413.

Behrens, T. E., Berg, H. J., Jbabdi, S., Rushworth, M. F., and Woolrich, M. W. (2007). Probabilistic diffusion tractography with multiple fibre orientations: What can we gain? neuroimage, 34(1):144–155.

Behrens, T. E., Woolrich, M. W., Jenkinson, M., Johansen-Berg, H., Nunes, R. G., Clare, S., Matthews, P. M., Brady, J. M., and Smith, S. M. (2003). Characterization and propagation of uncertainty in diffusion-weighted mr imaging. Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine, 50(5):1077–1088.

Beliveau, V., Ganz, M., Feng, L., Ozenne, B., Højgaard, L., Fisher, P. M., Svarer, C., Greve, D. N., and Knudsen, G. M. (2017). A high-resolution in vivo atlas of the human brain's serotonin system. Journal of Neuroscience, 37(1):120–128.

Beuter, A., Balossier, A., Vassal, F., Hemm, S., and Volpert, V. (2020). Cortical stimulation in aphasia following ischemic stroke: Toward model-guided electrical neuromodulation. Biological Cybernetics, 114(1):5–21.

Block, N. (1980). Troubles with functionalism. Readings in philosophy of psychology, 1:268–305.

Block, N. (1995). On a confusion about a function of consciousness. Behavioral and brain sciences, 18(2):227–247.

Block, N. (2011). Perceptual consciousness overflows cognitive access. Trends in cognitive sciences, 15(12):567–575.

Bocaccio, H., Pallavicini, C., Castro, M. N., Sánchez, S. M., De Pino, G., Laufs, H., Villarreal, M. F., and Tagliazucchi, E. (2019). The avalanche-like behaviour of large-scale haemodynamic activity from wakefulness to deep sleep. Journal of the Royal Society Interface, 16(158):20190262.

Breakspear, M. (2017). Dynamic models of large-scale brain activity. Nature neuroscience, 20(3):340–352.

Breakspear, M., Terry, J. R., and Friston, K. J. (2003). Modulation of excitatory synaptic coupling facilitates synchronization and complex dynamics in a biophysical model of neuronal dynamics. Network: Computation in Neural Systems, 14(4):703.

Cabral, J., Hugues, E., Kringelbach, M. L., and Deco, G. (2012). Modeling the outcome of structural disconnection on resting-state functional connectivity. Neuroimage, 62(3):1342–1353.

Cabral, J., Kringelbach, M. L., and Deco, G. (2014). Exploring the network dynamics underlying brain activity during rest. Progress in neurobiology, 114:102–131.

Cardeña, E. and Winkelman, M. J. (2011). Altering Consciousness: Multidisciplinary Perspectives [2 volumes]: Multidisciplinary Perspectives. ABC-CLIO.

Carhart-Harris, R. and Nutt, D. (2017). Serotonin and brain function: a tale of two receptors. Journal of Psychopharmacology, 31(9):1091–1120.

Carhart-Harris, R. L. (2018). The entropic brain-revisited. Neuropharmacology, 142:167–178.

Carhart-Harris, R. L., Bolstridge, M., Rucker, J., Day, C. M., Erritzoe, D., Kaelen, M., Bloomfield, M., Rickard, J. A., Forbes, B., Feilding, A., et al. (2016a). Psilocybin with psychological support for treatment-resistant depression: an open-label feasibility study. The Lancet Psychiatry, 3(7):619–627.

Carhart-Harris, R. L., Erritzoe, D., Williams, T., Stone, J. M., Reed, L. J., Colasanti, A., Tyacke, R. J., Leech, R., Malizia, A. L., Murphy, K., et al. (2012). Neural correlates of the psychedelic state as determined by fmri studies with psilocybin. Proceedings of the National Academy of Sciences, 109(6):2138–2143.

Carhart-Harris, R. L. and Friston, K. (2019). Rebus and the anarchic brain: toward a unified model of the brain action of psychedelics. Pharmacological reviews, 71(3):316–344.

Carhart-Harris, R. L., Leech, R., Hellyer, P. J., Shanahan, M., Feilding, A., Tagliazucchi, E., Chialvo, D. R., and Nutt, D. (2014). The entropic brain: a theory of conscious states informed by neuroimaging research with psychedelic drugs. Frontiers in human neuroscience, page 20.

Carhart-Harris, R. L., Muthukumaraswamy, S., Roseman, L., Kaelen, M., Droog, W., Murphy, K., Tagliazucchi, E., Schenberg, E. E., Nest, T., Orban, C., et al. (2016b).

Neural correlates of the lsd experience revealed by multimodal neuroimaging. Proceedings of the National Academy of Sciences, 113(17):4853–4858.

Casali, A. G., Gosseries, O., Rosanova, M., Boly, M., Sarasso, S., Casali, K. R., Casarotto, S., Bruno, M.-A., Laureys, S., Tononi, G., et al. (2013). A theoretically based index of consciousness independent of sensory processing and behavior. Science translational medicine, 5(198):198ra105–198ra105.

Chalmers, D. J. (1999). First-person methods in the science of consciousness. Consciousness Bulletin.

Chalmers, D. J. (2000). What is a neural correlate of consciousness. Neural correlates of consciousness: Empirical and conceptual questions, pages 17–39.

Chang, J.-Y., Pigorini, A., Massimini, M., Tononi, G., Nobili, L., and Van Veen, B. D. (2012). Multivariate autoregressive models with exogenous inputs for intracerebral responses to direct electrical stimulation of the human brain. Frontiers in human neuroscience, 6:317.

Chennu, S., O'Connor, S., Adapa, R., Menon, D. K., and Bekinschtein, T. A. (2016). Brain connectivity dissociates responsiveness from drug exposure during propofol-induced transitions of consciousness. PLoS computational biology, 12(1):e1004669.

Chialvo, D. R. (2010). Emergent complex neural dynamics. Nature physics, 6(10):744–750.

Cofré, R., Herzog, R., Mediano, P. A., Piccinini, J., Rosas, F. E., Sanz Perl, Y., and Tagliazucchi, E. (2020). Whole-brain models to explore altered states of consciousness from the bottom up. Brain Sciences, 10(9):626.

Cohen, M. A. and Dennett, D. C. (2011). Consciousness cannot be separated from function. Trends in cognitive sciences, 15(8):358–364.

Coronel-Oliveros, C., Cofré, R., and Orio, P. (2021). Cholinergic neuromodulation of inhibitory interneurons facilitates functional integration in whole-brain models. PLoS computational biology, 17(2):e1008737.

Cover, T. M. (1999). Elements of information theory. John Wiley & Sons.

Crick, F. and Koch, C. (1990). Towards a neurobiological theory of consciousness. In Seminars in the Neurosciences, volume 2, pages 263–275. Saunders Scientific Publications.

Crick, F. and Koch, C. (2003). A framework for consciousness. Nature neuroscience, 6(2):119–126.

Damoiseaux, J. S., Rombouts, S., Barkhof, F., Scheltens, P., Stam, C. J., Smith, S. M., and Beckmann, C. F. (2006). Consistent resting-state networks across healthy subjects. Proceedings of the national academy of sciences, 103(37):13848–13853.

Deco, G., Cabral, J., Saenger, V. M., Boly, M., Tagliazucchi, E., Laufs, H., Van Someren, E., Jobst, B., Stevner, A., and Kringelbach, M. L. (2018a). Perturbation of whole-brain dynamics in silico reveals mechanistic differences between brain states. Neuroimage, 169:46–56.

Deco, G., Cruzat, J., Cabral, J., Knudsen, G. M., Carhart-Harris, R. L., Whybrow, P. C., Logothetis, N. K., and Kringelbach, M. L. (2018b). Whole-brain multimodal neuroimaging model using serotonin receptor maps explains non-linear functional effects of lsd. Current biology, 28(19):3065–3074.

Deco, G., Cruzat, J., Cabral, J., Tagliazucchi, E., Laufs, H., Logothetis, N. K., and Kringelbach, M. L. (2019). Awakening: Predicting external stimulation to force transitions between different brain states. Proceedings of the National Academy of Sciences, 116(36):18088–18097.

Deco, G., Jirsa, V. K., Robinson, P. A., Breakspear, M., and Friston, K. (2008). The dynamic brain: from spiking neurons to neural masses and cortical fields. PLoS computational biology, 4(8):e1000092.

Deco, G. and Kringelbach, M. L. (2014). Great expectations: using whole-brain computational connectomics for understanding neuropsychiatric disorders. Neuron, 84(5):892–905.

Deco, G. and Kringelbach, M. L. (2020). Turbulent-like dynamics in the human brain. Cell reports, 33(10):108471.

Deco, G., Perl, Y. S., Vuust, P., Tagliazucchi, E., Kennedy, H., and Kringelbach, M. L. (2021). Rare long-range cortical connections enhance human information processing. Current Biology, 31(20):4436–4448.

Deco, G., Ponce-Alvarez, A., Hagmann, P., Romani, G. L., Mantini, D., and Corbetta, M. (2014). How local excitation–inhibition ratio impacts the whole brain dynamics. Journal of Neuroscience, 34(23):7886–7898.

Dehaene, S. and Changeux, J.-P. (2011). Experimental and theoretical approaches to conscious processing. Neuron, 70(2):200–227.

Dehaene, S. and Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework. Cognition, 79(1-2):1–37.

Demirtaş, M., Burt, J. B., Helmer, M., Ji, J. L., Adkinson, B. D., Glasser, M. F., Van Essen, D. C., Sotiropoulos, S. N., Anticevic, A., and Murray, J. D. (2019). Hierarchical heterogeneity across human cortex shapes large-scale neural dynamics. Neuron, 101(6):1181–1194.

Dennett, D. (2003). Who's on first? heterophenomenology explained. Journal of Consciousness Studies, 10(9-10):19–30.

Dennett, D. C. (1993). Consciousness explained. Penguin uk.

Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., Buckner, R. L., Dale, A. M., Maguire, R. P., Hyman, B. T., et al. (2006). An automated labeling system for subdividing the human cerebral cortex on mri scans into gyral based regions of interest. Neuroimage, 31(3):968–980.

Destexhe, A. and Sejnowski, T. J. (2009). The wilson–cowan model, 36 years later. Biological cybernetics, 101(1):1–2.

Deutsch, D. (1986). A musical paradox. Music Perception, 3(3):275–280.

Doerig, A., Schurger, A., Hess, K., and Herzog, M. H. (2019). The unfolding argument: Why iit and other causal structure theories cannot explain consciousness. Consciousness and cognition, 72:49–59.

El Houssaini, K., Bernard, C., and Jirsa, V. K. (2020). The epileptor model: a systematic mathematical analysis linked to the dynamics of seizures, refractory status epilepticus, and depolarization block. Eneuro, 7(2).

Ermentrout, G. B. and Cowan, J. D. (1979). A mathematical theory of visual hallucination patterns. Biological cybernetics, 34(3):137–150.

Escrichs, A., Sanz, Y., Uribe, C., Camara, E., Türker, B., Pyatigorskaya, N., López-González, A., Pallavicini, C., Panda, R., Annen, J., et al. (2021). Unifying turbulent dynamics framework distinguishes different brain states. bioRxiv.

Fernández-Espejo, D., Bekinschtein, T., Monti, M. M., Pickard, J. D., Junque, C., Coleman, M. R., and Owen, A. M. (2011). Diffusion weighted imaging distinguishes the vegetative state from the minimally conscious state. Neuroimage, 54(1):103–112.

Frankish, K. (2016). Illusionism as a theory of consciousness. Journal of Consciousness Studies, 23(11-12):11–39.

Friston, K. J., Harrison, L., and Penny, W. (2003). Dynamic causal modelling. Neuroimage, 19(4):1273–1302.

Friston, K. J., Mechelli, A., Turner, R., and Price, C. J. (2000). Nonlinear responses in fmri: the balloon model, volterra kernels, and other hemodynamics. NeuroImage, 12(4):466–477.

Gao, R., Peterson, E. J., and Voytek, B. (2017). Inferring synaptic excitation/inhibition balance from field potentials. Neuroimage, 158:70–78.

Gatica, M., Rosas, F. E., Mediano, P. A., Diez, I., Swinnen, S. P., Orio, P., Cofre, R., and Cortes, J. M. (2021). High-order functional interactions in ageing explained via alterations in the connectome in a whole-brain model. bioRxiv.

Gerstner, W., Sprekeler, H., and Deco, G. (2012). Theory and simulation in neuroscience. science, 338(6103):60–65.

Grasso, C., Barresi, M., et al. (2016). Serotonin modifies the spontaneous spiking activity of gracile nucleus neurons in rats: role of 5-ht1a and 5-ht2 receptors. Archives italiennes de biologie, 154(2-3):39–49.

Guennebaud, G., Jacob, B., et al. (2010). Eigen v3. http://eigen.tuxfamily.org.

Hagmann, P., Cammoun, L., Gigandet, X., Meuli, R., Honey, C. J., Wedeen, V. J., and Sporns, O. (2008). Mapping the structural core of human cerebral cortex. PLoS biology, 6(7):e159.

Haimovici, A., Balenzuela, P., and Tagliazucchi, E. (2016). Dynamical signatures of structural connectivity damage to a model of the brain posed at criticality. Brain connectivity, 6(10):759–771.

Haimovici, A., Tagliazucchi, E., Balenzuela, P., and Chialvo, D. R. (2013). Brain organization into resting state networks emerges at criticality on a model of the human connectome. Physical review letters, 110(17):178101.

Hansen, E. C., Battaglia, D., Spiegler, A., Deco, G., and Jirsa, V. K. (2015). Functional connectivity dynamics: modeling the switching behavior of the resting state. Neuroimage, 105:525–535.

Hellyer, P. J., Jachs, B., Clopath, C., and Leech, R. (2016). Local inhibitory plasticity tunes macroscopic brain dynamics and allows the emergence of functional brain networks. NeuroImage, 124:85–95.

Hellyer, P. J., Scott, G., Shanahan, M., Sharp, D. J., and Leech, R. (2015). Cognitive flexibility through metastable neural dynamics is disrupted by damage to the structural connectome. Journal of Neuroscience, 35(24):9050–9063.

Hermann, B., Raimondo, F., Hirsch, L., Huang, Y., Denis-Valente, M., Pérez, P., Engemann, D., Faugeras, F., Weiss, N., Demeret, S., et al. (2020). Combined behavioral and electrophysiological evidence for a direct cortical effect of prefrontal tdcs on disorders of consciousness. Scientific reports, 10(1):1–16.

Herzog, R., Mediano, P. A., Rosas, F. E., Carhart-Harris, R., Perl, Y. S., Tagliazucchi, E., and Cofre, R. (2020). A mechanistic model of the neural entropy increase elicited by psychedelic drugs. Scientific reports, 10(1):1–12.

Honey, C. J., Sporns, O., Cammoun, L., Gigandet, X., Thiran, J.-P., Meuli, R., and Hagmann, P. (2009). Predicting human resting-state functional connectivity from structural connectivity. Proceedings of the National Academy of Sciences, 106(6):2035–2040.

Horn, A. and Blankenburg, F. (2016). Toward a standardized structural–functional group connectome in mni space. Neuroimage, 124:310–322.

Horn, A., Neumann, W.-J., Degen, K., Schneider, G.-H., and Kühn, A. A. (2017). Toward an electrophysiological "sweet spot" for deep brain stimulation in the subthalamic nucleus. Human brain mapping, 38(7):3377–3390.

Howes, O. D. and Kapur, S. (2009). The dopamine hypothesis of schizophrenia: version iii—the final common pathway. Schizophrenia bulletin, 35(3):549–562.

Ipiña, I. P., Kehoe, P. D., Kringelbach, M., Laufs, H., Ibañez, A., Deco, G., Perl, Y. S., and Tagliazucchi, E. (2020). Modeling regional changes in dynamic stability during sleep and wakefulness. NeuroImage, 215:116833.

Jirsa, V. K., Proix, T., Perdikis, D., Woodman, M. M., Wang, H., Gonzalez-Martinez, J., Bernard, C., Bénar, C., Guye, M., Chauvel, P., et al. (2017). The virtual epileptic patient: individualized whole-brain models of epilepsy spread. Neuroimage, 145:377–388.

Jobst, B. M., Hindriks, R., Laufs, H., Tagliazucchi, E., Hahn, G., Ponce-Alvarez, A., Stevner, A., Kringelbach, M. L., and Deco, G. (2017). Increased stability and breakdown of brain effective connectivity during slow-wave sleep: mechanistic insights from whole-brain computational modelling. Scientific reports, 7(1):1–16.

Jouvet, M. (1972). The role of monoamines and acetylcholine-containing neurons in the regulation of the sleep-waking cycle. In Neurophysiology and neurochemistry of sleep and wakefulness, pages 166–307. Springer.

Kim, H., Hudetz, A. G., Lee, J., Mashour, G. A., Lee, U., Avidan, M. S., Bel-Bahar, T., Blain-Moraes, S., Golmirzaie, G., Janke, E., et al. (2018). Estimating the integrated information measure phi from high-density electroencephalography during states of consciousness in humans. Frontiers in human neuroscience, 12:42.

Kitazono, J., Kanai, R., and Oizumi, M. (2018). Efficient algorithms for searching the minimum information partition in integrated information theory. Entropy, 20(3):173.

Kobeleva, X., López-González, A., Kringelbach, M. L., and Deco, G. (2021). Revealing the relevant spatiotemporal scale underlying whole-brain dynamics. Frontiers in neuroscience, page 1367.

Koch, C., Massimini, M., Boly, M., and Tononi, G. (2016). Neural correlates of consciousness: progress and problems. Nature Reviews Neuroscience, 17(5):307–321.

Kong, X., Kong, R., Orban, C., Wang, P., Zhang, S., Anderson, K., Holmes, A., Murray, J. D., Deco, G., van den Heuvel, M., et al. (2021). Sensory-motor cortices shape functional connectivity dynamics in the human brain. Nature communications, 12(1):1–15.

Kraehenmann, R., Pokorny, D., Aicher, H., Preller, K. H., Pokorny, T., Bosch, O. G., Seifritz, E., and Vollenweider, F. X. (2017). Lsd increases primary process thinking via serotonin 2a receptor activation. Frontiers in pharmacology, 8:814.

Kriegeskorte, N. and Douglas, P. K. (2018). Cognitive computational neuroscience. Nature neuroscience, 21(9):1148–1160.

Kriegeskorte, N. and Kievit, R. A. (2013). Representational geometry: integrating cognition, computation, and the brain. Trends in cognitive sciences, 17(8):401–412.

Kriegeskorte, N., Mur, M., and Bandettini, P. A. (2008). Representational similarity analysis-connecting the branches of systems neuroscience. Frontiers in systems neuroscience, 2:4.

Kringelbach, M. L., Cruzat, J., Cabral, J., Knudsen, G. M., Carhart-Harris, R., Whybrow, P. C., Logothetis, N. K., and Deco, G. (2020). Dynamic coupling of whole-brain neuronal and neurotransmitter systems. Proceedings of the National Academy of Sciences, 117(17):9566–9576.

Krohn, S. and Ostwald, D. (2017). Computing integrated information. Neuroscience of consciousness, 2017(1):nix017.

Kubicki, M., Park, H., Westin, C.-F., Nestor, P. G., Mulkern, R. V., Maier, S. E., Niznikiewicz, M., Connor, E. E., Levitt, J. J., Frumin, M., et al. (2005). Dti and mtr abnormalities in schizophrenia: analysis of white matter integrity. Neuroimage, 26(4):1109–1118.

Kunzendorf, R. G. and Wallace, B. (2000). Individual differences in conscious experience, volume 20. John Benjamins Publishing.

Lamme, V. A. (2006). Towards a true neural stance on consciousness. Trends in cognitive sciences, 10(11):494–501.

Laureys, S., Goldman, S., Phillips, C., Van Bogaert, P., Aerts, J., Luxen, A., Franck, G., and Maquet, P. (1999). Impaired effective cortical connectivity in vegetative state: preliminary investigation using pet. Neuroimage, 9(4):377–382.

Laureys, S., Owen, A. M., and Schiff, N. D. (2004). Brain function in coma, vegetative state, and related disorders. The Lancet Neurology, 3(9):537–546.

Lebedev, A. V., Kaelen, M., Lövdén, M., Nilsson, J., Feilding, A., Nutt, D. J., and Carhart-Harris, R. L. (2016). Lsd-induced entropic brain activity predicts subsequent personality change. Human brain mapping, 37(9):3203–3213.

LeDoux, J. E., Michel, M., and Lau, H. (2020). 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.

Licata, F., Volsi, G. L., Maugeri, G., Ciranna, L., and Santangelo, F. (1993). Serotonin-evoked modifications of the neuronal firing rate in the superior vestibular nucleus: a microiontophoretic study in the rat. Neuroscience, 52(4):941–949.

Lizier, J. T. (2014). Jidt: An information-theoretic toolkit for studying the dynamics of complex systems. Frontiers in Robotics and AI, 1:11.

Lord, L.-D., Expert, P., Atasoy, S., Roseman, L., Rapuano, K., Lambiotte, R., Nutt, D. J., Deco, G., Carhart-Harris, R. L., Kringelbach, M. L., et al. (2019). Dynamical exploration of the repertoire of brain networks at rest is modulated by psilocybin. Neuroimage, 199:127–142.

Ludwig, A. M. (1966). Altered states of consciousness. Archives of general Psychiatry, 15(3):225–234.

Luppi, A. I., Mediano, P. A., Rosas, F. E., Allanson, J., Pickard, J. D., Williams, G. B., Craig, M. M., Finoia, P., Peattie, A. R., Coppola, P., et al. (2021). Paths to oblivion: common neural mechanisms of anaesthesia and disorders of consciousness. bioRxiv.

Lutz, A., Lachaux, J.-P., Martinerie, J., and Varela, F. J. (2002). Guiding the study of brain dynamics by using first-person data: synchrony patterns correlate with ongoing conscious states during a simple visual task. Proceedings of the national academy of sciences, 99(3):1586–1591.

Madsen, M. K., Fisher, P. M., Burmester, D., Dyssegaard, A., Stenbæk, D. S., Kristiansen, S., Johansen, S. S., Lehel, S., Linnet, K., Svarer, C., et al. (2019). Psychedelic effects of psilocybin correlate with serotonin 2a receptor occupancy and plasma psilocin levels. Neuropsychopharmacology, 44(7):1328–1334.

Marinazzo, D., Pellicoro, M., Wu, G., Angelini, L., Cortés, J. M., and Stramaglia, S. (2014). Information transfer and criticality in the ising model on the human connectome. PloS one, 9(4):e93616.

Markov, N. T., Ercsey-Ravasz, M. M., Ribeiro Gomes, A., Lamy, C., Magrou, L., Vezoli, J., Misery, P., Falchier, A., Quilodran, R., Gariel, M.-A., et al. (2014). A weighted and directed interareal connectivity matrix for macaque cerebral cortex. Cerebral cortex, 24(1):17–36.

Markram, H. (2006). The blue brain project. Nature Reviews Neuroscience, 7(2):153–160.

Marsden, J. E. and McCracken, M. (2012). The Hopf bifurcation and its applications, volume 19. Springer Science & Business Media.

Mashour, G. A., Roelfsema, P., Changeux, J.-P., and Dehaene, S. (2020). Conscious processing and the global neuronal workspace hypothesis. Neuron, 105(5):776–798.

McLean, R. A., Sanders, W. L., and Stroup, W. W. (1991). A unified approach to mixed linear models. The American Statistician, 45(1):54–64.

Mediano, P. (2019). Integrated Information Theory in Complex Neural Systems. PhD thesis, Imperial College London.

Mediano, P. A., Farah, J. C., and Shanahan, M. (2016). Integrated information and metastability in systems of coupled oscillators. arXiv preprint arXiv:1606.08313.

Mediano, P. A., Rosas, F., Carhart-Harris, R. L., Seth, A. K., and Barrett, A. B. (2019). Beyond integrated information: A taxonomy of information dynamics phenomena. arXiv preprint arXiv:1909.02297.

Mehrmann, V., Sima, V., Varga, A., and Xu, H. (1999). A matlab mex-file environment of slicot. SLICOT Working Note, 11.

Messé, A., Rudrauf, D., Benali, H., and Marrelec, G. (2014). Relating structure and function in the human brain: relative contributions of anatomy, stationary dynamics, and non-stationarities. PLoS computational biology, 10(3):e1003530.

Michel, M. (2019). Consciousness science underdetermined: A short history of endless debates. Ergo, an Open Access Journal of Philosophy, 6(28).

Mindt, G. (2017). The problem with the'information'in integrated information theory. Journal of Consciousness Studies, 24(7-8):130–154.

Morch, H. H. (2019). Is consciousness intrinsic?: A problem for the integrated information theory. Journal of Consciousness Studies, 26(1-2):133–162.

Mosher, J. C., Leahy, R. M., and Lewis, P. S. (1999). Eeg and meg: forward solutions for inverse methods. IEEE Transactions on biomedical engineering, 46(3):245–259.

Müller, F., Dolder, P. C., Schmidt, A., Liechti, M. E., and Borgwardt, S. (2018). Altered network hub connectivity after acute lsd administration. NeuroImage: Clinical, 18:694–701.

Munn, B. R., Müller, E. J., Wainstein, G., and Shine, J. M. (2021). The ascending arousal system shapes neural dynamics to mediate awareness of cognitive states. Nature communications, 12(1):1–9.

Muthukumaraswamy, S. D., Carhart-Harris, R. L., Moran, R. J., Brookes, M. J., Williams, T. M., Errtizoe, D., Sessa, B., Papadopoulos, A., Bolstridge, M., Singh, K. D., et al. (2013). Broadband cortical desynchronization underlies the human psychedelic state. Journal of Neuroscience, 33(38):15171–15183.

Naskar, A., Vattikonda, A., Deco, G., Roy, D., and Banerjee, A. (2021). Multi-scale dynamic mean field (mdmf) model relates resting-state brain dynamics with local cortical excitatory–inhibitory neurotransmitter homeostasis. Network Neuroscience, 5(3):757–782.

Natsoulas, T. (1981). Basic problems of consciousness. Journal of Personality and Social Psychology, 41(1):132.

Nichols, D. E. (2016). Psychedelics. Pharmacological reviews, 68(2):264–355.

Nida-Rümelin, M. (2016). The illusion of illusionism. Journal of Consciousness Studies, 23(11-12):160–171.

Noë, A. and Thompson, E. (2004). Are there neural correlates of consciousness? Journal of Consciousness studies, 11(1):3–28.

Oizumi, M., Albantakis, L., and Tononi, G. (2014). From the phenomenology to the mechanisms of consciousness: integrated information theory 3.0. PLoS computational biology, 10(5):e1003588.

Overgaard, M. (2017). The status and future of consciousness research. Frontiers in psychology, 8:1719.

Overgaard, M. and Overgaard, R. (2010). Neural correlates of contents and levels of consciousness. Frontiers in psychology, 1:164.

Pasricha, S. and Stevenson, I. (1986). Near-death experiences in india: A preliminary report. The Journal of nervous and mental disease, 174(3):165–170.

Peduto, V., Concas, A., Santoro, G., Biggio, G., and Gessa, G. (1991). Biochemical and electrophysiologic evidence that propofol enhances gabaergic transmission in the rat brain. Anesthesiology, 75(6):1000–1009.

Perl, Y. S., Bocaccio, H., Pérez-Ipiña, I., Zamberlán, F., Piccinini, J., Laufs, H., Kringelbach, M., Deco, G., and Tagliazucchi, E. (2020a). Generative embeddings of brain collective dynamics using variational autoencoders. Physical Review Letters, 125(23):238101.

Perl, Y. S., Pallavicini, C., Ipiña, I. P., Kringelbach, M., Deco, G., Laufs, H., and Tagliazucchi, E. (2020b). Data augmentation based on dynamical systems for the classification of brain states. Chaos, Solitons & Fractals, 139:110069.

Quednow, B. B., Kometer, M., Geyer, M. A., and Vollenweider, F. X. (2012). Psilocybin-induced deficits in automatic and controlled inhibition are attenuated by ketanserin in healthy human volunteers. Neuropsychopharmacology, 37(3):630–640.

Ramsey, J. D., Hanson, S. J., Hanson, C., Halchenko, Y. O., Poldrack, R. A., and Glymour, C. (2010). Six problems for causal inference from fmri. neuroimage, 49(2):1545–1558.

Reardon, S. (2019). Rival theories face off over brain's source of consciousness.

Renart, A., Brunel, N., and Wang, X.-J. (2004). Mean-field theory of irregularly spiking neuronal populations and working memory in recurrent cortical networks. Computational neuroscience: A comprehensive approach, pages 431–490.

Revonsuo, A., Kallio, S., and Sikka, P. (2009). What is an altered state of consciousness? Philosophical Psychology, 22(2):187–204.

Richardson, M. P. (2012). Large scale brain models of epilepsy: dynamics meets connectomics. Journal of Neurology, Neurosurgery & Psychiatry, 83(12):1238–1248.

Ritter, P., Schirner, M., McIntosh, A. R., and Jirsa, V. K. (2013). The virtual brain integrates computational modeling and multimodal neuroimaging. Brain connectivity, 3(2):121–145.

Robinson, P. and Roy, N. (2015). Neural field theory of nonlinear wave-wave and wave-neuron processes. Physical Review E, 91(6):062719.

Rosas, F. E., Mediano, P. A., Gastpar, M., and Jensen, H. J. (2019). Quantifying high-order interdependencies via multivariate extensions of the mutual information. Physical Review E, 100(3):032305.

Rubinov, M. and Sporns, O. (2010). Complex network measures of brain connectivity: uses and interpretations. Neuroimage, 52(3):1059–1069.

Saggio, M. L., Ritter, P., and Jirsa, V. K. (2016). Analytical operations relate structural and functional connectivity in the brain. PloS one, 11(8):e0157292.

Sanz Perl, Y., Pallavicini, C., Pérez Ipiña, I., Demertzi, A., Bonhomme, V., Martial, C., Panda, R., Annen, J., Ibañez, A., Kringelbach, M., et al. (2021). Perturbations in dynamical models of whole-brain activity dissociate between the level and stability of consciousness. PLoS Computational Biology, 17(7):e1009139.

Schaefer, A., Kong, R., Gordon, E. M., Laumann, T. O., Zuo, X.-N., Holmes, A. J., Eickhoff, S. B., and Yeo, B. T. (2018). Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity mri. Cerebral cortex, 28(9):3095–3114.

Schartner, M. M., Pigorini, A., Gibbs, S. A., Arnulfo, G., Sarasso, S., Barnett, L., Nobili, L., Massimini, M., Seth, A. K., and Barrett, A. B. (2017). Global and local complexity of intracranial eeg decreases during nrem sleep. Neuroscience of consciousness, 2017(1):niw022.

Schirner, M., McIntosh, A. R., Jirsa, V., Deco, G., and Ritter, P. (2018). Inferring multi-scale neural mechanisms with brain network modelling. Elife, 7:e28927.

Scott, G. and Carhart-Harris, R. L. (2019). Psychedelics as a treatment for disorders of consciousness. Neuroscience of consciousness, 2019(1):niz003.

Seager, W. (2017). Could consciousness be an illusion? Mind and Matter, 15(1):7–28.

Sergent, C. and Naccache, L. (2012). Imaging neural signatures of consciousness:'what','when','where'and 'how'does it work? Archives italiennes de biologie, 150(2/3):91–106.

Seth, A. (2007). Models of consciousness. Scholarpedia, 2(1):1328.

Seth, A. K. (2018). Consciousness: The last 50 years (and the next). Brain and neuroscience advances, 2:2398212818816019.

Seth, A. K., Izhikevich, E., Reeke, G. N., and Edelman, G. M. (2006). Theories and measures of consciousness: an extended framework. Proceedings of the National Academy of Sciences, 103(28):10799–10804.

Setsompop, K., Kimmlingen, R., Eberlein, E., Witzel, T., Cohen-Adad, J., McNab, J. A., Keil, B., Tisdall, M. D., Hoecht, P., Dietz, P., et al. (2013). Pushing the limits of in vivo diffusion mri for the human connectome project. Neuroimage, 80:220–233.

Shahriari, B., Bouchard-Côté, A., and Freitas, N. (2016). Unbounded bayesian optimization via regularization. In Artificial intelligence and statistics, pages 1168–1176. PMLR.

Shanahan, M. (2010). Metastable chimera states in community-structured oscillator networks. Chaos: An Interdisciplinary Journal of Nonlinear Science, 20(1):013108.

Shannon, C. E. (1948). A mathematical theory of communication. The Bell system technical journal, 27(3):379–423.

Shear, J. and Varela, F. J. (1999). The view from within: First-person approaches to the study of consciousness. Imprint Academic.

Sinha, N., Dauwels, J., Kaiser, M., Cash, S. S., Brandon Westover, M., Wang, Y., and Taylor, P. N. (2017). Predicting neurosurgical outcomes in focal epilepsy patients using computational modelling. Brain, 140(2):319–332.

Sitt, J. D., King, J.-R., El Karoui, I., Rohaut, B., Faugeras, F., Gramfort, A., Cohen, L., Sigman, M., Dehaene, S., and Naccache, L. (2014). Large scale screening of neural signatures of consciousness in patients in a vegetative or minimally conscious state. Brain, 137(8):2258–2270.

Solovey, G., Alonso, L. M., Yanagawa, T., Fujii, N., Magnasco, M. O., Cecchi, G. A., and Proekt, A. (2015). Loss of consciousness is associated with stabilization of cortical activity. Journal of Neuroscience, 35(30):10866–10877.

Sporns, O., Tononi, G., and Kötter, R. (2005). The human connectome: a structural description of the human brain. PLoS computational biology, 1(4):e42.

Stephan, K. E., Weiskopf, N., Drysdale, P. M., Robinson, P. A., and Friston, K. J. (2007). Comparing hemodynamic models with dcm. Neuroimage, 38(3):387–401.

Stinson, C. and Sullivan, J. (2018). Mechanistic explanation in neuroscience. The Routledge Handbook of Mecahnisms and Mechanical Philosophy, pages 375–387.

Swanson, L. R. (2018). Unifying theories of psychedelic drug effects. Frontiers in pharmacology, page 172.

Tagliazucchi, E. (2017). The signatures of conscious access and its phenomenology are consistent with large-scale brain communication at criticality. Consciousness and cognition, 55:136–147.

Tagliazucchi, E., Chialvo, D. R., Siniatchkin, M., Amico, E., Brichant, J.-F., Bonhomme, V., Noirhomme, Q., Laufs, H., and Laureys, S. (2016a). Large-scale signatures of unconsciousness are consistent with a departure from critical dynamics. Journal of The Royal Society Interface, 13(114):20151027.

Tagliazucchi, E. and Laufs, H. (2014). Decoding wakefulness levels from typical fmri resting-state data reveals reliable drifts between wakefulness and sleep. Neuron, 82(3):695–708.

Tagliazucchi, E., Roseman, L., Kaelen, M., Orban, C., Muthukumaraswamy, S. D., Murphy, K., Laufs, H., Leech, R., McGonigle, J., Crossley, N., et al. (2016b). Increased global functional connectivity correlates with lsd-induced ego dissolution. Current Biology, 26(8):1043–1050.

Tart, C. (1976). _.'the basic nature of altered states of consciousness, a system approach'. Journal of Transpersonal Psychology, 1.

Tart, C. T. (1972). Altered states of consciousness. Doubleday.

Tassi, P. and Muzet, A. (2001). Defining the states of consciousness. Neuroscience & Biobehavioral Reviews, 25(2):175–191.

Thom, R. (1992). Prédire n'est pas expliquer. Revue Philosophique de la France Et de l, 182(2).

Timmermann, C., Roseman, L., Schartner, M., Milliere, R., Williams, L. T., Erritzoe, D., Muthukumaraswamy, S., Ashton, M., Bendrioua, A., Kaur, O., et al. (2019). Neural correlates of the dmt experience assessed with multivariate eeg. Scientific reports, 9(1):1–13.

Toker, D. and Sommer, F. T. (2019). Information integration in large brain networks. PLoS computational biology, 15(2):e1006807.

Tononi, G. (2004). An information integration theory of consciousness. BMC neuroscience, 5(1):1–22.

Tononi, G. (2008). Consciousness as integrated information: a provisional manifesto. The Biological Bulletin, 215(3):216–242.

Tononi, G. and Edelman, G. M. (1998). Consciousness and complexity. science, 282(5395):1846–1851.

Tononi, G., Sporns, O., and Edelman, G. M. (1994). A measure for brain complexity: relating functional segregation and integration in the nervous system. Proceedings of the National Academy of Sciences, 91(11):5033–5037.

Tsuchiya, N., Andrillon, T., and Haun, A. (2020). A reply to "the unfolding argument": Beyond functionalism/behaviorism and towards a science of causal structure theories of consciousness. Consciousness and Cognition, 79:102877.

Tsuchiya, N., Wilke, M., Frässle, S., and Lamme, V. A. (2015). No-report paradigms: extracting the true neural correlates of consciousness. Trends in cognitive sciences, 19(12):757–770.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., and Joliot, M. (2002). Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain. Neuroimage, 15(1):273–289.

Ulmasov, D., Baroukh, C., Chachuat, B., Deisenroth, M. P., and Misener, R. (2016). Bayesian optimization with dimension scheduling: Application to biological systems. In Computer Aided Chemical Engineering, volume 38, pages 1051–1056. Elsevier.

Vaitl, D., Birbaumer, N., Gruzelier, J., Jamieson, G. A., Kotchoubey, B., Kübler, A., Lehmann, D., Miltner, W. H., Ott, U., Pütz, P., et al. (2005). Psychobiology of altered states of consciousness. Psychological bulletin, 131(1):98.

Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E., Yacoub, E., Ugurbil, K., Consortium, W.-M. H., et al. (2013). The wu-minn human connectome project: an overview. Neuroimage, 80:62–79.

Váša, F., Shanahan, M., Hellyer, P. J., Scott, G., Cabral, J., and Leech, R. (2015). Effects of lesions on synchrony and metastability in cortical networks. Neuroimage, 118:456–467.

Wacker, D., Wang, S., McCorvy, J. D., Betz, R. M., Venkatakrishnan, A., Levit, A., Lansu, K., Schools, Z. L., Che, T., Nichols, D. E., et al. (2017). Crystal structure of an lsd-bound human serotonin receptor. Cell, 168(3):377–389.

Wang, P., Kong, R., Kong, X., Liégeois, R., Orban, C., Deco, G., Van Den Heuvel, M. P., and Thomas Yeo, B. (2019). Inversion of a large-scale circuit model reveals a cortical hierarchy in the dynamic resting human brain. Science advances, 5(1):eaat7854.

Wischnewski, K. J., Eickhoff, S. B., Jirsa, V. K., and Popovych, O. V. (2022). Towards an efficient validation of dynamical whole-brain models. Scientific reports, 12(1):1–21.

Wong, K.-F. and Wang, X.-J. (2006). A recurrent network mechanism of time integration in perceptual decisions. Journal of Neuroscience, 26(4):1314–1328.

Wu, F., Mao, X., and Kloeden, P. E. (2013). Discrete razumikhin-type technique and stability of the euler–maruyama method to stochastic functional differential equations. Discrete & Continuous Dynamical Systems, 33(2):885.

Zhang, X.-S., Roy, R. J., and Jensen, E. W. (2001). Eeg complexity as a measure of depth of anesthesia for patients. IEEE transactions on biomedical engineering, 48(12):1424–1433.