Theoretical models and measures of conscious brain network dynamics

An integrative approach

Camilo A. Miguel Signorelli

SORBONNE UNIVERSITÉ & UNIVERSITAT POMPEU FABRA
THESIS DOCTORAL/ YEAR 2020

THESIS SUPERVISOR
Gustavo Deco
Bechir Jarraya

Brain-Cognition-Behaviour Doctoral School, Sorbonne Universié, France.
Department of Experimental and Health Sciences, Universitat Pompeu Fabra, Spain.
Cognitive Neuroimaging Unit, Institut National de la Santé et de la Recherche Médicale U992, France.
Center for Brain and Cognition, Computational Neuroscience Group, Universitat Pompeu Fabra, Spain.

To my family

Acknowledgment

Agradezco con lo más profundo de su corazón, ha cada persona que a contribuido con discusiones, consejos, motivación y apoyo dentro de las mas variadas formas.

I thank with the deepest of my heart, each person who has contributed with discussions, advice, motivation and support in the most varied ways.

Sommaire

Dans le domaine des neurosciences de la conscience, il existe actuellement une tendance à opposer et comparer les modèles de conscience. Même si le dernier mot est empirique, des efforts théoriques sont essentiels pour replacer les hypothèses conceptuelles et les résultats expérimentaux dans leur contexte. À partir de là, nous pouvons concevoir de meilleures évaluations et répondre à la question de savoir quel modéle est optimal. Sur cette voie, cette thèse explore les modèles et les approches intégratives computationnelles. Le document classe les modèles scientifiques de conscience selon leur "profil explicatif". Les données empiriques sont décrites à la lumière de la théorie des réseaux. Ensuite, des outils informatiques inspirés de l'intégration conceptuelle de deux modèles influents sont mis en œuvre pour quantifier les différences entre l'état éveillé et anesthésié. Enfin, la thèse introduit de nouveaux concepts pour éviter le réductionnisme actuel de certains modèles, ce qui entraîne le texte vers des discussions controversées. Cette thèse est un travail théorique et conceptuel inspiré de résultats empiriques, qui tente de révéler la puissance des modèles computationnels et mathématiques afin de développer des hypothèses testables et de mieux comprendre les neurosciences de la conscience.

Resumen

En el área de la neurociencia de la conciencia, actualmente existe una tendencia de contrastar y comparar modelos de conciencia. Aunque la palabra final es empírica, esfuerzos teóricos son esenciales para poner en contexto tanto suposiciones conceptuales como resultados experimentales. Desde allí, podemos diseñar mejores pruebas y responder a la pregunta sobre que modelo es optimo. En esta dirección, esta tesis explora modelos y enfoques computacionales integrativos. El documento clasifica modelos científicos de conciencia de acuerdo a sus "perfiles explicativos". Resultados empíricos son describidos a la luz de teoría de redes. Luego, herramientas computacionales inspiradas por la integración conceptual de dos de los mas influyentes modelos son implementados para cuantificar las diferencias entre la condiciones de despierto y anestesiado. Finalmente, la tesis introduce nuevos conceptos para evadir el actual reducionismo de algunos modelos, orientando el texto hacia polémicas discusiones. Esta tesis es un trabajo teórico y conceptual inspirado por resultados empíricos que intenta revelar el poder de modelos computacionales y matemáticos en la búsqueda de desarrollar hipótesis testeables y entender mejor la neurociencia de la conciencia.

Abstract

In the field of neuroscience of consciousness, there is a current trend to contrast and compare existing models of consciousness. Even though the final word is empirical, theoretical efforts are essential to place both, conceptual assumptions and experimental results in context. From that, we can design better assessments and answer the question about what model is optimal. In this direction, this thesis explores models and computational integrative approaches. The document classifies scientific models of consciousness according to their "explanatory profile". Empirical data is described in light of network theory. Then, computational tools inspired by the conceptual integration of two influential models are implemented to quantify differences between awake and anaesthetic conditions. Finally, the thesis introduces new concepts to avoid the current reductionism of some models, pushing the text to controversial discussions. This thesis is a theoretical and conceptual work inspired by empirical results, attempting to reveal the power of computational and mathematical models in order to develop testable hypotheses and understand better the neuroscience of consciousness.

Contents

List of Figures

List of Tables

Chapter 1

INTRODUCTION

Understanding consciousness is not only an intellectual challenge, but it is also a mandatory step forward to improve diagnosis of mental illnesses, rehabilitation treatments, optimal human-robot interactions, as well as answering controversial questions about multidimensional measures of consciousness [Bayne et al., 2016] or animal consciousness [Barron and Klein, 2016].

During centuries, the differences between subjective experiences and objective reality were mostly discussed within philosophy. William James was one of the pioneer scientists moving the question to a more empirical domain. Later, thinkers such as Bernard Baars, Gerald Edelman, or Francis Crick, among others, approached the question from the lens of cognitive neuroscience. Currently, the most influential theories are global neuronal workspace (GNW) [Mashour et al., 2020] and integration information theory (IIT) [Tononi et al., 2016]. Competing accounts include Higher-Order Thought Theory (HOT), Recurrent Processing Theory (RPT) [Lamme, 2010], mechanisms such as Predictive Processing (PP) [Parr et al., 2019], or thalamocortical loops [Redinbaugh et al., 2020]. Besides, novel methods to explore conscious experience from a first-person perspective [Petitmengin et al., 2019] have been proposed to complement neuroscientific studies. However, no current model is unequivocally accepted and the different methodologies used to study conscious experience are hardly integrated.

In an early scientific phase, it is natural to expect many competing models to co-exist. Moving forward, one would like to compare them. The most widespread experimental method to study consciousness involves the *neural correlates of consciousness* (NCCs). Since their introduction more than 20 years ago, science has made considerable progress [Koch et al., 2016], but there exists no agreement on how to interpret them. For example, some evidence supports the GNW [Van Vugt et al., 2018], while other experiments argue in favour of IIT [Siclari et al.,

2017]. Decisive answers are expected to result from *crucial experiments* [Reardon, 2019], sometimes focusing on predictions such as localization of NCCs [Boly et al., 2017], but often applying concepts out of context from which they arise (e.g [Noel et al., 2019]), and neglecting the models' underlying philosophical or physical assumptions.

A more attentive reading might instead reconcile models and delimit the level of descriptions where theories can be compared [Maillé and Lynn, 2020]. One example pertains to the role of pyramidal neurons in cortical layer V as a mechanism to integrate the influence of feedback connectivity in cortico-cortical loops and higher-order thalamocortical loops [Suzuki and Larkum, 2020]. In this line, dynamical concepts such as criticality [Tagliazucchi, 2017] and harmonic modes [Atasoy et al., 2019] aim to reconcile different aspects of consciousness and brain function. Another example pertains to formal models that directly approach the phenomenology of experience [Prentner, 2019, Yoshimi, 2007, Signorelli et al., 2020b] and more abstract axiomatic models that uncover the hidden assumptions of cognitive models and make their relation to subjective experience more transparent [Northoff et al., 2019, Tull and Kleiner, 2020].

This strategy, however, necessitates a deep theoretical and methodological understanding as well as placing results in a broader context. Across this document, we will construct a first attempt of such a framework, with the explicit aim of explaining, comparing and integrating current models and supporting evidence in the field of consciousness research.

The following main chapters are based on published papers or papers under review by specialized journals in cognitive neuroscience (see list of publications and references). Specifically, the second chapter classifies and compares the main assumptions and premises of major models of consciousness [Signorelli et al., 2021b]. The third chapter suggests a parsimonious multilayer approach to integrate current neural and biological evidence about signatures of consciousness [Signorelli et al., 2021a]. The fourth chapter is devoted to applying the measure of intrinsic ignition as a measurement that integrates concepts from two of the most influential neural models of consciousness [Signorelli et al., 2021c]. Finally, the fifth chapter introduces and discusses new concepts to integrate more complex theoretical and empirical accounts [Signorelli and Meling, 2021]. As a comprehensible whole, the main arguments and results are combined and discussed in the conclusion section, with the hope to lay the foundations for an exciting future research program.

Chapter 2

MODELS OF CONSCIOUSNESS

Models of consciousness aim to interpret empirical evidence and inspire new experimental protocols to reveal the structure of conscious experience. Nevertheless, no current model is univocally accepted on either theoretical or empirical grounds. Moreover, a straightforward comparison is difficult for conceptual reasons. In particular, we argue that different models explicitly or implicitly subscribe to different notions of what constitutes a satisfactory explanation, use different tools in their explanatory endeavours and even aim to explain very different problems. We thus present a framework to compare existing models in the field with respect to what we call their ”explanatory profiles”. We focus on the following minimal dimensions: mode of explanation, mechanisms of explanation, and target of explanation. We also discuss the empirical consequences of the discussed discrepancies among models. This approach may eventually lead to a better design of future testing experiments, identifying driving assumptions, theoretical and experimental predictions, strengths and weaknesses. Finally, our conclusion points to more integrative theoretical research, where axiomatic models may play a critical role in solving current theoretical and experimental contradictions.

2.1 Introduction

Models of consciousness set out to provide a principled description of how the physical domain relates to conscious experience [Seth, 2007, Seth, 2009, Durham et al., 2020]. In the last decades, consciousness researchers put forward an abundance of conceptual and formal proposals, drawing from neuroscience, physics, mathematics, philosophy or experimental psychology. In an early scientific phase, it is natural to expect many competing models to develop in parallel to each other. A more mature stage should entail a substantial cross-talk between

them, aiming at distilling critical similarities and differences between them, extracting precise empirical predictions [Boly et al., 2017] and lastly, eliminating falsified frameworks through a set of crucial empirical experiments, as presently envisioned in ref [Reardon, 2019].

An alternative would be to demand that competing theories need to *converge* to a unified, synthesized account in order to make progress [Northoff and Lamme, 2020]. In the following paragraphs, we posit that there are currently several, serious impediments to both crucial experiments but also convergence approaches. Arguably, more conspicuous and empirically tangible differences in the theories (such as their postulated neural-correlates: prefrontal cortex or posterior hot-zone) derive from much deeper, implicitly held deviations in theoretical and philosophical assumptions. In particular, proponents of different theories seem to substantially disagree on what would constitute a *satisfactory explanation* of consciousness in the first place. Therefore, the aims of these theories are sometimes different. Once all these discrepancies are fully made explicit, the major models can start to enrich each other in a meaningful way. At least in some cases, theoretical misalignments between the models might boil down to different angles of looking at the same problem.

Although navigating through such a highly diversified theoretical landscape remains challenging, there is hope that one could account for the large variation in the field by using only relatively few axes of comparison. To this date, there have been only a couple of systematic attempts to thoroughly compare the contemporary models of consciousness [Klink et al., 2015, Block, 2009, Northoff and Lamme, 2020]. Arguably, however, all these endeavours have not gone beyond simply collating the theories' different explanatory targets and their main employed paradigms. In addition, the way the theories have been classified so far (such as targeting either *phenomenal* or *access* consciousness [Northoff and Lamme, 2020] or whether they are related to the *pre-* or *post*-stimulus neural activity) might have created more confusion than elucidation [Rosenthal, 2002b, Block, 2007, Rosenthal, 2020].

In order to identify the most critical points of contention in the field, we first introduce the philosophical landscape (Section 2.2), then we explicate three crucial directions in which frameworks diverge most noticeably with respect to their explanatory pretense (Section 2.3). Among these directions we distinguish: mode of explanation (mechanistic vs. unificationist), mechanism of explanation (functional vs. causal), and target of explanation (quality vs. quantity of consciousness). These dimensions are discussed along contemporary models of consciousness, which creates a multi-dimensional explanatory profile for each. We also dis-

cuss two empirical consequences of this (Section 2.4). We conclude by advocating for a more integrative approach, hinting at already existing empirical and theoretical (mathematical) tools (Section 2.5).

We narrow our scope to 13 influential models. While being aware that our selection of models is not exhaustive, we do hope that our work would still spark meaningful discussions and inspire orderly and structured comparisons in the field.

2.2 The problem of Consciousness

Philosophy does not solve problems, it rather helps to emphasize and reconceptualize them to make problems amenable to scientific investigation. Therefore, we first look at some of the main concepts relevant to the problem of consciousness, which are often implicitly held by current theories.

2.2.1 ”What-it-is-like” to have an experience

Most of the discussions of ”phenomenal consciousness” in the contemporary neuroscience of consciousness go back to work of Thomas Nagel from the 1970ies that emphasizes the ”what-it-is-likeness” of conscious experience – the subjective raw feeling that seems to evade scientific inquiry [Nagel, 1986, Thomas Nagel, 1974]. Nagel argued that the purely objective study of an entity, such as the one science provides, does not allow any inference about the subjective character of being such an entity. This has sometimes been misconstrued as the claim that a purely subjective phenomenon such as consciousness cannot be studied at all within the objective framework of science and should best be left alone, an idea which has been forcefully disputed by John Searle [Searle, 1998, Searle, 2000].

Nevertheless, the idea that consciousness includes an inner, subjective perspective poses perhaps the biggest challenge to any model of consciousness. One possible reply denies the reality of this inner, subjective perspective. Most prominent here is the eliminativism proposed early on by Daniel Dennett [Dennett, 1988, Dennett, 1993], which got traction recently but still comes with its own issues [Frankish, 2017, Chalmers, 2018]. Another possibility is to account

for what-it-is-likeness within science. We will mention, later in this paper, some of the early frameworks and the most recent models still relevant today.

Another related distinction is the one between the notions of ”phenomenal” and ”access” consciousness [Block, 2005, Block, 1995]. The former refers explicitly to this what-is-likeness, sometimes assumed to correspond to a ”minimal sense” of conscious experience without necessarily requiring reportability [Metzinger, 2020]. In particular, some theories of consciousness claim to be about exactly this phenomenal aspect and thus carry a distinct explanatory pretense. The latter notion, ”access” consciousness, corresponds to centralized availability for processing of information and the reportability of a conscious experience. It also refers to phenomena which are closely related to consciousness in other aspects (e.g. attention or meta-cognition). It is an open debate in the scientific study of consciousness whether an explanation purely in terms of access consciousness is truly satisfactory or whether it is not in fact the *only* scientifically rigorous approach.

Much of the prominence that *phenomenal* consciousness received as a potential scientific topic can be traced back to the work of David Chalmers who introduced the notion of the ”hard problem of consciousness” [Chalmers, 1995b], the difficulty to explain why certain forms of physical information processing should feel like anything at all or how physical and phenomenal facts are related to each other. Chalmers argued against the reducibility of consciousness and initially advocated a ”natural dualism” [Chalmers, 1997]. Chalmers later made the case for ”panpsychism” [Strawson, 2006, Chalmers, 2013b, Goff, 2019], the view that consciousness is irreducible and ubiquitous in nature. Panpsychism is related to ”dual-aspect” monisms that consider consciousness and physics merely as two aspects of a single underlying reality [Atmaspacher, 2014].

An alternative way to approach the problem of consciousness is the ”biological naturalism” of John Searle [Searle, 2000], an example of a ”monist” reply to Chalmers’ proposed dualism. While Searle acknowledges the phenomenal character of consciousness, he finds it ”obvious” that it emerged from the brain similar to the way bile is produced by the liver. However, to date, there no viable mechanism has been identified for this process. The problems for Searle’s approach are representative of all materialistic approaches to consciousness that are discussed later in the paper, at least where they pretend to shed light on phenomenal consciousness.

Another monistic response is to invert the hard problem and argue that the physical world

is a product of consciousness. This idealistic position was once the dominant world-view in much of Western and Eastern culture but received a massive blow in the 20th century. Most philosophers and scientists do not take this option seriously anymore, but the climate seems to change with more and more scholars advocating such a view [Hoffman, 2008, Hoffman and Prakash, 2014, Kastrup, 2017, Chalmers, 2019].

A different response to the dualist proposal of Chalmers was given by Francisco Varela. Whereas Chalmers postulated that it seems as if there was a need for ”extra ingredients” to physical theory and whereas monisms express a metaphysical presupposition, Varela suggested to regard consciousness and brain processes as mutually constraining phenomena [Thompson, 2007, Rodríguez, 2008, Varela, 1996] that ground an *empirical* approach to consciousness. For example, to stay with the mainstream idea in neuroscience, even if consciousness was an emergent phenomenon, the empirical question about the causal efficacy of consciousness would still be left unanswered, i.e. whether or not consciousness was able to influence its physical substrate [Thompson and Varela, 2001] or how any supposed ”backreaction” would manifest itself in scientific data. More generally, Varela's ”neurophenomenology” is an adaption of earlier continental approaches (”Phenomenology” [Gallagher and Zahavi, 2008, Kaufer and Chemero, 2015]) to cognitive neuroscience that seek to uncover the necessary structures of all experience (including the one that gives rise to scientific knowledge).

After all these proposals a scientific problem remains – even though it comes in very different guises. We either need to explain why the illusion of consciousness appears as something real and vivid or argue that there is no problem of consciousness over and above the problems of ”access” consciousness (problems which still need to be solved though!). Alternatively, we would need to specify the reductive causal relation between ”phenomenal” consciousness and matter (e.g. solve the ”hard problem” and give a model of how consciousness actually emerges from the brain), or explain how explanations in terms of brain dynamics and conscious experience mutually constrain each other.

2.2.2 From philosophy to scientific models of consciousness

A framework for the study of consciousness refers to a group of premises and assumptions to guide experiments and interpret general results. More specifically, a model of consciousness conveys concrete hypotheses, predictions, mechanisms, and explanations of the associated phe-

nomena. A proper theory for consciousness consists, therefore, in a set of assumptions plus a concrete model to enable the testing of (empirical or theoretical) predictions and eventually its implementation and manipulation. The approaches discussed in this paper correspond to models that mostly operate on implicit assumptions. They do not resemble proper theories in the sense just outlined, but we will use both the terms ”theory” and ”model” interchangeably in the remainder of this article to better conform to the literature.

A first step to better understand models of consciousness is to make explicit their underlying philosophical assumptions. These assumptions inform and influence models of consciousness. In figure 2.1, we summarize the main relationships between philosophy, early and modern models of consciousness.

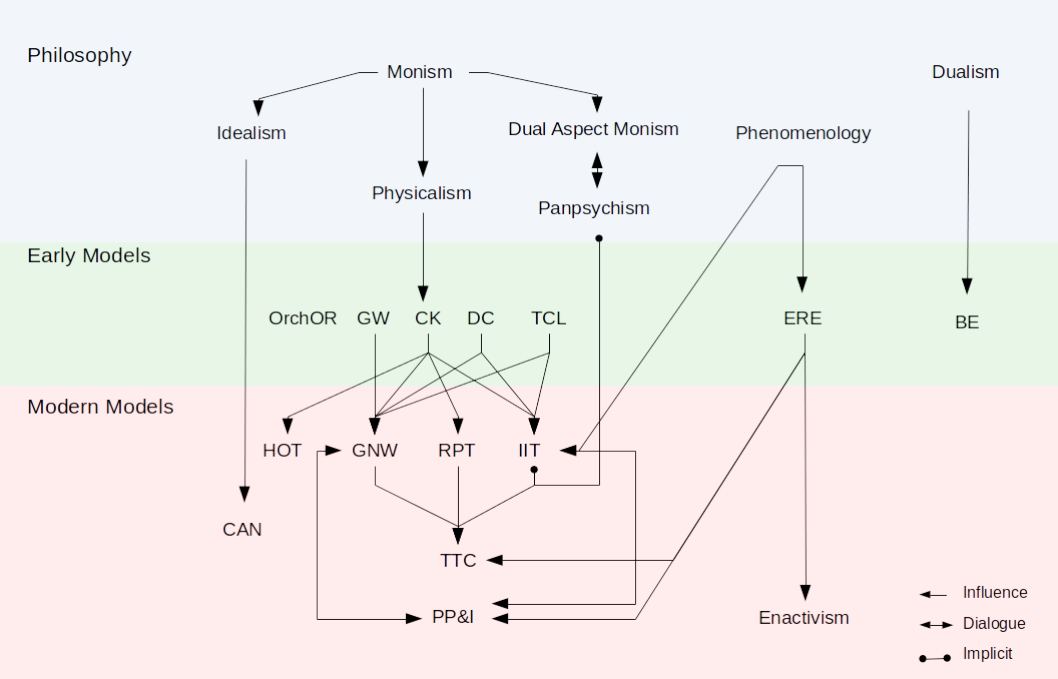


Figure 2.1: Philosophy and Models of consciousness. Relationship and influences between philosophy, early and modern models of consciousness. For acronyms see text.

In this article we consider early models such as the mechanistic model of Crick and Koch (CK), Dynamical Core (DC), Orchestrated Objective Reduction (OrchOR), Global workspace (GW), Thalamo-Cortical loops and Sensorimotor Couplings (TCL), the dualist proposal by Beck and Eccles (BE), and Enactive and radical Embodiment (ERE). Modern models are Global Neuronal Workspace (GNW), Higher-Order Thought Theory (HOT), Recurrent Processing Theory (RP), Predictive Processing and Interoception (PP&I), Integrated Information Theory (IIT), Conscious Agents Networks (CAN), and Temporo-spatial Theory of Consciousness (TTC).

Early models typically inform later ones, sometimes via direct succession (e.g. GW/GNW), or via integration of concepts developed in previous models (e.g. TTC). Models may also initiate dialogues and remain under dynamic influences with each other. For example, both postulates of GNW and IIT do remain consistent to some extent with mechanisms put forward by the predictive coding approach (PP&I).

2.3 Method of classification

Our initial examination of the current models of consciousness involves placing every theory within 3 orthogonal dimensions corresponding to the positions they espouse with regard to explanation. The first axis stands for the **mode of explanation** assumed by the model (mechanistic vs. unificationist), the second for the **mechanism of explanation** (functional vs.causal), and the third for the **target of explanation** (quantity vs. quality of consciousness). This results in ”explanatory profiles” for each theory, which accounts for a substantial amount of variance in the theoretical landscape.

Axis	**First direction**	**Second Direction**
Mode of explanation	Mechanistic	Unificationist
Mechanism of explanation	Function	Causal
Target of explanation	Quality	Quantity

Table 2.1: Axes for a three-dimensional classification map. The axes include three main axes with their two opposite directions, represented by positive or negative values.

2.3.1 Mode of explanation

One of the most important hallmarks of a successful theory is its explanatory power. Nonetheless, the very notion of explanation (what it means to successfully *explain* something), despite its deceptive simplicity, can be unpacked as highly heterogeneous [Nagel, 1961, Salmon, 1990, Woodward, 2019, Woodward, 2013, Woodward, 2004, Strevens, 2004, Colombo, 2017]. In particular, the very notion of ”explaining consciousness”, although ubiquitous in the literature, constitutes a deceptive ”umbrella term”, under which one can identify strikingly different theoretical goals. That ”consciousness” is frequently used as an umbrella term for various mental phenomena has been often stated in the literature (e.g. [Van Gulick, 2018]). Here we wish to emphasize that the same is true for the seemingly innocent notion of its ”explanation”. The field

of consciousness science has been flooded with numerous works contemplating whether the full explanation of this phenomenon is even feasible [McGinn, 1988, Chalmers, 1995a, Hohwy and Frith, 2004]. Here we focus on two alternatives that seem particularly relevant in the domain of consciousness studies: **mechanistic approaches** and **unification** [Salmon, 1990, Strevens, 2004].

A mechanistic approach (sometimes known as "causal-mechanical") posits that a particular occurrence is *explained* once we demonstrate how it "fits into" the causal network of the world. A *causal process* can be defined as a physical, typically spatiotemporally continuous process, which is capable of transmitting a "mark". A mark is understood as a counterfactual, local modification to the structure of a process itself [Woodward, 2004]. A theory of consciousness that holds the mechanistic view would accordingly imply that the satisfactory explanation of subjective experience requires meticulous unfolding of the chain of causes and effects occurring in the nervous system that unambiguously leads to a conscious experience of some kind. Such a position imposes a strictly empirical agenda, investigating which neurophysiological events precede and give rise to a conscious experience. Importantly, the driving force that impels such theoretical accounts is a philosophical *assumption* that such a chain of causes is *available* at some objective level of description at some spatiotemporal scale (of the brain). Naturally, these frameworks tend to cluster together under branches of ontological or methodological reductionism.

On the other hand, the unification stance puts less emphasis on seeking a discrete set of causal interactions. In this case, a successful explanation is a matter of providing a *unified account* of a range of different phenomena, previously thought as unrelated - or related in a mysterious or seemingly arbitrary way (core historical examples are Maxwell's unification of electricity and magnetism or Newton's unification of terrestrial and celestial motion). Explaining consciousness under the unificationist framework would give priority to demonstrating how the phenomenon of consciousness is embedded into a parsimonious, coherent framework. Unificationist theories typically strive to replace a theory which explains a phenomenon by making special assumptions about parameter values with another theory that avoids making these assumptions (by either removing them or showing that the explananda follow for all parameter values, or by making adequate mathematical identifications [Glymour, 1980]). The particular aim of these explanations can be unpacked as "reducing contingencies" [Woodward, 2013]. As a result, the unificationist seeking a satisfactory explanation would be inclined to associate less

with empiricism, and more with formal tools, mathematics and non-reductive philosophy. On the other hand, unifactionary explanations are often conceived to be most relevant for physics but at odds with biology and neuroscience [Anderson and Chemero, 2013, Bayne, 2018].

Such a view is often not compatible with the mechanistic approach. Mechanistic accounts, implicitly or explicitly, always assume the realist notion of causation in space and time. Conversely, unificationist accounts postulate that explaining consciousness cannot be exhausted by studying the spatiotemporal chain of causes and effects localized in the brain. In particular, assumptions about the causal order and spatiotemporal descriptions are often being put on hold, and treated as an *explanandum* (a phenomenon to be explained) rather than *explanans* (the explanation) [Barnes, 1992].

Numerous examples show that both types of explanation have contributed to significant progress in science [Barnes, 1992, Woodward, 2004]. Our aim is not to assess which approach produces better explanations, but to simply recognize that different frameworks of consciousness will be inclined to differ already on this very basic assumption.

2.3.2 Mechanism of explanation

We define the term *mechanism* after Illari & Williamson as ”entities and activities organized in such a way that they are *responsible* for the phenomenon” [Illari and Williamson, 2012]. Importantly, a mechanism of explanation is not to be conflated with the mode of explanation, i.e. while the mode informs about the ultimate *aim* of the explanation (what constitutes a satisfactory answer to the ”why” question), the mechanism demonstrates which particular *tool* can get us towards that predefined aim.

In the context of consciousness science, an increasingly popular division, introduced in ref [Doerig et al., 2019] and specifically addressing different kinds of mechanisms, distinguishes theories as either **functional** or **causal**. The stance of functionalism primarily states that consciousness can be generated as long as a particular function is realized [Block, 1996], without any specific constraint on the exact causal machinery behind it. In principle, any system may become conscious as long as it executes the functions associated with conscious experience. On the opposite side, causal theories ascribe the utmost relevance to the physical implementation: establishing what are the elementary mechanisms and what they *do* in terms of causal

actions. Interestingly, causal theories do not necessarily unsubscribe from the view that a particular function might be typically associated with consciousness, but the burden of explanation is placed on *how* such a function is implemented. In those terms, causal models support the idea that only a system with the right causal structure *implemented* will lead to conscious experience.

One could further illustrate this distinction by introducing the idea of a structure-preserving map (M) between two objects. Causal models insist that the causal structure of the system (S) gives rise to consciousness, $M : S \rightarrow C$, where the arrow refers to a mapping that preserves causal structure. Functional models, however, would argue that the functional structure of a system (F) gives rise to consciousness, $M : F \rightarrow C$, and thus it is the function that is preserved by the arrow. These two types of models appear exclusive, in the sense that S and F are different objects. However, if we focus on the nature of the map, independently of their objects, the difference becomes a question of "degree". Causal theories assume there is only one way to preserve the causal structure and the phenomenology of subjective reports. This is tantamount to postulating an isomorphism between domains (there exists only one such arrow) [Tsuchiya et al., 2016]; while functional theories would claim/assume that there are multiple ways (arrows) to preserve the function and therefore giving rise to consciousness, i.e. $M^{'}$,$M^{''}$, $M^{'''}$, etc.

2.3.3 Target of explanation

Lastly, the target of an explanation corresponds to the aspects of a particular phenomenon that scientists intend to explain. The most basic distinction that emerged over the last decades is the one between **quality** and **quantity**. The quality of consciousness is what makes consciousness feel *the way* it does (cf. also Section 2.2.1), and the quantity corresponds to what makes the system conscious rather than unconscious. A model targeting quality should therefore account for why any stimulus should feel a *particular way*, what makes an experience spatial, visual, auditory, painful, or temporal, while a model targeting quantity intends to account for global markers differentiating conscious vs unconscious systems.

A full-fledged theory of consciousness needs to explain both the quantity and quality aspects of subjective experience. Several theories tend to focus exclusively on global markers differentiating conscious vs unconscious systems; and the problem of quality in such cases is

delegated to the external world (i.e. sensory cues feel the way they do solely because they 'carry' their quality from the environment, or stimulate 'correct', labeled receptors). However, there is a great deal of philosophical grounds, all pointing to many reasons on why such a "delegation" is problematic, see arguments on brains in vats [Horgan et al., 2004], inverted spectrum [Shoemaker, 2000], actual cases of perceptual variation [Block, 1999], the fact that sensory characters correlate much more with neural patterns than anything else in the external world [Pautz, 2014], dramatic perceptual alterations in psychedelic experiences [Bayne and Carter, 2018], perceptual illusions, and the generic non-preservation of phenomenal and environmental structure [Prakash et al., 2020], etc. A much more ambitious project would strive to explain why an experience feels the way it does, for example based on the internal architecture and dynamics of the brain, without deflectionary referrals to the external world as something that 'stores' or 'produces' any qualities.

2.4 Classifying models of consciousness

According to our predefined analytical dimensions, we now classify and discuss selected models of consciousness in a three dimensional map (Figure 2.2).

2.4.1 Mode of explanation

Based on our classification, we unpack the explanatory profile of each theory. To this end, we have first focused on reviewing the relevant articles to describe the preselected theories specifically with respect to their mode of explanation, i.e. whether they seek mechanistic or unificationism types of explanations. Consequently, we have scoured representative works for statements pinpointing what a satisfactory explanation would look like according to each theory.

Already at this stage we have found discrepancies, most of the theories leaning clearly towards one or another mode. Some frameworks, including GNW, and HOT, gave us consistent clues to classify them under the mechanistic cluster. A statement that would drive us to the mechanistic classification could be, for example, the following statement by Rosenthal, a proponent of HOT: "we understand something only when we can explain it, and explaining a

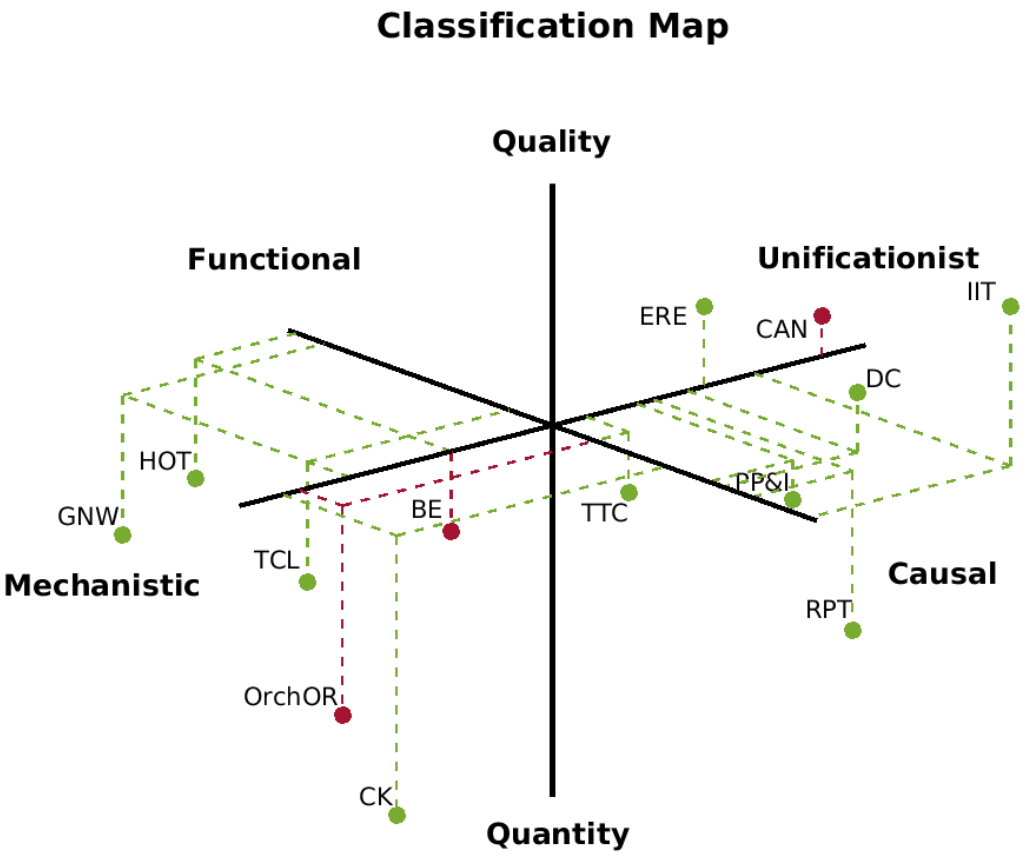


Figure 2.2: Classification of the models. Models inhabit different explanatory spaces according to our three analytical dimensions. Acronyms stand for Crick and Koch (CK), Dynamical Core (DC), Orchestrated Objective Reduction (OrchOR), Thalamo-Cortical loops and Sensorimotor Couplings (TCL), Enactive and Radical Embodiment (ERE), Global Neuronal Workspace (GNW), Higher-Order Thought Theory (HOT), Recurrent Processing Theory (RP), Predictive Processing and interoception (PP&I), Integrated Information Theory (IIT), Orchestrated Objective Reduction (OrchOR), Conscious Agents Network (CAN), Beck and Eccles (BE), and Temporo-spatial theory of consciousness (TTC).

natural phenomenon typically if not always means locating it in its distinctive causal nexus" [Rosenthal, 2008], or by Dehaene, a proponent of GNW: "tools of cognitive psychology and neuroscience may suffice to analyze consciousness" [Dehaene and Naccache, 2001]. On the other hand, unificationist propensities are more apparent in frameworks such as IIT: "IIT provides a principled explanation for several seemingly disparate facts about the PSC"; PP "... it could unify existing approaches under a single overarching principle (i.e., the FEP)" [Wiese and Friston, 2020]; or CAN "if we want to go beyond this 'applied science' and understand the true nature of the mind and the reality beyond it, we can't look to neurons or brains" [Fields et al., 2018]. The full list of relevant quotes, justifying each model's classification can be found in Table A.1.

This first dimension clearly constitutes a parsimonious dividing force, introducing tensions between the models' individual goals. Arguably, identifying which mode of explanation a the-

ory supports can also predict how a certain theory would pragmatically proceed in its investigations. Although both groups of theories would not deny the relevance of empirical research, the mechanistic models would take on a much more ”bottom-up” approach, withholding the drawing of any firm conclusions without substantial accumulation of incoming data [Michel et al., 2019]. Unification accounts, on the other hand, would be much more prone towards a ”top-down” approach, giving priority to those empirical predictions which can illuminate how to make sense of seemingly disparate phenomena from the perspective of the framework itself.

2.4.2 Mechanism of explanation

The second dimension along which one could classify a model is its proposed mechanism of explanation, i.e. what is *responsible* for the phenomenon according to a theory [Illari and Williamson, 2012]. Specifically, there is an appreciable difference in which functional and causal explanations provide explanatory power, by either concentrating on specific goals of phenomena, or by describing network parts and interactions between them. Pertinent examples for functional theories are HOT and GNW, according to which, consciousness can arise in a physical system as long as it realizes ”meta-representation” or ”global broadcasting”, respectively. On the other hand, models such as IIT and RPT lean more directly towards elucidating causal interactions at the level of network analysis. For a full list of relevant quotes disclosing either functional or causal inclinations in explaining consciousness, see Table A.2.

Although we believe that most theories do explicitly or implicitly differ in these explanatory assumptions, Fahrenfort and van Gaal point out that *most* empirical theories would eventually aim at explanations involving causal implementations, rather than functions [Fahrenfort and van Gaal, 2020]. In line with that view, we have indeed come across many examples in which models of consciousness, even those labeled as functional, in actuality resort frequently to the language of *causal interactions*. For example, as stated by Rosenthal: ”On the HOT hypothesis, a conscious state is a compound state, consisting of the state one is conscious of together with a HOT. So the causal role a conscious state plays is actually the interaction of two causal roles: that played by the state itself and that played by the HOT” [Rosenthal, 2008]. Another relevant instance could be predictive processing theory (PP). Despite having been classified as functional in [Doerig et al., 2019], PP clearly states that implementing the adequate computational/functional principles is only a necessary, but not sufficient condition for consciousness. As explicitly mentioned by Wiese and Friston in [Wiese and Friston, 2020], computations need

to be physically instantiated in the right architecture and not all virtual simulations that simply realize the appropriate Bayesian inference should be considered conscious. The reason they state is that a Markov blanket of the physical system must be based on the system's dynamics and the dynamics strictly depends on the system's structure. That drives us to conclude that even one of the most popular frameworks has recently moved significantly away from functionalism and can be firmly classified as a causal theory. One might speculate that while some functional theories are still valuable in informing about the functions of consciousness, they will not be equipped to contribute to true explanatory legwork bridging physics and subjective experience.

Additionally, some models seem difficult to classify as either functional or causal in the above sense. For example TTC aims to accommodate both functional and causal types of explanations, making it difficult to identify its commitments (although according to some key quotes it seems to lean towards causal approaches). Other models such as ERE, CAN and BE are neither functional nor causal. This is because they do not share the underlying philosophical assumptions of emergence and physicalism. ERE assumes a form of dynamical co-emergence, i.e. consciousness co-arises with the system in such a way that there is a contextual constraint between the biological system (living body) and the experience (lived body), making both interdependent [Thompson, 2007]. In other words, according to ERE consciousness does not only emerge from the brain and the body, but it also actively constrains them creating a causally reciprocal relationship between neural events and experience [Thompson and Varela, 2001]. Lastly, other problematic cases include CAN (consciousness in this case is fundamental and the physical realm is what emerges from it) and BE (which implies a dualist account, thus eschewing emergence).

2.4.3 Target of explanation

The final dimension corresponds to the target of explanation, i.e. quality vs quantity of consciousness. We analyzed relevant papers looking for statements revealing *what* the theory aims to explain. Theories that focus mainly on the contrast between wakefulness and other impaired conditions (TCL) or on what neural activity might underlie the transition from the stimulus being "unseen" to "seen" (GNW, HOT, CK) can be classified as theories of quantity. We encourage the reader to notice that the "why" question posed in case of such theories is not "why a stimulus feels this way", but rather "what constitutes the switch rendering content visible". The

stimulus here can be replaced with any other stimulus, as the quality (though not necessarily the structure) of content remains irrelevant and out of the explanatory target. Importantly, answering the first question leads us to phenomenal content, while the second question corresponds to content access.

An attempt to explain the quality rather than sole quantity is exemplified by IIT. IIT tries to identify the mechanisms behind the phenomenal character of experience, i.e. the quality of its content (*what makes an experience visual, auditory, colourful, painful?*). In that vein, it poses that the quality of consciousness is in one-to-one correspondence with the geometry, concepts and relations encapsulated by the Maximally Irreducible Conceptual Structures (MICS) [Haun and Tononi, 2019]. IIT predicts that the contents of consciousness are entirely specified by the internal workings of elementary mechanisms of the main complex . Notably though, IIT's agenda to target the problem of quality is still in its nascent stages and has so far not gone beyond trying to explain the spatiality of experience [Haun and Tononi, 2019].

Still, IIT might pave the way for other theories' proponents to appreciate different kinds of explanatory aims and recognize that a mature theory of consciousness should address both the quantity and quality of experience. Novel questions inspired by the focus on quality for other theories could be, *inter alia*: how exactly does the *pattern* of global ignition inform phenomenal content? What distinguishes visual or auditory reentrant processing, or in other words, what mechanism might account for the fact that these two modalities *feel* different? (Notice that deflecting the problem by saying that phenomenal content is specified simply by the *primary* sensory cortices is not tenable simply because it does not inform *what precisely makes* a primary cortex *visual* or *auditory* or *somatosensory* in the first place.)

2.4.4 Empirical consequences

The fact that different theories subscribe to different modes, mechanisms, and targets of explanation has several empirical consequences. One of them pertains to the localization of neural correlates of consciousness (NCCs) [Boly et al., 2017, Klink et al., 2015]. First off, there is a substantial disagreement between the so called first order (e.g. IIT, RPT) and higher order theories (e.g. HOT) insofar that the former typically assign NCCs to the early sensory cortices while the latter to the frontal-parietal network. It is worth pointing out that the "frontal theories" tend to systematically cluster under the mechanistic mode of explanation, often interested

in questions of quantity rather than quality (i.e under what conditions a particular stimulus would be classified as consciously perceived or not) (see Figure 2.2). Notwithstanding these distinctions, most of the models reviewed above are dynamical global network approaches and therefore non-localist in principle [Dehaene and Naccache, 2001]. The network is relevant, the nodes alone are not. For example, GNW is often misunderstood as a fixed architecture encompassing the fronto-parietal cortices. In actuality, it comprises dynamic neural contributions that define this hypothetical global network [Dehaene and Changeux, 2005]. The only anatomical constraint is that relevant regions should be connected by long axons of pyramidal neurons. During decades of research, GNW identified pyramidal neurons in layer II and III as candidate mechanisms. In light of recent relevant evidence [Suzuki and Larkum, 2020], they have put forward layer V as a more likely GW substrate [Mashour et al., 2020]. On the contrary, IIT, guided by both theoretical and empirical clues, postulates that the true NCC lies in the posterior hot zone; one of the main points of divergence with GNW being the counterintuitive role of inactive units, contributing the cause-effect information just as the active ones [Oizumi et al., 2014, Siclari et al., 2017]. Theories such as HOT do not even specify network mechanisms in sufficiently rigorous terms, leaving a lot of space to accommodate ambiguous empirical data. Other inconclusive findings include the study of posterior hot zone involvement in dreams [Siclari et al., 2017, Mashour et al., 2020] or PCI index [Casarotto et al., 2016]. In the first case the reduction of low-frequency activity in posterior zones of the brain correlates with dreams during rapid-eye-movement (REM) sleep and non-REM sleep [Siclari et al., 2017]. At the same time, content specific dreams involve high-frequency activity in the frontal and prefrontal cortex. Some authors interpret these results as evidence of posterior zones for phenomenal consciousness and support of IIT, while others view them as clear evidence of the role of prefrontal cortex and GWN [Mashour et al., 2020]. In the second case, the PCI index inspired by the IIT framework seems compatible with the global ignitory activity of GNW [Mashour et al., 2020], as well as with other models such as ERE. Another recent study inspired by GNW found that the thalamic nucleus of monkey brains under deep brain stimulation (DBS) restores signatures of consciousness and reactivate nodes of the GNW that remain inactive under anesthesia [Tasserie, 2020]. However, as the authors also pointed out, these results are also compatible with thalamocortical loops theories and IIT.

Therefore, the postulation of either prefrontal regions and posterior regions to be mainly responsible for consciousness does not by itself speak to different explanatory approaches and probably merely reflects the availability of experimental techniques. However, the question of

localization makes sense only if one endorses a causal-mechanistic explanatory framework. If one, by contrast, looks for unification, then the fact that, say, posterior regions are most relevant for the NCC, is contingent. This is particularly true for those views which do not adhere to the explanatory primacy of causal chains within space-time (e.g. CAN). As such, experiments that try to identify the "locus" of the NCC do not provide evidence for or against such a theory. But even within the causal-mechanist framework, localization approaches are problematic: evidence for some mechanisms at any particular level does not falsify the relevance of other mechanisms at other scales. An interesting example is a controversial recent comparison between IIT and GNW at the single-unit level. At first glance, results suggest that GNW is supported by the evidence, while IIT is not [Noel et al., 2019]. However, considering the active single-unit level as the optimal spatiotemporal scale for testing IIT remains problematic. This would force a convenient assumption of GNW (i.e. that consciousness is correlated to only active neurons) onto IIT, which is *prima facie* not warranted. But this is a conceptual issue having to do with IIT's explanatory commitments – and not an empirical one.

Another empirical consequence of the way a theory emphasizes either quality or quantity has implications for the methodology used. In general, it has been long recognized that the scientific study of consciousness utilizes first-person and third-person approaches [Olivares et al., 2015]. The former includes subjective reports and phenomenological interviews [Chalmers, 2013a], the latter refers to objective measures of physical states, using different techniques such as Electroencephalography (EEG), functional magnetic resonance imaging (fMRI), Magnetoencephalography (MEG), among others. Most models of consciousness claim to employ both subjective and objective accounts. However, their assigned importance varies across the models. In particular those theories which seek to explain quality will wish to utilize a method that is specifically suited to make the qualitative aspect of experience precise. Particularly promising accounts are called "second-person" methods, referring to interview techniques that incorporate verbal and non-verbal reports in order to obtain a well informed subjective report [Olivares et al., 2015, Petitmengin et al., 2019]. This approach is motivated by earlier research in neurophenomenology [Thompson, 2004, Varela, 1996, Lutz et al., 2002]. The second-person method is different from the first-person method in that the former is guided by an interviewer who reads and interprets various indicators from the first-person subjective report. Given these indicators, the interviewer is able to ask more refined questions that force a subject to closely specify her reports.

Related to the quality versus quantity distinction is the one between levels and contents of consciousness [Bachmann and Hudetz, 2014, Bayne et al., 2016, Storm et al., 2017]. Levels of consciousness convey global signatures of consciousness, from which different paradigms contrast awake neural activity against non-awake or disrupted conditions such as sleep, chronic disorders of consciousness, anaesthesia, among others [Signorelli et al., 2021a]. By contrast, paradigms looking for contents of consciousness survey conscious experiences through contrasting perceptual analysis (perceived vs unperceived) and multiple psychophysical reporting paradigms. Examples include masked stimuli, high-contrast figures, binocular rivalry, flash suppression, motion-induced blindness, attentional paradigms, among others [Klink et al., 2015] (Klink et al., 2015). However one must not conflate the study of contents with the study of quality that focuses on the specific phenomenology of such contents. It is more appropriate in this case to distinguish between an ”access content” and a ”phenomenological content”.

2.5 Conclusions: integrative methods for the future

In light of our previous sections, the empirical testing of models of consciousness is far from trivial. Comparing models of consciousness is not only difficult due to experimental limitations, but also due to the fact that different models operate on very different and mostly implicit assumptions about modes, mechanisms, and targets of explanation. We thus introduced a classification scheme to make the different explanatory profiles of leading models explicit. To our knowledge, this is the first time that all these models and perspectives are reviewed systematically, focusing on different explanatory aspects, analytical dimensions, and organized in one single and comprehensible classification.

Having perused the relevant literature around 13 popular frameworks, we suggest that a number of points of contentions in the field of consciousness studies might stem from differently set explanatory goals and targets. Although both mechanistic and unificationist accounts have their advantages, there are certain areas of research where one of them might turn out to be more suitable than the other. Some types of explanations might require to step back from investigating the specific empirical details and look more at the overall mathematical structure: the reason ”why” a person cannot untie a particular knot may stem from a topological fact about the knot, rather than from a detailed causal trajectory illustrating the attempt of its disentanglement. Numerous examples show that a full-fledged explanation of the phenomenon might

sometimes require something more than a causal story [Reutlinger, 2017]. On the other hand, unificationist accounts might be facing its own problems, such as the problem of asymmetric structure of explanation[1] [Barnes, 1992]. It is yet to be established which type of explanation is most adequate for the science of consciousness or whether even different modes of explanation would be required (e.g. manipulationist accounts [Woodward, 2004, Woodward, 2019]).

Nonetheless, one of the messages of this chapter is to emphasize that the first step towards a more mature science of consciousness is the recognition that the question of "why" the brain generates subjective experience can be understood in a variety of ways. We suggest that to fully explain *what it means to be conscious*, one needs to first be precise about what it means to *explain* something.

Remaining aware of the disagreements within the field (and its philosophical pre-history), one might also try to extract what most models agree on. Some empirical approaches intend to follow this kind of pragmatism. For example, the concept of criticality in dynamical system theory shows to be compatible with evidence for both access and phenomenal consciousness [Tagliazucchi, 2017]. The analytical method of connectome harmonics also aims to unify different signatures of consciousness, from a more general perspective of brain functioning and physical system theory [Atasoy et al., 2017, Atasoy et al., 2019, Luppi et al., 2020]. Large scale models using different anatomical, functional and molecular layers of description also present promising features to integrate different mechanisms at different scales [Kringelbach et al., 2020], as well as signatures of consciousness [Signorelli et al., 2021a]. Recently, optogenetic experiments demonstrated that the biophysics of pyramidal neurons in cortical layer V integrates two contentious mechanisms associated with consciousness, i.e. cortico-cortical loops and higher-order thalamocortical loops [Suzuki and Larkum, 2020, Aru et al., 2020]. More research aiming at synthesizing different findings is currently underway.

Recent mathematical works too have recognized the need for integration within a sound theoretical foundation. This new trend, "the mathematics of consciousness", employs formal and rigorous methods to explore ways to distinguish various models and derive new empirical predictions. Some examples are the mathematical developments, based on IIT [Oizumi et al., 2016, Tsuchiya et al., 2016, Kleiner and Tull, 2020] or mathematized phenomenology

[1]The problem of asymmetric structure of explanation can be summarized as follows: if A explains B, B does not explain A. In the mechanistic approach this asymmetry is easily explained away, as inherited from the intrinsic asymmetry in the direction of causation itself. It might seem much more problematic from the stance of unification, as clearly lacking any commitment to the realism of causation [Barnes, 1992].

[Yoshimi, 2007, Ehresmann and Gomez-Ramirez, 2015, Prentner, 2019]; other approaches are based on symmetry [Kleiner, 2020], category theory [Northoff et al., 2019] or on the compositionality of processes [Signorelli and Meling, 2021, Signorelli et al., 2020b]. Some models explicitly address different metaphysical starting points, such as idealism [Signorelli et al., 2021d] or decompositional approaches of dual-aspect monism [Atmanspacher, 2020]. Common to all these approaches is that, inspired by transparency of mathematics, they explicitly define their core assumptions. In the end, whether they are of any value will be determined by how much explanatory power they bring into the constantly accruing experimental evidence.

The science of consciousness needs integrative frameworks and integrative frameworks by definition are multidisciplinary. In the future, and going beyond empirical methods and mathematics, a dialogue with artists, meditators and sociologists may help us to think out of the box, and rediscover some aspects of conscious experience that have been largely unattended to.

Chapter 3

A MULTILAYER APPROACH

Scientific studies of consciousness require both experimental and theoretical models. In this chapter, we discuss a variety of empirical results inspiring a multilayer conceptual model. Evidence points out to the multilevel brain-body organization and diverse brain regions involved in different aspects of consciousness. Our framework parsimoniously integrates those findings and generates new experimental and theoretical hypotheses. This theoretical exercise conveys a reconciliation framework to target both the brain-body function and its relationship with consciousness.

3.1 Introduction

We propose an integrative meta-framework for consciousness. The framework consists of a multilayer network [Kivela et al., 2014], where layers represent self-sustained systems organized by their intrinsic dynamics. In this framework, *intra* and *inter* interactions drive dynamic scenarios, coupling and decoupling the whole brain-body system. Intra interactions are the internal network interactions across nodes, while inter interaction corresponds to in-between networks exchanges. Interactions are any relevant biological exchange for these biological networks and commonly represented by functional, metabolic or anatomical connections. Here, we introduce and discuss how these conceptual layers and their intra and inter interactions integrate the variety of signatures of consciousness, requirements for brain functioning and consciousness as the summit of brain evolution.

In the following sections, relevant evidence is discussed in light of the complex brain organization and signatures of consciousness. On the one hand, there are multilevel and multiscale

interactions ranging from molecular neurotransmitters to electrical impulses inside neurons and large scale activity through multiple brain assemblies. Different cell types, such as glia and astrocytes, molecular diffusion, among other biological interactions add complex dimensions to the brain organization. On the other hand, there are a variety of brain signatures associated with consciousness, multiple regions involved, and rival theories pointing out to some relevant neural networks instead of others.

Two groups of evidence inform us about signatures of consciousness [Bachmann and Hudetz, 2014, Storm et al., 2017]. One group of evidence convey global signatures of consciousness, i.e. brain activity related to global states such as sleep, awake, or vegetative state. In this line, the most robust global signature is the slow-wave activity ($\leq$4 Hz) observed in the electroencephalogram (EEG) signals during lack of consciousness. Another group of evidence surveys conscious experience through contrasting analysis (perceived vs unperceived) and multiple reporting paradigms. Examples include masked stimuli, high-contrast figures, binocular rivalry, flash suppression, motion-induced blindness, attentional paradigms, among others [Klink et al., 2015]. On top of that, non-reporting paradigms appear in order to avoid possible confounding factors [Aru et al., 2012, Tsuchiya et al., 2015]. In summary, the first group of experiments is thought to inform about *global modes of consciousness*, while the second about *content access*. Additionally, recent paradigms also include neural correlates of dreams, psychedelics and meditation.

Unfortunately, most of these experimental paradigms (see [Sandberg et al., 2010] for exceptions) tend to describe consciousness in one continuous dimension of interest [Bayne et al., 2016, Bayne and Carter, 2018]. Consciousness is reduced to one quantifiable dimension, instead of a richer multi-dimensional structure.

In order to deal with this issue and integrate all the evidence collected, the neuroscience of consciousness requires dynamical integrative frameworks. One crucial insight is that consciousness involves temporal and dynamical brain organization. Brain states associated with consciousness are characterized by a richer repertoire of functional configurations [Barttfeld et al., 2015, Demertzi et al., 2019], while disruptions of this dynamic exploration seem to imply loss of consciousness. Transient lapses of awareness are also present during healthy wakefulness and they might be addressed by empirical and theoretical models [Demertzi et al., 2019, Naccache, 2018]. Therefore, dynamical scenarios, such as criticality, become important ingredients to reconcile different aspects of consciousness and brain function [Tagliazucchi,

2017]. However, dynamical scenarios are descriptions of the dynamical evolution of a system, but they do not explain how such a system reaches those regimes. To answer the mechanistic question we need to focus on the causal forces driving the dynamics of the system [Werner, 2013]. In next sections, we construct a conceptual theory and discuss how biological layers interacting may become those causal forces, reconciling different types of evidence from consciousness research. This is important in order to integrate multiple dimensions of interaction, better describe brain-body systems, the embodiment of consciousness, and study, both conceptually and mathematically, the system coupling via intra and inter interactions observed as signatures of consciousness. Eventually, this framework may better serve to reason about the neuroscience of consciousness, its embodiment and empirical paradigms, becoming a more appropriate framework to explore the intertwined *new physics* emerging from layers interacting.

3.2 The dynamic Brain

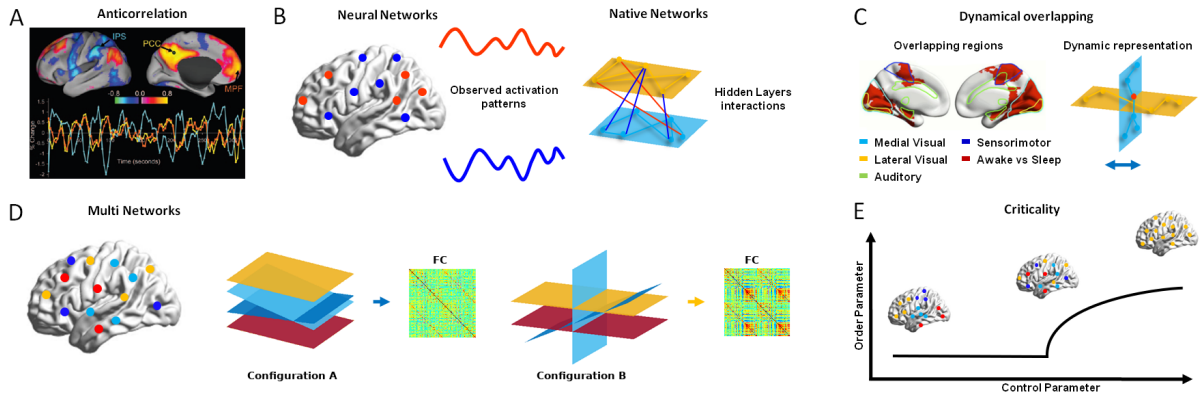


Figure 3.1: The dynamic brain. (**A**) Anticorrelated networks, adapted from [Fox et al., 2005]. (**B**) Each node represents a region of interest and each colour one family of observed anticorrelated networks. Colour lines are their hypothetical activities. We suggest these observed anticorrelated networks are the result of more basic or native networks (yellow and light blue layers). These layers are virtual forms to visualize physical brain structures. (**C**) Example of overlapping regions. The contrast between awake and sleep conditions (right). Different layers of a multilayer network may dynamically share nodes. Here a dynamical overlapping representation in the form of rotation layers interacting (left). Orange dot corresponds to one overlapping region. Adapted from [Tagliazucchi et al., 2016b] and [Signorelli and Meling, 2021]. (**D**) Different hypothetical native networks and their configurations convey different functional observed patterns of connectivity. (**E**) Criticality is the transition point where a system presents fluctuations between order and disorder patterns. In our conceptual model, regions previous a critical phase transition (disorder) represent uncoupled native layers (given by coloured dots). Critically then emerges through the couplings of these native layers generating the functional patterns observed in resting-state. Finally, regions that continue beyond criticality (order) correspond to extremely coupled layers, such that the intrinsic layer dynamics are lost in favour of a homogeneous global dynamics. Here, the control parameter (i.e. the strength of interactions) mirrors the global interactions between layers (interintercations), while the order parameter (i.e. the degree of order) represents the degree of disruption of internal layer interactions (intrainteractions). Reinterpreted and adapted from [Cocchi et al., 2017].

3.2.1 Resting states networks

Brain networks are the informative unit at the whole-brain scale [Pessoa, 2014]. Brain networks are groups of non-overlapping brain regions according to organized patterns interacting by function or structure. One example is the resting-state networks (RSNs), discovered by Biswal and colleagues [Biswal et al., 1995]. RSNs appear from the spontaneous brain activity at rest, i.e. without performing any particular task. The default mode network (DMN) is one concrete example. DMN is an anatomical well-defined network, associated with resting states and introspective tasks [Buckner et al., 2008]. Another example is the dorsal attention network (DAN), mostly activated by tasks requiring spatial attention.

In resting state, the brain seems organized into dynamical anti-correlated functional networks [Buckner et al., 2008, Fox et al., 2005, Fransson, 2006]. In other words, the high correlation between areas of one network corresponds to the low correlation between areas of the other network (Figure 3.1A). For instance, the relationship between DMN and DAN is anticorrelated. These results extend to task-related activity, such as the suggested anticorrelation for cingulated-operculum and frontal-parietal network in goal-directed executive control [Dosenbach et al., 2008].

To explain anticorrelated networks, we propose they are the product of more basic networks interacting (Figure 3.1B). Anticorrelations between networks suggest competing systems [Deco and Corbetta, 2011]. Nevertheless, this competition is not a direct antagonism between nets but triggered by inner spatiotemporal structure [Deco et al., 2011]. This brain spatiotemporal structure partially relays on anatomical structure [Vincent et al., 2007], however, spontaneous functional activity is not fully explained by anatomy [Deco et al., 2009, Deco et al., 2013, Ghosh et al., 2008, Honey et al., 2009]. RSNs are not explained by simple random noise either [Deco and Corbetta, 2011], despite noise-driven-transitions playing an important role in underlying mechanisms of spontaneous activity [Deco et al., 2009, Ghosh et al., 2008]. Therefore, one alternative is disentangling RSNs by invoking more basic independent or semi-independent networks that may or may not interact according to each context. This *native networks* would define the co-dependent networks later observed as brain networks and their dynamical organization.

The resulting co-dependent networks become multi-faced dynamical networks with over-

lapping regions [Mesulam, 1990, Pessoa, 2014] (Figure 3.1C). To describe this overlapping, we allow virtual rotation networks (for visual purpose), such that different regions dynamically overlap. It allows the same and different areas participate in the same and different behaviours (many to many instead of one to one relation). These emergent networks are contextual, namely, the region affiliation varies according to the time and task [Cole et al., 2013, McIntosh, 2000]. As such, one brain region is part of a different process according to its network affiliation at a given time. Overlapping networks in space and time may better account for cognitive processes and its relationships with brain function [Pessoa, 2014].

3.2.2 At the edge of criticality

Structure-function relationships are not static but dynamic [Kiverstein and Miller, 2015, Varela et al., 2001, Allen et al., 2014]. The observed functional patterns may reflect the dynamical organization of different native overlapping networks in a more general multi-network structure (Figure 5.3D). This evolving dynamic structure may generate different dynamic scenarios.

One relevant scenario is criticality (Figure 5.3E). For dynamical systems, criticality corresponds to transition points, or bifurcations between stable equilibrium and multistable states with coexisting multiple attractors [Deco and Jirsa, 2012, Cocchi et al., 2017]. These dynamical systems at criticality serve to model large scale brain activity. Large-scale dynamical models use different neural signals and combinations of structural (SC) and functional connectivity (FC) to search for links between spatial (anatomy) and temporal (functional) brain dynamics [Deco et al., 2013, Cabral et al., 2017, Breakspear, 2017]. These models teach us that spontaneous brain activity presents deviations from equilibrium trajectories, resulting in rapid switching between a discrete number of functional states [Hansen et al., 2015, Lynn et al., 2020]. Functional activity evolves through multiple and recurrent discrete functional states [Allen et al., 2014, Cabral et al., 2017], lasting around 100-200 ms [Vidaurre et al., 2016, Deco et al., 2019]. These deviations from equilibrium suggest that brains maximize their possible microstates at the edge of criticality [Cocchi et al., 2017]. In fact, RSNs are described by critical points [Deco et al., 2009, Ghosh et al., 2008] and different models find their optimal fitting points at the edge of the critical instability [Cabral et al., 2017].

Criticality seems to underlay a fundamental principle of brain organization. At criticality, small extrinsic perturbations can trigger the dynamic of task-related networks, while intrinsic

perturbations may generate the exploration of functional resting states [Deco et al., 2013]. Criticality also seems supported by scale-free dynamics. Therefore, the brain at criticality is maximally sensitive to internal and external fluctuations.

3.2.3 Criticality and Consciousness

Criticality may also unify different evidence regarding consciousness [Werner, 2013, Tagliazucchi, 2017]. For instance, departure of criticality is reported in conditions of general anesthesia [Scott et al., 2014, Tagliazucchi et al., 2016a], deep sleep [Priesemann et al., 2013, Tagliazucchi et al., 2013] and epileptic seizures [Meisel et al., 2012]. In deep sleep, the dynamic of the whole brain presents an increase of stability and decrease of effective interactions [Jobst et al., 2017]. Contrary, perturbations in computational simulations showed that larger recoveries of integration properties at the whole-brain scale are associated with shifts in the model operation point. This result indicates that in awake condition, brains operate far from a stable equilibrium [Deco et al., 2018a].

Criticality also helps to reinterpret current evidence. For instance, in awake condition, complex patterns appear after transcranial magnetic stimulation (TMS), but they disappear in non-conscious conditions [Casali et al., 2013]. One dynamical reinterpretation of these results convey criticality: in non-conscious conditions, the brain moves away from the edge of criticality and therefore becomes only locally responsive. The disruption of criticality might be due to the switch of different systems from coupled to uncoupled intrinsic dynamical modes that disconnect functional correlations (Figure 3.1E). Therefore, the question about ”mechanisms” of consciousness might turn into the question about driven forces of criticality, from which network integration becomes a consequence of this critical dynamic scenario [Werner, 2013].

3.3 The Brain-Body coupling

If criticality plays a role in functional brain organization, what makes the brain work at these critical points? [Cocchi et al., 2017]. One would expect intrinsic neural activity involved, however, more complex body physiological signals also participate [Breakspear, 2017], e.g. diffuse neurotrasmitter realise, among others. In [Laumann et al., 2017], the authors show

how removing physiological confounds have an enormous impact on neural signals, decreasing non-stationarities up to the level of almost disappearing. Physiological fluctuations have subtly but important cognitive effects [Allen et al., 2016]. This forces us to distinguish between the confounding role (e.g. BOLD signal) of first-order physiological signal (e.g. heart rate, respiration), and the second-order effects (e.g. heart rate variability), which co-vary with regions such as insula in anticipation tasks [Nguyen et al., 2016].

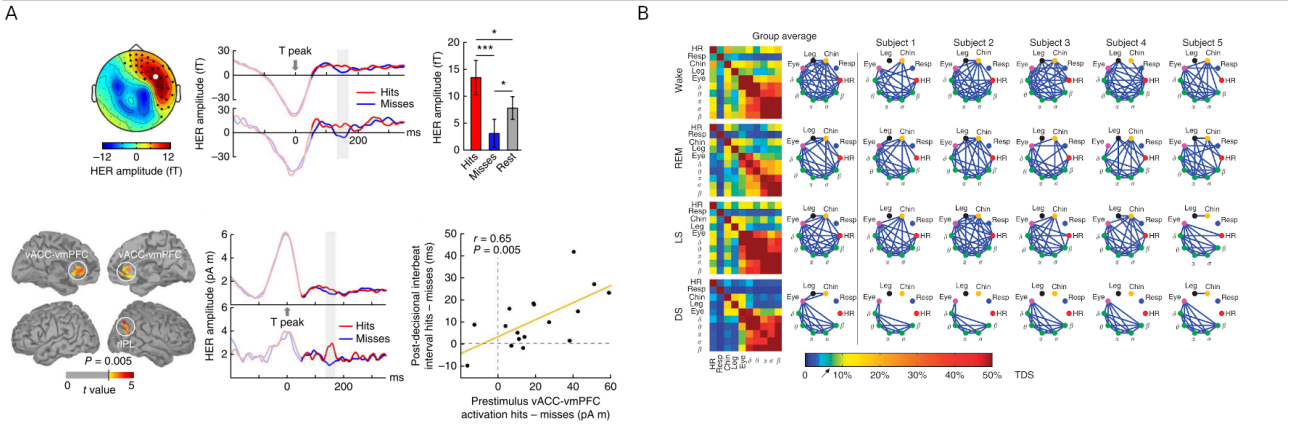


Figure 3.2: Brain-body couplings. (**A**) Conscious perception is predicted by neural events locked to heartbeats. Upper, the average of heartbeat-evoked response (HER) is shown for all black dots, followed by the HER in a single node (white dot). The amplitude of HER (at 135-171 ms) changes for observed versus missed stimuli, also compared under open eye rest. Bottom, differential activation for bilateral vACC-vmPFC and rIPL regions, followed by their HER curves across hits and misses before stimulus onset. Correlation across subjects between the hit-miss difference in cardiac interbeat and the hit-miss difference in HER before stimulus onset in vACC-vmPFC. Figures from [Park et al., 2014]. Shaded areas highlight the significant difference. (**B**) Brain-body couplings across different sleep stages: Wake, rapid eye movement (REM, light sleep (LS) and deep sleep (DS). Figures from [Bashan et al., 2012].

3.3.1 Physiological coupling

One example of body-brain interaction is the gut's influence in neurological conditions like anxiety, depression, and autism spectrum disorder [Critchley and Harrison, 2013, Mayer, 2011, Sharon et al., 2016]. The bi-directional signalling between the gut and the central nervous system include different paths, the vagus nerve, the enteric nervous system (ENS), sympathetic and parasympathetic branches of the autonomic nervous system (ANS), gut hormone signalling, the immune system, neuroendocrine signalling, tryptophan metabolism, and microbial metabolites such as short-chain fatty acids [Foster et al., 2017, Grenham et al., 2011]. These paths ensure the coordination of gastrointestinal functions to support behaviour, as well as feedback from the gut to influence motivated behaviour and high cognitive functions [Foster et al., 2017, Mayer et al., 2015, Rhee et al., 2009]. For instance, the vagus nerve targets the nucleus of the solitary tract in the caudal brainstem, which mediates polysynaptic inputs to higher brain regions, such as the hypothalamus, limbic forebrain, ventromedial basal nucleus of the thalamus [Saper,

2002] and the ventral anterior cingulate cortex [Vogt and Derbyshire, 2009]. While descending neural projections from cingulate and insular cortices, amygdala, bed nucleus of stria terminalis and hypothalamus provides a bi-directional control [O'Mahony et al., 2011]. Other examples of visceral interactions also include the heart and lungs.

Recent results strengthen and extend the link between brain-body interactions to conscious perception (Figure 3.2A). Neural spontaneous fluctuations locked to heartbeats seem to predict and shape visual detection, suggesting the influence of heartbeats in conscious perception [Park et al., 2014]. These findings suggest propagated activity from ventromedial prefrontal cortex and ventral anterior cingulate cortex (known to receive cardiac inputs) to the posterior right inferior parietal lobe [Park et al., 2014]. Interoceptive heartbeats activity sent to the insula also has a systemic effect on conscious perception, modulating exteroceptive awareness [Salomon et al., 2016] and emotional experiences [Nguyen et al., 2016]. Explicit cardiac perception also influences activity in regions such as the posterior and anterior insula, dorsal anterior cingulate, somatomotor cortices, among others, supporting interoceptive awareness [Critchley et al., 2004].

Studies of anaesthesia and sleep also indicate the importance of brain-body couplings and its connection with consciousness. In [Stankovski et al., 2016], the authors demonstrated alterations on the coupling functions of cortical and cardio-respiratory oscillations under sevoflurane and propofol. Sevoflurane affects the respiratory-theta coupling more than propofol, while heart-theta coupling presents complex forms suggesting influence from the heart to theta neural oscillations. Delta activity influences alpha oscillations [Stankovski et al., 2017], which are thought to play a key role in consciousness. These couplings are significantly stronger in anaesthesia than awake condition, suggesting a reduction of the dynamic brain repertoire [Stankovski et al., 2016].

Using other coupling definition, analyses in sleep show that sleep-stage transitions are correlated by changes in the topology of dynamical brain-body physiological networks [Bartsch et al., 2015, Bashan et al., 2012] (Figure 3.2B). The coupling of cardiac, respiratory, cerebral, ocular and muscle signals convey higher network interactions in awake and light sleep, intermediate values in REM and much lower in deep sleep [Bashan et al., 2012]. Moreover, brain couplings characterized by different oscillations exhibit a decoupling from the other physiological systems through different sleep-stages, at the same time that the strength on intra-connectivity links was stronger in light and deep sleep, intermediate in awake and low in REM.

All these data suggest that brain-body couplings have indeed an effect in different brain states associated with consciousness, as well as brain and more general physiological systems act sometimes connected and other times disconnected.

3.3.2 Cell types

Astrocytes and glial cells also play an important role. The hemodynamic functions of the heart partially determine the heart-brain coupling and the cardiac-theta coupling. This coupled system provides metabolic substances and oxygen through the blood flux. At the neural level, astrocytes and other glial cells are responsible for mediation of these processes [Haydon and Carmignoto, 2006, Zonta et al., 2003]. For instance, mice with knockout of mitochondrial astrocyte-specific proteins take longer times to recover from volatile anaesthetics than mice control [Ramadasan-Nair et al., 2019]. It suggests that astrocytes' mitochondrial function modulates the emergence or recovery from anaesthesia. Astrocytes also seem involved in sleep-wake cycles, mechanisms and functions of sleep [Haydon, 2017, Petit and Magistretti, 2016]. The interaction neuron-astrocytes is crucial to maintain neural energy consumption, making astrocytes the supplier of brain glucose [Bélanger et al., 2011, Jha and Morrison, 2018, Magistretti and Allaman, 2015].

In fact, glia and astrocytes cells, as producers of glucose, may influence the whole-brain energy states and its relation with consciousness, as discussed in the following sections.

3.3.3 Brain-Body and Consciousness

Unfortunately, consciousness studies have paid little attention to all these interactions. The neuroscience of consciousness focuses on the cascade of neural events, giving for granted those basic brain-body couplings.

Nevertheless, these findings are important because they add new complex dimensions on the understanding of anaesthesia [Perouansky et al., 2019], sleep mechanisms, and their relation with the conscious phenomenon. On the one hand, brain-body couplings add new types of interactions, from neuron-neuron, neuron-glia [Velazquez, 2020], to more complex molecular and kinetic physiological systems [Thompson and Varela, 2001, Merleau-Ponty, 2005]. On

the other hand, these interactions might be essential to support the basis of any conscious experience.

In [Park and Tallon-Baudry, 2014], Park and Tallon-Baudry suggest a neural subjective frame made of a group of basic biological mechanisms. This frame defines the subject as a biological entity and this biological entity as the building blocks of first-person and subjective experience [Park and Tallon-Baudry, 2014]. The intrinsic connection between body signals and higher brain areas would generate the seed of the "I" and subsequent subjective experiences through an interaction of neural responses to visceral inputs and stimuli responses. These primary brain-body interactions are not explicitly experienced, they form a necessary, but not sufficient, condition to create a conscious perceptual experience.

Taking the above evidence together, a more integrative approach demands a switch from a "neurocentric" perspective to a whole picture of the brain-body function and consciousness [Thompson and Varela, 2001, Merleau-Ponty, 2005, Signorelli and Meling, 2021]. This integration justifies the introduction of multilayers networks as a mathematical structure to describe brain and body coupled organization.

3.4 Multilayers and Brain organization

The discussion above led us to define a multilayer structure for brain organization.

Multilayer Networks

Multilayer network is an extension of the widely used graph theory. A graph is a tuple $G = (V, E)$, where V is the set of nodes and $E \subseteq V \times V$ is the set of edges that connect pairs of nodes. A graph is a network of interactions, also called *single-layer network*. Then, a multilayer network is a quadruplet $M = (V_M, E_M, V, \mathbf{L})$. $\mathbf{L} = \{L_k\}_{k=1}^{d}$ is a sequence of sets L_k of *elementary layers*, being d the total number of aspects, these aspects are given in our pictures by the colour of the edges and they represent the different types of interactions. $V_M \subseteq V \times L_1 \times ... \times L_d$ is the subset of all tuples containing mixed information about the node and the layer present in every case. Finally, $E_M \subseteq V_M \times V_M$ is the set of pairs of possible

nodes and elementary layers [Kivela et al., 2014].

If a node u is present on different layers $(l_1, ..., l_d)$, we can use the notation:

$$(u, \mathbf{l}) \equiv (u, l_1, ..., l_d)$$

The set of edges is partitioned into *intra-layer edges* as those belonging to sets

$$E_A = \{((u, \mathbf{l}), (v, \mathbf{h})) \in E_M | \mathbf{l} = \mathbf{h}\}$$

and *inter-layer edges* as those in $E_C = E_M - E_A$.

This simple extension allows us to incorporate more complex brain-body structure. For example, one group of elementary layers might convey types of neurotransmitters, while another represents cell types. Each of these elementary layers will have their own aspect given by the specifications of their elements. On top of that, there are intra-layers edges and inter-layer edges representing the internal interactions inside nodes in the same layer and between nodes of different layers, respectively.

This framework brings new physics in the form of new dimensions of analyses [Kivela et al., 2014]. For instance, the relationships between anatomical and functional brain connectivity. One might consider that anatomical connectivity between brain regions accounts for all the spatiotemporal complexity of the brain. This is the analysis of one-dimensional network given by only anatomical connectivity, i.e. one layer network. If we now turn to another dimension of interaction, the functional one, we realize that anatomy is not enough to explain the richness of the brain dynamic at resting state. However, the functional network by itself is also insufficient to explain its own richness. Only the dynamical link between both dimensions of analyses brings a more compelling explanation [Cabral et al., 2017], i.e. the analysis of interconnectivity between anatomy and function.

This structure conveys an integrative meta-framework to account for further empirical evidence on brain-body function and consciousness research. As such, one might wonder what kind of new effects we are able to explain adding new dimensions of interaction.

Multilayers and Consciousness

Evidence supporting the introduction of layers and multilayers in conscious research comes from the reinterpretation of two recent empirical findings. One of them is the role of pyramidal cortical layer V as a mediator mechanism between the influence of feedback connectivity in cortico-cortical loops and higher-order thalamocortical loops in mice [Suzuki and Larkum, 2020]. In that work, the authors demonstrated that three different types of anaesthetics decouple the apical signalling between dendrites and cell body in layer V (Figure 3.3A). In terms of our conceptual model, cortical layers and thalamic regions would form part of a more general multilayer, where intra and inter interactions are modelled differently.

The second finding is the optimal coupling of neurotransmitter molecular system and cellular whole-brain system [Kringelbach et al., 2020]. The brain is approached as a three-layer system, where layers are defined by types of interactions (Figure 3.3B). Specifically, functional connectivity, anatomical connectivity and molecular diffusion. Then, the resting-state activity under psilocybin is fitted by coupling the systems, while decoupling systems produce a breakdown in the fitting of the empirical data [Kringelbach et al., 2020]. This result, and particularly the addition of a third dynamical dimension (molecular density and diffusion), is crucial to understand how functional activity is modulated under a mostly unchangeable anatomical neural substrate. In this case, the neurotransmitter layer given by its intrinsic molecular dynamic brings a new physical dimension not considered before. In other words, this new layer and its intrinsic interactions account for the modulation and dynamical switch of whole-brain repertoire via inter interactions.

Under our conceptual framework, the above example implies to study different layer configurations on a 3-layer system: a) defining and modifying the internal dynamic of the system-layer, e.g. defining different layer systems and their dynamical evolution via differential equations [Deco et al., 2018b, Kringelbach et al., 2020], b) modulating the interconnectivity via coupling functions, e.g. exploring the coupling parameter (inter-edges) to fit the empirical data (Figure 3.3B, non-modulation and cut dynamic), and c) studying the effects of intra connectivity in the whole system, e.g. changing the anatomical connectivity map or the neuromodulatory map (Figure 3.3B, with different receptors maps). These are three possibilities already explored, but our framework is open to further implementations.

Importantly, the multilayer framework subsumes the two types of empirical layers discussed above (Figure 3.3C), namely layers defined by cell types such as in Suzuki et al. experiments, and layers given by types of interactions such as the functional, anatomical and molecular layer division in Kringelbach et al. analyses. In our general framework, layer organization represents the whole architecture of living systems, while layer interactions, intra and inter, the processing inside and between layers. These layers present and should be defined by their particular self-sustained activity, given by, for example, their intrinsic oscillatory activities, or by metabolic exchanges.

Furthermore, we can hypothesize brain-body systems represented by layer networks considered as fully independent systems, i.e. under isolation, they behave exactly how they behave under minimal consciousness interactions (e.g. deep sleep). Therefore, a unique empirical layer division related to consciousness, and the concrete regions in one layer, shall be defined under ideal conditions of minimal coupling. Then, we hypothesize, this variety of layers and couplings would trigger the complex configurations and global brain dynamics observed in awake conditions, becoming the native networks mentioned in the previous section. Some of these layers are molecular systems connecting the otherwise anatomically disconnected regions. Different types of layers may couple and decouple from time to time and correlate with a reduction of awareness even under global awake conditions. For example, in our conceptual model, this reduction is due to the recovery of intrinsic layers dynamic (i.e. decouple of layers), in detriment of inter interactions that would interfere with those intrinsic layer dynamics. Interestingly, this reduction of awareness seems to be present between two moments of consciousness [Ward and Wegner, 2013, Naccache, 2018], as well as dynamical brain states associated with unconsciousness, are present even during conscious resting-states [Demertzi et al., 2019]. Our conceptual model explains this reduction by the natural need to go back to the intrinsic dynamics of decoupled layers. It makes multiple interlayer interactions fundamental forces supporting consciousness.

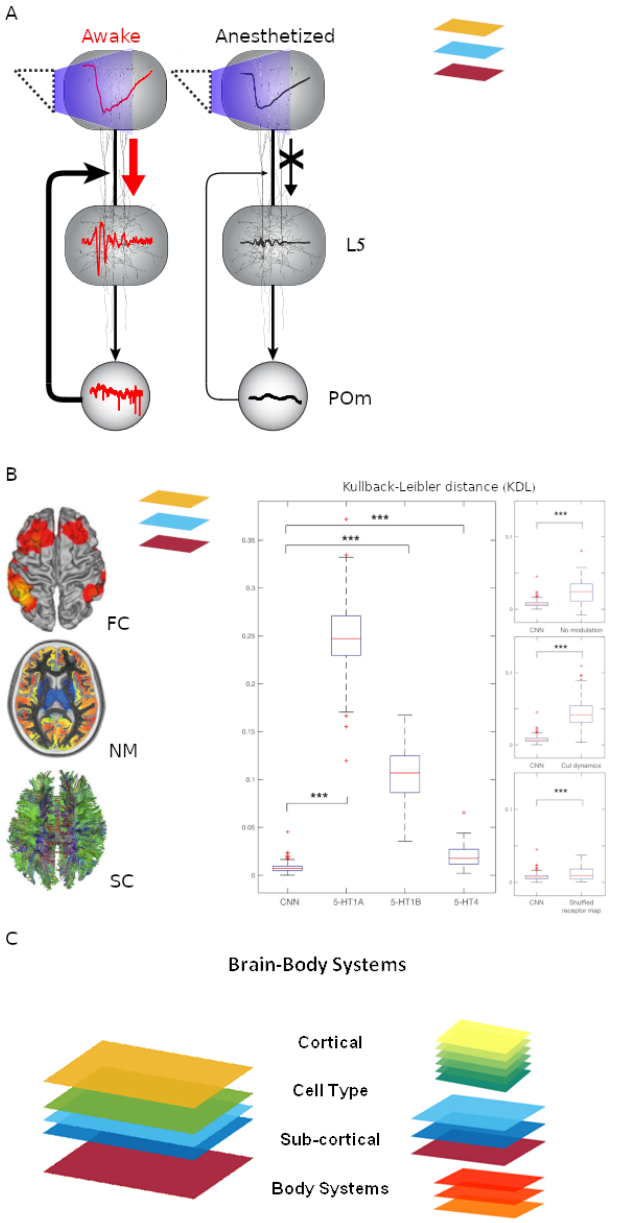


Figure 3.3: Brain-body systems in a multilayer network. (**A**) Neural activity in the posteromedial nucleus (POm) correlates with coupling activity of layer V pyramidal neurons during awake conditions, independently of optogenetic stimulation of distal apical dendrites. In turns, inactivation of POm impaired the coupling. Adapted from [Suzuki and Larkum, 2020]. (**B**) Functional (FC), neuromodulatory (NM) and structural connectivity (SC) may be represented by three interacting layers. The optimal dynamical coupled neuronal-neurotransmission (CNN) is measured by Kullback-Leibler distance (KLD). The coupling neural and neuromodulatory system of $5 - HT_{2A}$ fits the observed brain functional activity much better than other receptors maps, uncoupled neuromodulatory system, cutting the dynamic feedback dynamic between systems, as well as the result using shuffled receptors maps. Adapted from [Kringelbach et al., 2020]. (**C**) Hypothetical native networks may correspond to brain-body organizations, ranging from different neural layers, such cortical or subcortical networks, to different cells types layers, molecular layers and body systems in general.

Algebraic Multilayer Networks

If now one restricts our discussion to only one set of elementary layers, where aspects match a set of colours of dimension d and a time index t is external to the multilayer, we can give the next formal definition of a multilayer model based on the work of categorical networks models

[Baez et al., 2018]. For an extended discussion, see [Signorelli and Joaquin Diaz Boils, 2020].

First, consider a functor that produces many-coloured networks out of non-directed multigraphs. A multigraph G on a set $V(G)$ of *vertices* is a multisubset $E(G)$ of pairs of elements of $V(G)$, called *edges*, with a function $m : E(G) \to \mathbb{N}$ that calculates the multiplicity of every edge. Then, an *edge-colored multigraph* is a multigraph G together with a (non-necessarily surjective) function $col : E(G) \to \mathcal{P}(\mathcal{C})$ where $\mathcal{P}(\mathcal{C})$ denotes the set of subsets of the colour set $\mathcal{C}$. A graph is said to be *monochromatic* if all its edges are of the same colour. Secondly, we denote by $\mathcal{S}$ the *permutation groupoid*, that is, a skeleton of the groupoid of finite sets and bijections. $\mathcal{S}$ is a (strict) symmetric monoidal category.

Definition 1 *A one-coloured network model is a lax symmetric monoidal functor $F : \mathcal{S} \to Mon$ where Mon is the category of monoids.*

Now, let $MG(n)$ be the set of multigraphs on $\mathbf{n} = \{1, ..., n\}$. Then, we define a network model as $MG : \mathcal{S} \to Mon$ with values $(MG(n), +)$ where $+$ is multiset sum, that is, addition of multiplicities of edges sharing the same vertices. $(MG(n), +)$ is denoted by $MG(n)$ and called layers. The edges into every layer has a single c_i from a set of colours $\mathcal{C}$.

Finally, we consider tuples of layers by constructing a colour-indexed tensor of the monoids $MG(n)$. Elements of these tensors are called *multilayers*. For that, we use the category of network models over a fixed colour set $\mathcal{C}$, see [Baez et al., 2018]. That category is symmetric monoidal and therefore we can tensor functors such as those introduced above.

Definition 2 *A network model for multigraphs with coloured edges is a functor*

$$MG^{\otimes \mathcal{C}} : \mathcal{S} \to Mon$$

where $MG^{\otimes \mathcal{C}}(n)$ is a product of $|\mathcal{C}|$ copies of the monoid $MG(n)$.

At the initial time, every layer into a multilayer is identified both with a colour and a position in the tensor product. They also contain edges of a single colour each.

Composition of Layers

To account for the complexity of brain networks and their interactions, we introduce one formal way to compound the colours of different layers into multilayers. This is given by the operation of $\odot$. After using the operator $\odot$, layers become many-coloured.

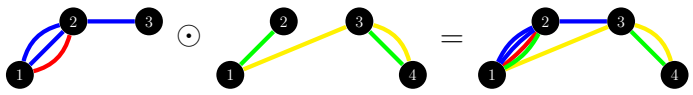
The following definition is a recursive operation:

Definition 3 *Let $\mathcal{C}$ be a fixed set of colours and $\mathcal{C}_1, \mathcal{C}_2 \subseteq \mathcal{C}$ two disjoint sets. For every s-coloured layer G in $MG^{\otimes \mathcal{C}_1}(n)$ and q-coloured layer H in $MG^{\otimes \mathcal{C}_2}(m)$ the distributive operation*

$$MG^{\otimes \mathcal{C}_1}(n) \times MG^{\otimes \mathcal{C}_2}(m) \xrightarrow{\odot} MG^{\otimes \mathcal{C}_1 \cup C_2}(n + m - p)$$

produces a new $(s + q)$-colored layer $G \odot H$ where $p = |V(G) \cap V(H)|$.

Example 1 *For $n = 3, m = 4, s = q = 2$ and $p = 3$:*



Note that in $t = 0$ we have $s = q = 1$. New colours appear in a layer after more applications of $\odot$.

We speak of a single $|\mathcal{C}'|$-*coloured layer* into a $|\mathcal{C}|$-coloured multilayer for $\mathcal{C}' \subseteq \mathcal{C}$ from the interaction through $\odot$ among all the different layers. Roughly speaking, $\odot$ shows the way by which one can combine colours into a multilayer by the process of *merging layers* whenever they interact (see below).

Note that, in the sense of [Kivela et al., 2014], $\odot$ does not generate any *inter-layer edge* among layers but rather the *intra-layer edges* of a $(s + q)$-aspect layer from previous s-aspect and q-aspect. For instance, for $m_G(uv) = k$ and $m_H(uv) = k'$ with $col_G(uv) = \{c_1, ..., c_s\}$ and $col_H(uv) = \{c'_1, ..., c'_q\}$ we have $m_{G \odot H}(uv) = k + k'$ and $col_{G \odot H}(uv) = \{c_1, ..., c_s, c'_1, ..., c'_q\}$.

In other words, one can only consider the *inter-layer edges* that become new *intra-layer edges* under composition of layers, such as these interactions add up as new layers are incorporated.

This property is important because it emphasises the irreducibility of brain processes, i.e. the difficulty to isolate brain-body interactions as independent exchanges under awake conditions. In our framework, the conscious brain acts as an entangled system of layers. Moreover, the fact that every layer contains a unique and different colour in an initial unconscious time $t = 0$ ensures the distinction of different layers, even if later they are composed employing $\odot$. This particular *direct product* of multigraphs is chosen as a way of avoiding any information lost whenever forming new layers, i.e. $\odot$ keeps track of the whole picture of the multilayer when mixing colours.

Interaction of Layers

Another feature of these multilayers is the time indexation. To formally describe these time-evolving interactions, we introduce a notion of a multigraph endowed with a rotation angle.

Definition 4 *A* rotation graph *is a pair* $[G, \alpha]$ *where* G *is an edge-coloured multigraph and* $\alpha \in [-\pi/2, \pi/2]$.

Given an interval $T \subseteq \mathbb{R}^+ \cup \{0\}$ the angles considered in the sequel are of the form of continuous functions $\alpha : T \to [-\pi/2, \pi/2]$. Then, we obtain:

Definition 5 *Two rotation graphs* $[G, \alpha]$ *and* $[H, \beta]$ interact in a time t *if* $\alpha(t)$ *and* $\beta(t)$ *have different sign, considering* 0 *as a sign in itself, and* $|\alpha(t)| + |\beta(t)| \geq \pi/2$.

By interacting, new multigraphs are constructed by merging the edges present in the involved multigraphs. This is done by several applications of $\odot$.

Definition 6 *Given* $t \in T$ *and* PMon *the category of partial monoids, a* network rotation model *is a functor*

$$RMG^t : \mathcal{S} \to PMon$$

giving partial monoids.

$$RMG^t(n) = \{\textit{rotation graphs } [G,\alpha] \textit{ with } |V(G)| = n\}$$

We set $RMG^0(n)$ to contain only 1-coloured multigraphs, when there is no interaction yet. Every element of an object $RMG^t(n)$ obtained from a network rotation model is called a *rotation layer* or simply a *layer*.

Finally, the $\odot$ operation in the non-rotation case is extended to the rotation. We take into account that this operation is partial since it works only in the case that the involved layers do interact.

Definition 7 *For every interacting s-coloured layer $[G,\alpha]$ in $[RMG^t]^{\otimes\mathcal{C}_1}(n)$ and q-coloured layer $[H,\beta]$ in $[RMG^t]^{\otimes\mathcal{C}_2}(m)$ with $\mathcal{C}_1,\mathcal{C}_2 \subseteq \mathcal{C}$ the assignation*

$$[RMG^t]^{\otimes\mathcal{C}_1}(n) \times [RMG^t]^{\otimes\mathcal{C}_2}(m) \xrightarrow{\bigodot} [RMG^t]^{\otimes\mathcal{C}_1\cup\mathcal{C}_2}(k)$$

where $k = n + m - p$ and $p = |V(G) \cap V(H)|$ produces a new $(s+q)$-coloured layer containing a graph

$$[G,\alpha]\bigodot[H,\beta] = [G\odot H, min(\alpha(t),\beta(t)]$$

Note that $\odot$ is now bold.

According to this extension, every rotation layer mixes colours as in the previous subsection. This mixing involves the notion of *coupling*: the $\odot$ commutative operator preserves the individuality of layers' colours at the cost of losing their angle independence. Additionally, edges whose constituent nodes appear both in more than one layer (coloured more than once) are allowed. They are called *coupling edges* and admit an underlying *coupling graph* containing the basic configuration of the system. In followed sections, these nodes might correspond to observed brain-body regions whose contribution to the conscious experience is diverse.[1]

[1] In the extreme case that a coupling edge is coloured with all colours into $\mathcal{C}$, we have a pair of nodes that participate in every conscious operation that system could have and they should be seen as part of a *core structure* into the brain-body structure. Importantly, they are not the effective cause of consciousness, but the consequence of conscious operations as new coupling edge activities appear.

The introduction of rotation layers allows us to reason about multilayers with rotation on its constituent layers and describe their interactions by means of simple rotation drawings such as in figures 3.1D and subsequent figures.

3.5 Global Signatures of Consciousness

Following the conceptual and mathematical multilayer framework inspired by dynamic brain organization and brain-body couplings, we now discuss global signatures of consciousness. First, we review relevant evidence from three experimental paradigms to later integrate them under the multilayer umbrella.

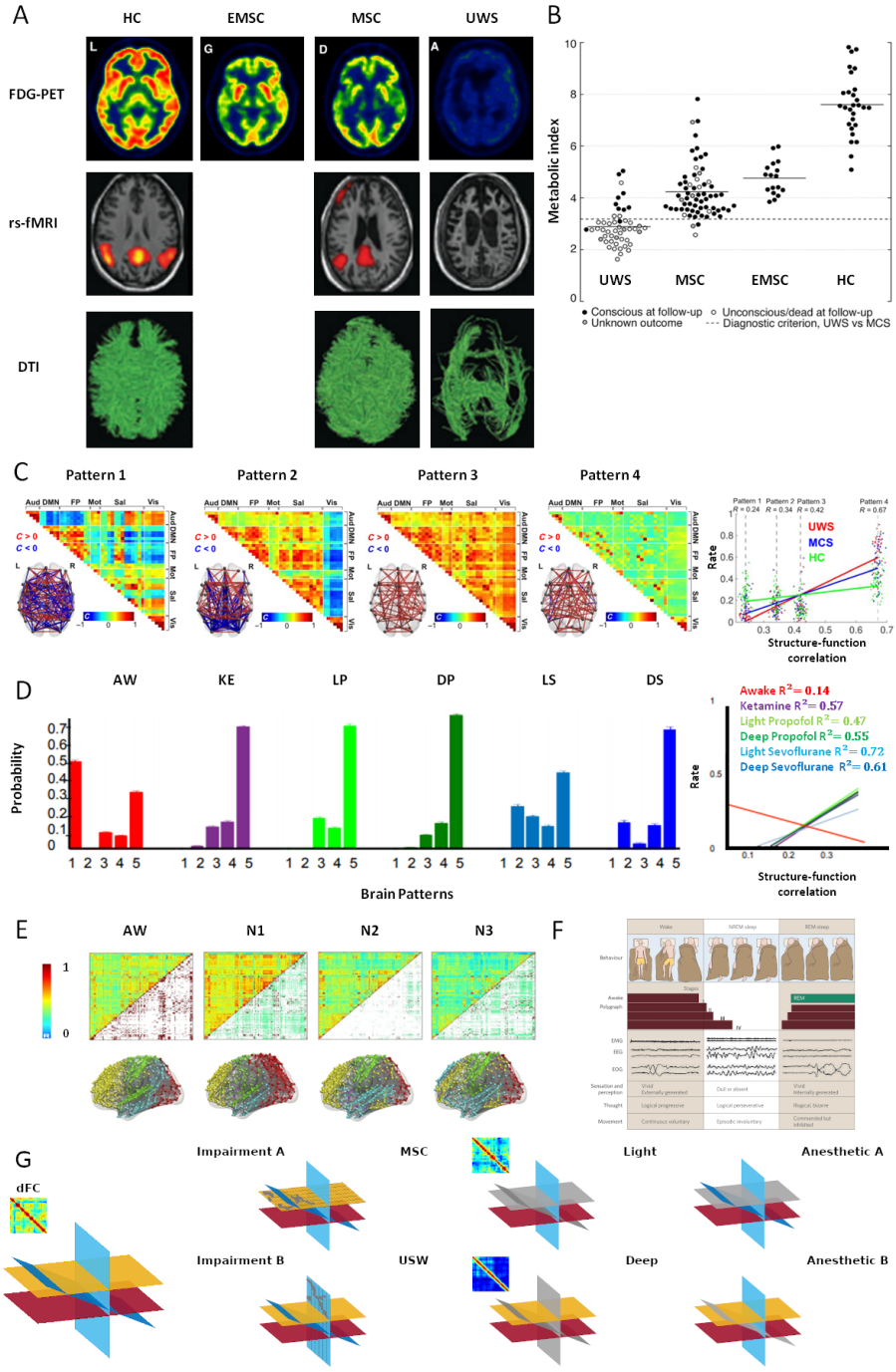


Figure 3.4: Global signatures of consciousness. (**A**) Multimodal neuroimages using [F]-fluoro-deoxyglucose positron emission tomography (FDG-PET), resting-state fMRI (rs-fMRI) and diffusion tensor imaging (DTI). Conditions from healthy controls, emergence from minimally conscious state (EMSC), minimally conscious state(MSC), to unresponsive wakefulness syndrome (UWS). Adapted from [Gosseries et al., 2014]. (**B**) Glucose Metabolism index for disorders of consciousness. Figure from [Stender et al., 2016]. (**C**) Dynamic brain patterns and disorders of consciousness. Pattern 1 is complex and far from anatomical driven activity. Contrary, pattern 4 is close to anatomical driven connectivity. Healthy controls presented a dynamical switch among these four patterns, while impaired conditions remained mainly in pattern 3 and 4 [Demertzi et al., 2019]. (**D**) Similar analysis with five patterns for awake (AW), ketamine (KE), light propofol (LP), deep propofol (DP), light sevoflurane (LS), deep sevoflurane (DS). Bars show the probability to find brain activity in one or another pattern. Adapted from the supplementary images in [Uhrig et al., 2018]. (**E**) In deep sleep stages, functional connectivity and brain dynamics come close to anatomical connectivity patterns, as well as the modular organization for those functional networks. Adapted from [Tagliazucchi et al., 2016b]. (**F**) Behavioural, polygraphic, multimodal brain measures, and psychological descriptions for awake, NREM and REM sleep in humans [Hobson, 2005]. (**G**) In our conceptual model, both layer impairment and/or anaesthetized paths lead to loss of consciousness. Impairments in one or another layer might lead to MSC or UWS, as well as different anaesthetics may target different layers. The concentrations among anaesthetics drive light or deep lack of consciousness, according to which layer configuration is further affected by them.

3.5.1 Disorders of consciousness

The characteristic global dynamic of awake conditions disappears with the loss of consciousness, regardless of clinical diagnosis, anaesthetic or deep sleep.

Characterizing disorders of consciousness

Chronic loss of consciousness involves different cortical and subcortical brain impairments. An fMRI meta-analysis of resting-state activity reveals that disorders of consciousness present reduction of activity in midline cortical and subcortical sites of default mode network. For instance, the vegetative state (VS) presents less activity than the minimally conscious state (MCS) patients [Hannawi et al., 2015]. Some areas involved are the left cingulate gyrus, posterior cingulate cortex, precuneus, medial temporal lobe, middle frontal lobe, and bilateral medial dorsal nuclei of the thalamus. Moreover, lesions in the intralaminar thalamus and the midline nuclei of the pons, inevitably lead to coma [Posner et al., 2008], while mesial parietal cortex, posterior cingulated cortex and precuneus are the first areas to reactivate when patients recover consciousness [Laureys et al., 2006].

Pathological impairment in anticorrelated networks impact on conscious cognition. MCS patients recover consciousness if anticorrelations are partially preserved [Di Perri et al., 2016]. Contrary, lack or pathological negative anticorrelations were related to remaining unconscious. Rather than linear transitions, these connectivity impairments look like exponential decay [Demertzi et al., 2014, Vanhaudenhuyse et al., 2010].

Morphometry indicates that differences in connectivity are not only attributable to structural damage [Di Perri et al., 2016], but also to impairment in brain metabolism. Under disorders of consciousness, the reduction of energy production and metabolism is well-documented [Laureys, 2005, Laureys et al., 1999, Di Perri et al., 2016, Stender et al., 2016]. Metabolic levels in whole-brain are based on glucose concentrations and measured by positron emission tomography (PET) with [F]-fluorodeoxyglucose (FDG) (Figure 3.4A). The recovery from vegetative and MCS correlates with whole-brain energetic states above a certain metabolic boundary [Stender et al., 2016]. These results support the idea of a minimal energetic requirement for awareness, making metabolism an indicator of recovery (Figure 3.4B). This requirement seems to be 42% of normal cortical activity, from which 94% of patients return to awareness. This

minimal energetic requirement is consistent with the need of a whole-brain energetic state, instead of the specific activity of a few networks or areas [Stender et al., 2016, Shulman et al., 2009].

Alterations of consciousness also present characteristic dynamical signatures (Figure 3.4C). Analyses on three human conditions, unresponsive wakefulness syndrome (UWS), MCS and healthy controls (HC), found different dynamical patterns across dynamical states. The most complex pattern (pattern 1), mainly appearing in HC conditions, seems relevant to support consciousnesses [Demertzi et al., 2019]. This pattern is characterized by higher spatial complexity, long-distant edges, high modularity, high efficiency and low similarity with respect to the anatomical structure. Another pattern (pattern 4) appears predominantly during unconscious conditions and it is associated with reduced or absence of consciousness. This last pattern presents low interregional dynamic coordination, low efficiency and high similarity to anatomical connectivity [Demertzi et al., 2019].

Global measures of consciousness

Different clinical scales [Giacino et al., 2004, Harrison and Connolly, 2013] and measurement techniques try to quantify consciousness. One example is the Local-Global paradigm [Bekinschtein et al., 2009, King et al., 2013]. In this paradigm, global deviations of auditory patterns correlate with late and spatially distributed brain activity under attentive subjects. Local irregularities do not show this behaviour. The paradigm was tested in healthy humans, patients with disorders of consciousness as well as monkeys in anaesthesia [Uhrig et al., 2016]. In this last report, anaesthesia does not fully suppress the brain activity to the global sequence, but there is an absence of response to the global violations in areas such as the parietal cortex and thalamus in both propofol and ketamine. Moreover, in the auditory cortex, the local effect is replaced by a global effect in presence of propofol.

Another measure of consciousness is the Perturbational complexity index (PCI) [Casali et al., 2013]. This index is computed from the compression of EEG brain signals triggered by TMS stimulation. PCI exhibits high sensitivity and specificity recognizing conscious and unconscious conditions [Casarotto et al., 2016]. This sensitivity is higher than the 78% from other qualitative assessments of EEG data [Sitt et al., 2014]. It supports the idea that additional

diagnose information is not captured by only qualitative EEG signal description.

More complex measures are also in the process of development. Unfortunately, the two examples above assume one unique measurable dimension of consciousness: the levels of consciousness [Bayne et al., 2016]. This approach misses more complex or extra features of consciousness. Recently, some prototypes tried to confront this issue, incorporating multi-dimensional cognitive evaluations [Sergent et al., 2017]. Other studies add machine learning techniques and a combination of markers that let to a more reliable classification of conscious disorders as well as using economical and accessible EEG data [Engemann et al., 2018]. Nevertheless, the fundamental principles to define accuracy and an optimal threshold for measures of consciousness remain elusive [Harrison and Connolly, 2013].

A combined theoretical approach is still needed [Gosseries et al., 2014] and our multilayer integrative framework may give some light (Section 3.5.4).

Partial recovery in chronic loss of consciousness

After diagnosis, one would expect to manipulate the damaged brain and recover consciousness, i.e. therapeutic intervention.

There are some cases of modulation and partial recovery in chronic loss of consciousness. Electrical stimulation in thalamus presents some effects [Schiff et al., 2007, Schiff, 2010, Schiff, 2013]. For example, bilateral deep brain stimulation (DBS) of the central thalamus had a positive effect on behavioural responsiveness in one patient under MCS during six years [Schiff, 2013]. Stimulation of the vagus nerve also presents effects. One patient locked during 15 years in vegetative state presented signals of recovering, improving responsiveness and brain connectivity patterns after vagus nerve stimulation [Corazzol et al., 2017]. This stimulation also seems to increase metabolism in the forebrain, thalamus and reticular formation [Corazzol et al., 2017, Henry et al., 1999]. This particular case suggests beneficial effects of vagus nerve stimulation such as reactivating the sensory/visceral afferents and enhancing brain activity within a brain-body close-loop process.

On the other side, excessive stimulation also disrupts consciousness, like in epileptic seizures [Mateos et al., 2018]. In [Koubeissi et al., 2014], the authors report reversible disruption of con-

sciousness in one patient using electrical stimulation between the left claustrum and anterior-dorsal insula. This stimulation generates a loss of consciousness together with the increase of EEG synchrony within frontal-parietal networks. This network is associated with conscious perception and these findings suggest that increased correlations between these areas can also disrupt consciousness.

These results provide evidence that in principle is possible to promote functional recovery from traumatic brain injury by different paths.

3.5.2 Anesthesia

General anaesthesia is a unique reversible experimental paradigm to investigate the brain mechanisms of consciousness, their dynamic brain and behavioural disruptions.

Characterizing dynamics under anesthesia

Under anaesthesia, the brain patterns are non uniformly modulated. Some areas appear more deactivated than others, while most anaesthetics, with exception of ketamine [Langsjo et al., 2005], cause a global reduction of cerebral blood flow [Franks, 2008]. For example, in anaesthetized monkey, DMN and resting-state networks do not disappear at all, i.e. anticorrelations are still present [Atasoy et al., 2017, Vincent et al., 2007]. Specifically, precuneus cortex and lateral temporoparietal components of DMN persisted under anaesthesia [Vincent et al., 2007], but the connectivity of the posterior cingulate cortex (PCC) is reduced during sedation [Greicius et al., 2008]. Despite DMNs being preserved, under moderate propofol, changes of PCC connectivity are observed between areas such as somatomotor cortex, the anterior thalamic nuclei and the reticular activating system [Stamatakis et al., 2010]. Moreover, the connectivity of primary sensory cortices lightly increases while connectivity in high order areas (hippocampus and insula) decreases [Martuzzi et al., 2010]. It might partially explain the observed decrease in the spatial extension of anticorrelations [Barttfeld et al., 2015, Boveroux et al., 2010]. Together, the evidence indicates that anaesthesia modulates the strength of functional connectivity and although DMN is present under sedation, widespread changes in those thalamocortical patterns are observed [Boveroux et al., 2010, Martuzzi et al., 2010, Schrouff et al., 2011].

Dynamical resting-state analyses indicate similarities under different anaesthetics (Figure 3.4D). Recent studies in macaque brains compare dynamical pattern configurations between awake, anaesthetics and their concentrations [Barttfeld et al., 2015, Uhrig et al., 2018]: ketamine (KE), light propofol (LP), deep propofol (DP), light sevoflurane (LS) and deep sevoflurane (DS). In awake condition patterns of brain activity display richer and flexible functional brain configurations. These patterns are far from anatomical connectivity. Contrary, under anaesthetics conditions, brain patterns tend to remain close to brain anatomy. These results suggest the decrease of dynamic repertoire despite dissimilar anaesthetics undergoing different pharmacological paths [Hudetz and Mashour, 2016].

Loss and recovery under anaesthesia

Under anaesthesia, loss of consciousness (LOC) emulates sleep phase transitions and coma brain states. During the period of induction, the patient may enter in a paradoxical excitation, defensive movements, incoherent speech, euphoria or dysphoria and increase in heart rate and beta activity EEG [Bevan et al., 1997]. For instance, four EEG patterns define anaesthetic transitions phases. The first three present similarities with sleep transitions, but the final phase resembles brain death. The first phase shows a decrease of beta activity (13-30 Hz) and an increase of alpha activity (8-12 Hz) [Feshchenko et al., 2004]. In the second phase, alpha and delta activity (0-4 Hz) increase, similarly to sleep stage 3. For the third phase, EEG presents flat signals interspersed with alpha and beta activity (burst suppression). Finally, the last phase corresponds to an isoelectric or flat activity, similar, but not equal, to coma or brain death [Brown et al., 2010].

On the other hand, recovery of consciousness (ROC) behaviour seems to correlate with age and anaesthetic types. Emergence from anaesthesia is a passive process where the return of spontaneous respiration is one of the first indicators of recovery, the heart rate and blood pressure increase, salivation and tearing begin and EEG patterns reverse the order of phases to an active EEG. Furthermore, data shows that recovery from anaesthesia differs in elderly subjects [Canet et al., 2003, Lepousé et al., 2006] and it may generate confusion/delirium if it is followed by abrupt recovery. In a study with 393 subjects, the third part of them experienced abrupt slow-wave emergence trajectories, showing a bimodal distribution of abrupt and gradual emergence after surgery [Warnaby et al., 2017]. This type of slow-wave emergence trajectory was predicted in 78% of cases, only using clinically relevant parameters, such as age and type

of anaesthetic.

General anaesthesia also induces asymmetries as part of the modulation of LOC and ROC. One of them is the asymmetry between the concentration needed for induction and recovery from anaesthetic conditions [Sepúlveda et al., 2018, Engbers, 2018]. This phenomenon is associated with hysteresis and sometimes interpreted as support for neural inertia [Friedman et al., 2010, Tarnal et al., 2016, Warnaby et al., 2017], i.e. the tendency to resist transitions. In this line, observed abrupt and gradual awake-anaesthetic transitions (induction-emergence) support different associated neural processes for LOC and ROC [Chander et al., 2014, Lee et al., 2011, Warnaby et al., 2017, Priesemann et al., 2014]. For example, stochastic mathematical models need more than two states to explain asymmetries and variability on ROC [Proekt and Hudson, 2018]. Other analyses indicate that the brain, in order to recover consciousness, progresses through well organized discrete metastable states or hubs of patterns that connect others patterns otherwise disconnected [Hudson et al., 2014].

These transitions seem to have continuous and discrete dimensions. Despite propofol, LOC induction generates an abrupt neural network fragmentation [Lewis et al., 2012], topological analyses of EEG data in human demonstrate that both continuous and discrete components participate in transitions between awake and anaesthetized states [Lee et al., 2011]. At LOC and ROC, global efficiency has two main contributions, from structure topology and strength connectivity. The contribution of the former is similar in most subjects, while the contribution of the later showed two groups of responses: a pattern of slow decay and sudden recovery, and another pattern of sudden decay and slow return. These components were more evident in the parietal cortex than frontal regions [Lee et al., 2011]. Computational simulations, virtually injuring monkey brains, also show how parietal lobe disruptions break integrative aspects of cortical functions more notoriously than other areas [Honey and Sporns, 2008]. These results support dissociable network properties between structure and strength, and a relevant role of parietal cortex on mechanisms of anaesthesia [Lee et al., 2011].

3.5.3 Sleep and wake transitions

Although anaesthesia and sleep correspond to different brain organizations, they share common features [Franks, 2008]. For instance, under light concentrations of ketamine as well as REM sleep, a basic conscious experience seems to remain, despite the body being disconnected from

the external environment [Collier, 1972]. Another example, propofol-induced LOC is related to the emergence of slow waves resembling the slow waves in non-REM (NREM) sleep and sharing similar starting and propagation patterns [Lewis et al., 2012]. In NREM, for example, similar areas disrupted by anaesthesia present marked deactivation, some of them: thalamus, the brainstem, basal forebrain, basal ganglia, frontal and parietal cortices (particularly, anterior cingulate, orbitofrontal cortices, precuneus and posterior cingulate). Nevertheless, contrary to anaesthetic transitions, the daily transitions between awake and sleep stages are familiar, making sleep protocols a natural experimental paradigm for consciousness research.

Characterizing sleep transitions

During NREM, independent modules emerge, increasing intraconnectivity interactions, in detriment of the observed awake network interconnectivity (Figure 3.4E). This change suggests a reconfiguration of large scale brain network organization [Boly et al., 2012b], together with transitions from couple to decouple of semi-independent subsystems. Moreover, dynamic patterns increase their similarity with anatomical connectivity as one progresses into deeper states of sleep [Tagliazucchi et al., 2016b], like in previously discussed evidence for anaesthesia and disorders of consciousness. In this respect, comparisons of subjects in awake versus sleep, vegetative and anaesthesia condition reveal that neural activity driven by an external stimulation spreads through different areas of the cortex when subjects are awake, but remains local when they do not [Casali et al., 2013, Massimini et al., 2009, Rosanova et al., 2012, Sarasso et al., 2014]. For wake condition, pulses driven by TMS generate richer and sequential EEG signals in different cortical areas, while the absence of awareness, in general, does not.

Interestingly, almost every animal species exhibits sleep-like behaviour, making sleep a relevant and common function for the neural organization. For example, worm brains present sleep cycles characterized by brain state trajectories [Nichols et al., 2017]. In that study, the authors demonstrated that almost 75% of neurons in the Caenorhabditis elegans become inactive during sleep, while γ-aminobutyric acid-producing (GABAergic) and peptidergic head neurons such as the sleep-promoting interneurons remain active. A neuropeptide receptor (NPR-1) expressed in a hub interneuron regulates arousal cues, that together with environmental conditions led to sleep-wake transitions [Nichols et al., 2017]. In mammalians, the sleep-wake transition is characterized by a change from EEG low-amplitude high-frequency alpha oscillations (8-12 Hz) to high-amplitude low-frequency delta waves (0.5-4 Hz) [Simon and Emmons,

1956]. This basic signature is probably the most robust change between awake and sleep condition in mammalians (Figure 3.4F). Furthermore, during sleep, there are at least other two transitions between REM and NREM sleep. REM stages are characterized by low-amplitude high-frequency oscillations in a wakefulness-like manner, eye movements, dreaming, irregularities in respiration and heart rate together with skeletal-muscle hypotonia [Brown et al., 2010]. NREM stages are accompanied by slow waves and sleep spindles, waxing and waning muscle tone, a decrease of body temperature and heart rate. Metabolism and blood pressure also decrease in REM [Buchsbaum et al., 2001] and continue decreasing with the depth of NREM sleep [Maquet, 1995].

Neural bases of sleep

Cortical slow-waves are the predominant signature for sleep transitions and loss of consciousness. This new oscillatory mode apparently causes a breakdown of causal interactions among the cortex [Pigorini et al., 2015], and interactions associated with consciousness, as reviewed in previous sections.

The origins of slow-waves correspond to intricate networks of interactions, but in a broader sense, the activity of neocortical neurons is influenced by physiological, circadian, and pharmacological changes [Pace-Schott and Hobson, 2002]. Just as a way of example, circadian cycles are controlled by special genes inside thousands of cells in the suprachiasmatic nucleus of the hypothalamus. These interactions all together tend to hyperpolarized neocortical neurons, triggering oscillations between a bistable depolarized (up) and hyperpolarized (down) states at around 1Hz [Hobson and Pace-Schott, 2002]. These oscillations become travelling waves and emerge as the slow-wave activity observed in the scalp [Massimini et al., 2004, Nir et al., 2011]. This slow-wave activity is thought to represent the switch of the thalamus from tonic to a busting mode [Steriade and Timofeev, 2003, Steriade et al., 2001]. The tonic mode would facilitate sensory activity transmitted through the thalamus to higher cortical regions, while the bursting mode would break down that exchange.

Interestingly, the slow-wave cortical rhythm also appears in isolated cortical tissue, supporting the idea that it is an intrinsic dynamic mode of the cortex [Steriade et al., 1993]. Moreover, during the first NREM stage, slow-wave activity is greater in frontal than parietal and occipital areas [Aeschbach et al., 2001], evidence that together with another group of experiments indi-

cate that frontal cortex is the first region to ”fall asleep” and the latest to recover [Achermann et al., 1995, Hobson and Pace-Schott, 2002]. In waking and REM stages, ascending arousal systems through thalamocortical interactions seems to suppress the autonomous cortical slow waves which characterized the NREM stage.

Slow-wave sleep, as a marker of loss of consciousness in sleep transitions, therefore, corresponds to an intricate coupling and decoupling of the thalamus and cortical systems, either driven by frontal regions or ascending-thalamus modulations.

3.5.4 A common layer mechanism

Through previous sections, we have reviewed different sources of evidence that indicate the complex intertwined interactions supporting consciousness at different scales.

There are common fingerprints of loss and recovery of consciousness but also differences. Slow-waves of sleep activity displays similarities with the electrophysiological activity observed during anaesthesia and other impaired conditions [Horovitz et al., 2009]. The dynamical disruptions of long-distance networks are also a common sign of injuries-induced LOC, anaesthesia-induced LOC and sleep (Figure 3.4C-E). These global effects are independent of anaesthetic pharmacology [Uhrig et al., 2018], as well as the metabolic reduction is a common marker across different types of brain injury [Stender et al., 2016].

Nevertheless, the causally driven forces that generate these similar global disruptions seem of different nature. Different anaesthetics follow different molecular paths [Hudetz and Mashour, 2016], injuries might disrupt different regions generating similar global effects, while sleep transitions do not need any brain impairment to switch off the cortex every night. Some local network analyses suggest these changes might be through thalamocortical disruptions while others due to higher-order front-parietal associative network disruption [Boly et al., 2012a]. Therefore, it is still unclear how these functional disruptions are induced across different experimental paradigms.

Altogether, the evidence as a whole imposes the greatest challenge for current theories. For example, some theories of anaesthesia would predict graded, continuous transitions [Alkire et al., 2008] and other discrete phase transitions [Steyn-Ross et al., 2001, Steyn-Ross et al.,

2004]. However, the evidence reviewed above support both types of transitions at different network levels [Lee et al., 2011]. Moreover, the paradoxical excitation at LOC and the abrupt ROC correlated with elderly subjects [Warnaby et al., 2017] also challenges current theoretical frameworks, while a common global mechanism is desirable to explain how correlations and anticorrelations (which are not reduce to only anatomical connectivity) persist under certain conditions but not others.

Here, we offer a parsimonious explanation based on the impairment of hypothetical native neural, physiological and molecular layer interactions (Figure 3.4G). Independently of the pharmacological path or localization of brain-body impairments, disruptions of consciousness involve one or more layer disruption. For example, impairments may affect portions of one or another layer (grey surface in Figure 3.4G), leading to one or another injury state (e.g. MSC or UWS). Similarly, different anaesthetics may turn off one or another layer, and according to their concentrations, these layer impairments may or may not influence subsequent layers. Some of these combinations will lead to light or deep anaesthetic state (Figure 3.4G). In this model, unconscious modes are characterized by parallel non-interacting layers, while minimally conscious modes still allow minimal interinteractions among unparalleled layers. These two configurations drive different dynamical activities, as shown by the example of dynamical functional connectivity (dFC) in figure 3.4G.

This system-level approach integrates the previous evidence. The rapid fragmentation of propofol-induced unconsciousness [Lewis et al., 2012] and modularity changes under deep sleep [Tagliazucchi et al., 2016b] may correspond to the parallel layer configuration of minimal coupling. The slow-wave activity being the intrinsic activity of a decoupled cortex. Local brain integration observed across different unconscious conditions may represent parallel or minimally interacting layers, while the richer global dynamic of awake conditions may correspond to native layers interconnected configurations. In this line, the modulation of DMNs may parallel the transitions from interconnectivity to intraconnectivity driven dynamics of native semi-independent layers. During sleep, DMNs persist [Horovitz et al., 2009], but like anaesthesia, important changes in cortical coupling are observed at different stages. In early stages, anticorrelations do not disappear but are modulated [Fukunaga et al., 2006, Picchioni et al., 2008] and eventually, frontal cortex (medial prefrontal cortex) functionally decouple from the DMN [Horovitz et al., 2009] (Figure 3.5A). This observed decouple may correspond to the final stage of different cortical and subcortical layers turning to their independent or intrinsic

dynamics. The switch of many layer interacting would explain why there is a non-binary absence or presence of functional connectivity across consciousness modes [Atasoy et al., 2017]. Some layers may decouple first than others, and the order of this process changes according to anaesthetics, injuries or sleep stages. The paradoxical excitation of small anaesthetic dose may correspond to the switch from awake layer configurations to faster transitions of coupling and decoupling layers, generating experiences closer to dreams, psychedelics and hallucinations. In elderly subjects, the layer boundaries may naturally become fuzzy, as the brain decreases in size, the functional extensions of some layers contract, making some layers suddenly interact once anaesthetic effects drop off. Later, the recovery of consciousness and the discrete metastable state activity observed [Hudson et al., 2014] may represent the transitions between parallel semi-independent layers to interacting ones.

The constitution of these independent or semi-independent layers might involve different cortical layers, subcortical regions as well as more general brain-body systems, like molecular diffusion of neurotransmitters. Some examples were discussed in section 3.4, and we leave for future empirical developments the concrete definition of these layers, their regions and constituent systems. Further extensions may incorporate the native layer concept, in order to test their dynamic relevance for consciousness.

3.6 Conscious Experience

Disorders of consciousness, anaesthesia and sleep are useful experimental paradigms to study global correlates of consciousness. In addition, another set of paradigms is required to deepen the study of phenomenal experience, subjectivity and its experiential content: dreams, meditation, conscious perception and their alterations are some of them.

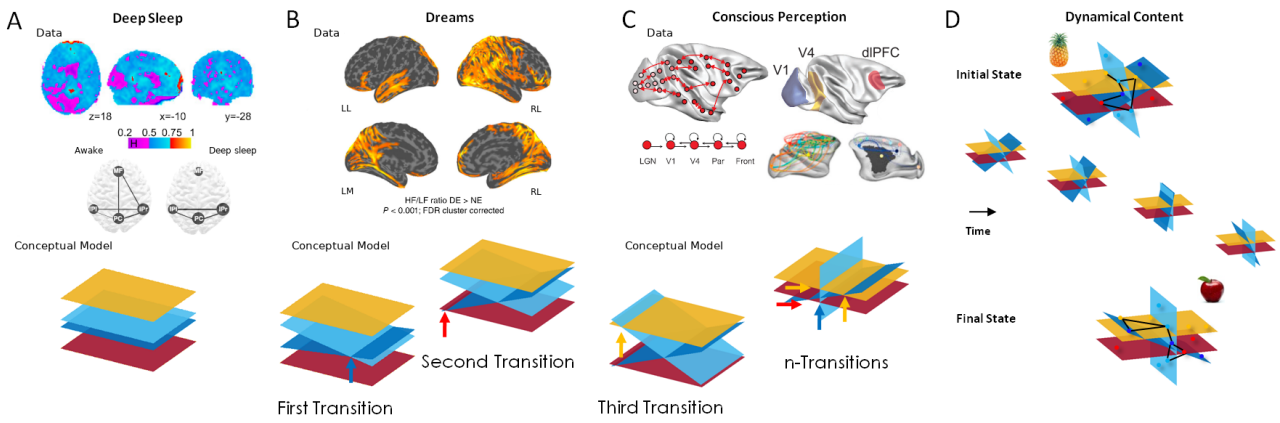


Figure 3.5: Brain signatures of conscious experience. (**A**) Hurst exponent map (all voxels, average all subjects) showed short-range temporal correlations in deep sleep [Tagliazucchi et al., 2013]. Moreover, deep sleep is characterized by short-range temporal correlations and decouple of observed default mode networks [Horovitz et al., 2009]. The posterior areas (bilateral IPC and PCC) strengthen their connectivity, whereas the connections between frontal and posterior regions are lost. In a multilayer model, this condition is described by independent layers (non-interinteractions). (**B**) Brain signatures of dreams [Siclari et al., 2017]. Dream experience is correlated with local decreases in low-frequency (LF) activity in posterior cortical regions and higher high-frequency (HF) versus LF power ratio than non-dreaming experiences. In the model, these signatures correspond to the activity after some layers start to interact. (**C**) Brain signatures of conscious perception reveal a complex network of interactions, together with ignition activity from PFC to the rest of the brain. Adapted figures from [Van Vugt et al., 2018, Uhrig et al., 2018, Signorelli and Meling, 2021]. In our model, these activities are triggered for further dynamical layer configurations. (**D**) Overlapping regions may correspond to the phenomenological content of conscious experience. Different dynamical configurations, number and types of layers involved, degrees of interaction, types of oscillation and oscillatory mechanisms, among others, would play a role in conscious content. Here, an example of layer configurations from the content of pineapple to apple, and their dynamical changes: the blue layers change position, while the others remain fixed. These changes inform about new dynamical states.

3.6.1 Dreams experience

Dreams appear in both REM and NREM sleep stages [Stickgold et al., 2001]. During decades dreams were associated only with REM and its electrophysiological similarities with awake conditions, such as global high-frequency EEG activity. These similarities, however, do not include the full awake brain activity. In REM, the forebrain is activated through the ascending arousal system but mainly driven by the cholinergic path of the brainstem [Steriade, 2000] and basal forebrain [Szymusiak, 1995]. Limbic and paralimbic cortical and subcortical areas such as the amygdala, anterior cingulate, orbitofrontal and insular cortices, are selectively activated [Braun et al., 1998, Maquet et al., 1996]. These selective activations are thought to mirror content specificity [Hobson and Pace-Schott, 2002]. Moreover, the disruption of the inferior parietal lobe and specifically the Brodmann area 40, by its own, can produce the absence of dreaming experiences [Solms, 1997]. More recently, dreams are also detected during NREM stages, where subjects reported dreaming-like experiences once interrupted in both REM and NREM stages [Siclari et al., 2017].

Dreams are a unique paradigm to avoid report confounding factors. In one recent study, subjective dream experience was correlated with a decrease of low-frequency activity in EEG

signals from the parietal-occipital cortex. This reduction was independent of remembering the contents, whereas a local increase in low-frequency power correlated with the absence of experience [Siclari et al., 2017]. When participants recalled the content of their dreams, a higher high-frequency power in medial and lateral frontoparietal cortex appeared. Using this information, the authors were able to predict with 87% accuracy if participants were dreaming or not. Moreover, depending on the areas involved, part of their content experience was also predicted successfully. The experiment suggests that the thinking dimension or thought-like experience is related to frontal areas, while perceptual experience alone would correspond to parietal, occipital and temporal regions. Together, cortical relevant regions for dreaming experience involve bilateral occipital, medial and lateral parietal, medial temporal and inferior frontal cortex (Figure 3.5B). Reports of lucid dreams support this claim, during dreams where subjects have some degree of control, frontoparietal regions were activated [Dresler et al., 2012, Voss et al., 2009].

Furthermore, dreams also represent a unique paradigm to study subjective experience itself [Hobson, 2005]. For instance, dreams are associated with a hallucinatory experience. In one study, subjects were asked about their mental experience in different periods such as active and quite awake, sleep onset, REM and NREM. Their answers were classified on a scale of bizarreness, from hallucinatory to thinking/manipulated perception. Dreams under REM presented higher hallucinatory mental content and less thinking components in comparison with quite awake condition. The opposite relation appears in awake conditions [Fosse et al., 2001]. Other experiments found that dreamsb́izarreness of healthy subjects were close to the awake reports of psychotic schizophrenic patients [Scarone et al., 2008]. This result suggests an interesting connection between dreams and hallucinatory experiences. One of these connections is also neural. The stimulation of precuneus, cingulate cortex and retrosplenial cortices can induce different feelings such as ”being in parallel worlds” or in a ”dream-like state” [Balestrini et al., 2016, Herbet et al., 2014]. These feelings are also found in hallucinations. In this sense, dreams may correspond to hyper associative conscious experience, or ”normal delirium”, presenting features such as hallucinosis, disorientation, confabulations and sometimes memory loss [Hobson and Pace-Schott, 2002].

3.6.2 Psychedelic experience

From the experiential similarity of dreams and hallucinations [Sanz and Tagliazucchi, 2018], a new paradigm has emerged: the use of psychedelics to study altered states of consciousness

(ASC). Under psychedelics, subjects experience visual hallucinations and alteration in many aspects of experience such as sense perception, emotions and cognition. Some of these dimensions are enhanced while others are compromised [Bayne and Carter, 2018]. For example, the experience of time and space unity increase, but the recognition of objective intervals is impaired. Psychedelics impose new restrictions and challenges for current unidimensional models of consciousness [Bayne and Carter, 2018, Bayne et al., 2016].

Under psychedelic, the resting brain increases its global functional connectivity [Tagliazucchi et al., 2016c], while the disintegration of DMN and decrease of alpha activity in the posterior cingulate cortex (PCC) correlate with ego-dissolution [Carhart-Harris et al., 2014, Carhart-Harris et al., 2016], i.e. the dissolution of boundaries between the ”self” and the environment. In the neural dimension, doses of ayahuasca reduce alpha (8-13 HZ) oscillatory power in EEG [Schenberg et al., 2015]. Something observed in all serotonergic psychedelics. This is also observed in the low-frequency range (0.01-0.1 Hz) in fMRI studies, together with new enhanced connectivity patterns [Tagliazucchi et al., 2014]. After psilocybin administration, MEG activity decreased in the ACC/medial prefrontal cortex (mPFC). This reduction is correlated with the intensity of the subjective effects. Experiments with lysergic acid diethylamide (LSD) reveal an increase in visual cortex cerebral blood flux (CBF) related to hallucinations, increase in resting-state FC between PH, dorsal mPFC and right mPFC, while decrease between parahippocampus (PH), retrosplenial cortex (RSC) and PCC connectivity correlated with ego-dissolution and ”alerted meaning”. These couplings seem relevant to the notion of ”self” and ”meaning”. In short, psychedelics reduced the stability and integrity of established brain networks [Muthukumaraswamy et al., 2013, Carhart-Harris et al., 2014].

A common mechanism points out to a disruption or unbalance between systems. Psychedelic effects seem to result from increasing cortical excitation via the stimulation of 5-HT2A receptors [Glennon et al., 1984]. However, this effect is different to most brain activity observed under anaesthetics, coma or deep sleep conditions. Under LSD, measures of MEG oscillatory activity decrease in power in delta (1-4 HZ), theta (4-8 HZ) alpha (8-15 HZ) and beta (15-30 HZ) [Carhart-Harris et al., 2016, Tagliazucchi et al., 2016c]. It comes across as a global increase in functional connectivity within high-level association cortical regions overlapping with default mode, salience, and frontoparietal attention network as well as thalamus [Tagliazucchi et al., 2016c]. Furthermore, network modularity decreases, carrying a disruption in intraconnectivity networks (within-modules) in favour of interconnectivity (between-modules)

[Tagliazucchi et al., 2016c].

Together, disruptions in consciousness are either driven by anaesthetics increasing inhibitory activity, anaesthetics/sleep decreasing excitatory activity, or psychedelics increasing excitatory activity [Atasoy et al., 2017].

3.6.3 Meditation experiences

Psychedelics present similarities but also differences with meditation states [Millière et al., 2018]. In both cases, there is a common experience of ego dissolution. However, while psychedelic experience corresponds to phenomenal excitation along with brain overexcitation, during meditation the phenomenal experience is calm and quiet. This is why meditation states are linked with the concept of pure awareness, awareness of awareness or minimal phenomenological awareness.

Meditation is the ancient practice of turning attention, perception and emotion into a calm and relaxed steady state. More recently, neuroimaging studies have found dissociable patterns of brain activation according to different meditation styles [Fox et al., 2016], where the most common and recurrent involved area is the insular cortex related to multisensory interoceptive awareness. This connection might be explained by the attentional control of breathing during different types of meditation [Millière et al., 2018]. Other studies also report a decrease on DMN nodes, such as the medial prefrontal cortex and posterior cingulate cortex [Garrison et al., 2015, Lutz et al., 2016, Scheibner et al., 2017].

However, the most interesting results point out long term effects of the practice of meditation in functional brain activity [Lutz et al., 2004, Rodríguez, 2008, Taylor et al., 2013]. In other words, a mental, intentional and deliberate sustained state can permanently modify the brain activity. These results are interpreted as submergent effects [Rodríguez, 2008] and associated with the circular causality of embodied cognition suggested by Merleau-Ponty and others [de Rezende, 1975, Merleau-Ponty, 2005, Thompson and Varela, 2001].

Circular causality means that the action and sensation of cognitive systems are both the cause and effect of any ”intentional arc” [Thompson and Varela, 2001, Merleau-Ponty, 2005, Gómez-Ramirez, 2014]. Therefore, conscious and intentional mental states are not merely

reduced to brain-body interactions, i.e. emergent properties, but they also trigger submergent processes that dynamically modify the cellular substrate. These submerged interactions are analogue to the observed structural and local modification by recurrent global Chladni patterns in a violin [Rodríguez, 2008]. According to this phenomenological view, consciousness *co-arise* with the conjunction of many and different brain-body interactions[Signorelli and Meling, 2021].

3.6.4 Conscious perception

Awake conscious perception requires time. Unconscious and automatic tasks take $\sim$40 ms, while under a different set of experiments, conscious perception is correlated with brain signals at $\sim$200-500 ms [Varela et al., 2001, Dehaene and Changeux, 2011]. One example is the positive peak of event-related potentials (ERPs) in a conscious report of a seen stimuli [Del Cul et al., 2007, Herzog et al., 2016]. These time windows inspire two-stage models, i.e. there is a preliminary unconscious stage processing, followed by a second conscious stage processing [Herzog et al., 2016]. Consciousness would correspond to the second stage.

However, the temporal nature of these stages is unknown. The mechanisms of conscious perception might be continuous or discrete [VanRullen and Koch, 2003, Wittmann, 2011, White, 2018]. Some experiments suggest discrete mechanisms [Chakravarthi and VanRullen, 2012, Herzog et al., 2016, VanRullen and Koch, 2003]. For example, the colour phi phenomenon [Koler and Von Grünau, 1976]. In two different locations, two disks of different colour are presented in rapid succession. The observer perceives one disk moving between both positions and changing the colour in the middle of the trajectory. One expects that the experience of changing colour would be after the second disk is seen, but this is not the case. Therefore, the perception seems to be constructed retrospectively, which is contrary to continuous theories [Bachmann et al., 2004, Herzog et al., 2016, Koler and Von Grünau, 1976]. Other discussions, however, point out the inconsistencies of discrete frames [Wittmann, 2011, White, 2018], questioning the definition of the frame itself, how they work and the variety of time scales experimentally observed. If there is such discrete processing, how new stimuli enter into the present psychological moment? Is it stored in a buffer prior to entering in the next conscious moment? How? [White, 2018].

In order to account for a) multiple timescales of perceptual integration, b) the variation

of time scales according to kind of stimulus and c) the multimodal sensory processing observed [White, 2018], the time-frame phenomena may correspond to local processings. Further mechanistic explanations probably convey a combination of continuous and dynamical discrete interactions [Herzog et al., 2016, Kozma and Freeman, 2017].

In line, recent experiments recognize breaks of brain dynamics in conscious perception [Baria et al., 2017]. Transient neural dynamic associated with acceleration of brain repertoire appears during visual conscious perception, in contrast with the opposite unseen condition. The speed of population activity, measured as a point trajectory in the state space versus time (ms), shows an acceleration activity and switch in dynamics after stimulus onset, with a clear peak around 400 ms. Using this technique, conscious stimuli perception was predicted from the activity up to 1 second before stimulus onset [Baria et al., 2017].

The mechanistic explanation for part of those observations is the sustained activity in recurrent cortical loops, mainly ignited by the frontal and prefrontal cortex. Experiments in monkeys provide evidence for strong sustained activity in the prefrontal cortex linked with reported stimuli, while weaker and decayed activity in the same region was associated with unreported stimuli [Van Vugt et al., 2018]. In these experiments, the signal propagation was tracked using microelectrodes in V1, V4 and dorsolateral prefrontal cortex (dlPFC) (Figure 3.5C). Signals related to unreported stimuli were lost between V1 and V4 and/or V4 and dlPFC, preventing reaching the frontal cortex. Moreover, the unreported activity was correlated with lack on feedforward as well as feedback activity from the prefrontal cortex, supporting the transient ignition and posterior broadcasting from prefrontal cortex as a mechanism of conscious access [Dehaene and Changeux, 2005, Mashour et al., 2020].

Unfortunately, these results give no light either regarding the continuous versus discrete debates on the mechanisms of access consciousness, or how lower-level temporal units form higher-level conscious ”continuous present moments”.

3.6.5 Consciousness and Cognition

Dual system theories of cognition argue in favour of two cognitive systems, one intuitive fast decision-making system and another rational slow system [Gilovich et al., 2002, Kahneman, 2003]. Evidence from cognitive science, behavioural economics, concept combinations, hu-

man judgment and decision making support those claims [Kahneman, 2003]. Under the first system, human thinking tends to be biased and "easily wrong" [Gilovich et al., 2002, Ellsberg, 1961, Moore, 2002, Machina, 2009]. This system is thought to work under uncertainty and its triggered behaviour following non-classical probability laws [Busemeyer and Bruza, 2012, Pothos and Busemeyer, 2013, Wang et al., 2014, Bruza et al., 2015]. The second system conveys extra analytical operations and more reliable behaviours.

These cognitive systems also find connections with different conscious processing. Different types of cognition are expected according to different modes of consciousness [Shea and Frith, 2016, Dehaene et al., 2017, Signorelli, 2018a]. Unconscious processes generate automatic responses in a first cognitive level, while conscious processes are differentiated into two other cognitive types. One of them is triggered by the awareness of contents (knowing something), i.e. perceive or becoming aware of something. Another corresponds to the awareness about the processing or manipulation of content (know that I know). The former conscious process is sometimes called global availability [Dehaene et al., 2017], associated with access consciousness and a fast cognitive system. The latter is related to slow cognition and confidence reports. It is sometimes called self-monitoring [Dehaene et al., 2017], or as we prefer, self-reference [Varela, 1975, Signorelli, 2018a].

These connections between conscious processing and cognitive systems bring another complex dimension regarding brain organization, conscious interactions and cognition.

3.6.6 Entangling Mechanisms

The neural correlates of consciousness (NCCs) described through this section have inspired the conceptual division of phenomenal consciousness and access consciousness [Block, 1995, Block, 2005]. The former defines the experience itself, the "what is like to be" in such experience [Thomas Nagel, 1974], namely, the qualitative elements of experience that allow us to distinguish between two different experiences. The latter corresponds to the explicit access to the content of that experience, also involving other cognitive systems. Additionally, one may unravel other distinctions: i) the ancient concept of pure awareness, i.e. awareness of the experience itself, ii) awareness of the content or awareness that one knows, iii) conscious access, the access of content of consciousness by cognition, iv) conscious processing, operations applied to conscious content and v) self-monitoring, the awareness about the processing

or manipulation of contents [Dehaene et al., 2017]. Further distinctions may include the neural correlates as prerequisites for conscious experiences (NCC-pr), content-specific NCC, and neural consequences of conscious experience (NCC-co) [Aru et al., 2012, Storm et al., 2017].

Though the distinctions are debatable [Naccache, 2018], they serve a pragmatic purpose: explain the disparate neural correlates discussed above. The non-report and dreaming paradigms that point out to posterior parietal-occipital regions would correspond to phenomenal consciousness [Siclari et al., 2017, Tsuchiya et al., 2015], while experiments in conscious perception suggesting prefrontal-parietal networks would correspond to access consciousness [Van Vugt et al., 2018, Mashour et al., 2020].

In those terms, phenomenal awareness may relate to the richness of dynamical resting brain activity, while conscious access to the acceleration of this characteristic awake brain dynamic [Baria et al., 2017]. One transition seems to be graded, while the other an all-or-nothing one. Unfortunately, no current neural mechanism dynamically accounts for all those, either conceptual or empirical processes of consciousness.

Our framework integrates this evidence through layer interactions and their dynamical transitions (Figure 3.5A-C). The parallel layer configuration corresponds to unconscious processing, the default activity observed in deep sleep, characterized by slow-wave cortical activity and decoupling of DMNs. Later and following intrinsic layer cycles, some of these hypothetical layers start to interact with, or increase their interactions. First transitions construct the mentioned subjective frame [Park and Tallon-Baudry, 2014], they correlate with first signals of arousal and waking. At this level, there is no conscious perception, but implicit contents related to body signals. These interactions may eventually correlate with NREM dreams. Gradually, further layers begin to interact, involving bodily and environmental signals. These transitions co-arise with REM dreams. After these transitions, the brain dynamic sets on the known resting-state dynamic and the first groups of layers interacting become the unspecific content background for further richer conscious experiences. Criticality is then the dynamical scenario realized by these layers dynamically interacting. Subsequent dynamical transitions or dynamical overlapping regions become the access consciousness of more specific content experience. As an example of this reconfiguration, the dynamical light blue layer moves from posterior cortical zones to middle and prefrontal regions in figure 3.5C. Finally, the phenomenal and access consciousness combined give rise to rich layer dynamic from which content of experience is co-defined (Figure 3.5D). In this sense, phenomenal awareness, construction of

content and access consciousness are different dynamics of one and the same process: the layer interaction at different levels and transitions.

Furthermore, the time-frame phenomena observed during conscious perception [White, 2018] corresponds to local processing layers with intrinsic time scale dynamics, involving the ”multiple perceptual cycles, in distinct brain networks, with different periodicities” suggested by [VanRullen, 2016]. These layers may correspond to different modalities and present continuous activity, but once they start to interact, their continuous processing is interfered by the other layers, generating phase transitions between disordered and ordered activity at different intervals [Kozma and Freeman, 2017]. This dynamical interaction generates variable cyclical periods of high amplitude followed by short periods of low analytical power. The first, representing coordinated layers, and the second, uncoordinated layers activity. These breaks or dynamical interference sustain the awareness background, while high amplitude sustained activity conveys metastable states of experienced content. These ideas may be related to the simulation in [McComas and Cupido, 1999], where different cortical layers interact to integrate activities across time scales and generate a unified percept.

This framework also integrates cognitive systems, stages processing and types of conscious processing (Figure 3.6A). First-stage processing is associated with the intrinsic layer dynamic given by parallel semi-independent layer configuration. A second stage is given by layers interacting through inter interactions and interference with intralayer dynamics. These two stages together define a first cognitive system associated with access consciousness. The recursive processing of access consciousness within the same contents will later correlate with a second cognitive system and self-referential or monitoring awareness. During unconscious modes, inter and intralayer interactions do not interfere with each other, or they are directly decoupled. This lack of inter and intra process interference is represented by continuous lines in figure 3.6B. Cognitive processes in this model range from automatic kinesthetic responses to fast recognition [Dehaene et al., 2017]. Then, aware cognitive systems rely on inter and intralayer interference, represented by dot lines in figure 3.6B. As these interactions move from layers to layers, they generate different sustained activity and conscious cognitive processes associated.

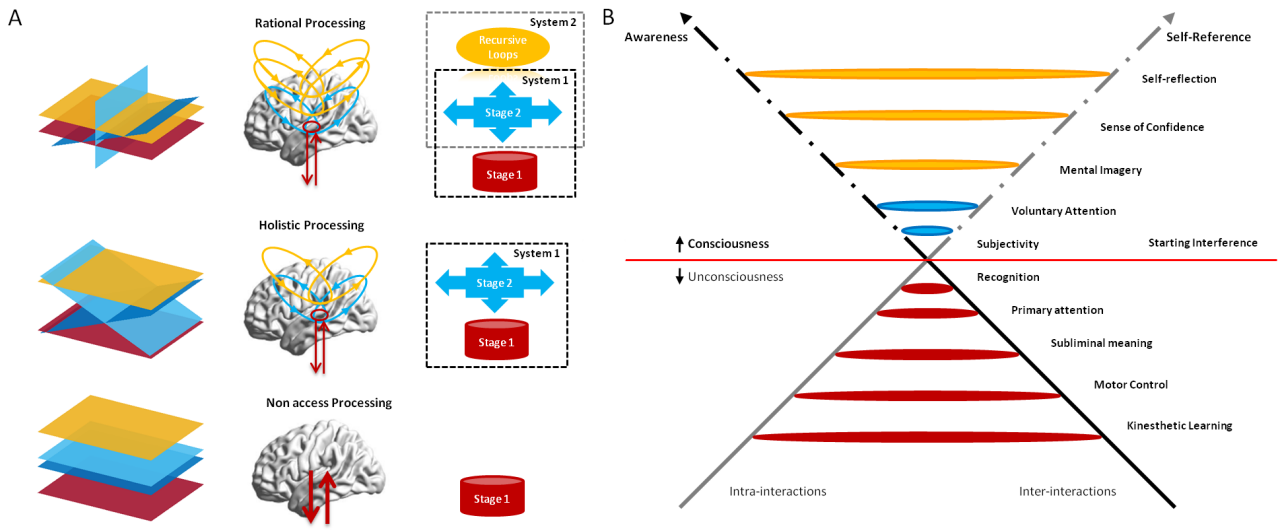


Figure 3.6: Entangling systems. (**A**) Multilayer configurations, types of cognitive systems and their relation with stages of conscious processing. From bottom to top: i) stage 1 corresponds to automatic and non-conscious processes (classical information) described by non-interacting layers. ii) Stage 2 is related to phenomenal awareness and conscious perception as holistic information. Here, the phenomenology of conscious experience co-arise with the number of layers interacting and types of interactions. Together, stage 1 and 2 form a non-classical complex system 1. iii) Recursive loops of stage 2 correspond to conscious manipulation or processing of contents. This is called self-monitoring [Dehaene et al., 2017], self-reference [Signorelli, 2018a], or meta-cognition. These recursive loops and re-entry form a second system 2, recovering classical characteristics. (**B**) The layer interaction is seen as a combination of intra-interactions inside layers and inter-interactions among layers. Unconscious processing corresponds to independent layer processing or lack of interference between intra and inter interactions (continuous line), while conscious processing conveys interference between layer activities (dot lines). As an example, we show some cognitive processes associated with unconscious and conscious conditions.

3.7 Conclusions

Through these pages, we introduced a multilayer framework for consciousness and the evidence supporting it. In this framework, consciousness co-arise with entangled mechanisms of brain-body interactions. These mechanisms compound a new irreducible whole-system, given by the inter-interactions that become intra-interactions under conscious conditions. This new whole-system is not reduced to the relationships between these mechanisms but co-defined by the whole process of interconnected layers [Signorelli and Meling, 2021].

As such, our framework seeks to integrate different signatures of consciousness. Multiple brain-body signatures were discussed inside of the main groups: global modes of consciousness and specific conscious experiences [Bachmann and Hudetz, 2014, Storm et al., 2017]. In our conceptual model, different distinctions such as pure awareness, phenomenal consciousness, access consciousness [Block, 1995, Block, 2005], among others [Dehaene et al., 2014, Dehaene et al., 2017, Signorelli, 2018a], corresponds to dynamical process configurations of layers interacting in a multilayer architecture. In such a framework, no single area is necessary and suf-

ficient for consciousness, but their complex interactions may be. For example, the brainstem, thalamus and other regions involving brain-body interactions might provide the background for consciousness, while others, such as cortical zones, become candidates for content specific conscious experience.

Some of these layers are brain networks, and molecular systems connecting the otherwise anatomically disconnected regions, but future empirical definitions might bring light to the optimal layer division regarding conscious experience. In future research, the suggested multi-structure approach may also attempt to respond to commonalities among different models of consciousness, such as global neuronal workspace model (GNW), integrated information model (IIT), dynamical systems theory, and embodiment models of conscious experience, among others. This might be important in order to consider all the elements that seem relevant for a sound theoretical and empirical framework.

Finally, our framework presents several advantages: i) it integrates different dimensions of interactions, accounting for the embodiment of consciousness [Thompson and Varela, 2001], ii) it is a dynamic approach, accounting for the rich spatiotemporal structure of consciousness [Deco et al., 2015, Deco et al., 2017a, Ipiña et al., 2020], iii) it is pragmatic framework already applied to relevant data, giving us new insight on the complex intertwined brain-body systems and their relationship with consciousness [Deco et al., 2018b, Kringelbach et al., 2020], iv) it offers simple concepts to reason about dynamical couplings and relevant systems [Signorelli and Joaquin Diaz Boils, 2020], inspiring new multidimensional experimental paradigms and new data set require to test the major hypotheses, v) it comes with a concrete mathematics, from which further perspectives may exploit their current implementations in other fields [Boccaletti et al., 2014], and finally vi) the semantic of multilayer networks interacting has the potential to integrate other theories which are based on, implicitly or explicitly, graph and network theory.

Chapter 4

INTRINSIC IGNITION AND ITS SIGNATURES

Anesthesia induces a reconfiguration of the repertoire of functional brain states leading to a high function-structure similarity. However, it is unclear how these functional changes lead to loss of consciousness. Here we suggest that the mechanism of conscious access is related to a general dynamical rearrangement of the intrinsic hierarchical organization of the cortex. To measure cortical hierarchy, we applied the Intrinsic Ignition analysis to resting-state fMRI data acquired in awake and anesthetized macaques. Our results reveal the existence of spatial and temporal hierarchical differences of neural activity within the macaque cortex, with a strong modulation by the depth of anesthesia and the employed anesthetic agent. Higher values of Intrinsic Ignition correspond to rich and flexible brain dynamics whereas lower values correspond to poor and rigid, structurally driven brain dynamics. Moreover, spatial and temporal hierarchical dimensions are disrupted in a different manner, involving different hierarchical brain networks. All together suggest that disruption of brain hierarchy is a new signature of consciousness loss.

4.1 Introduction

Recent studies suggest dynamical disruptions on brain activity during general anesthesia, sleep and disorders of consciousness [Dehaene and Changeux, 2011, Mashour et al., 2020]. Nevertheless, whatever the level of consciousness, the resting-state brain activity displays highly organized coherent networks [Biswal et al., 1995, Buckner et al., 2008, Fox et al., 2005, Fransson, 2006, Vincent et al., 2007]. Examples are anticorrelated networks still present under anes-

thesia [Boveroux et al., 2010, Vincent et al., 2007] and early stages of sleep [Fukunaga et al., 2006, Picchioni et al., 2008]. Evidence suggests that anesthesia modulates the strength of functional connectivity [Martuzzi et al., 2010, Barttfeld et al., 2015, Boveroux et al., 2010, Schrouff et al., 2011]. In Barttfeld et al. [Barttfeld et al., 2015], dynamical resting-state analyses of functional magnetic resonance imaging (fMRI) data acquired in awake and propofol anesthetized macaques, indicate that during the awake state, the brain activity at rest displays a rich repertoire of flexible functional patterns that is independent of the underlying anatomical connectivity. Conversely, during anesthesia-induced loss of consciousness, the resting-state brain activity is shifted toward a poor repertoire of rigid functional patterns with higher similarity to structural connectivity. A finding that was generalized to different anesthetic agents [Uhrig et al., 2018] and also applied to classify different categories of chronic loss of consciousness [Demertzi et al., 2019]. This dynamical disruption at long-distance networks might be the common fingerprint of all different types of loss of consciousness (anesthesia-induced, injuries-induced loss of consciousness and sleep).

Unfortunately, it is still unclear how these functional disruptions are induced and if they are causal or consequence of other factors. We hypothesised these dynamical disruptions are due to the breakdown of the hierarchical organization of the cortex and cortico-sub-cortical networks [Mesulam, 1998]. Independently of the molecular pathways of different anesthetics, stages of sleep or localization/types of brain injuries, if this is enough to disturb the hierarchical structure of the conscious brain, it will lead to a loss of consciousness. Differently than previous hierarchical auditory regularities studied in awake and anesthetized macaques [Bekinschtein et al., 2009, Uhrig et al., 2016], the causal driven forces that generate similar global disruptions would correspond to any local or global disturbance with enough power to reorganize the network hierarchy. These disruptions become a common signature for loss of consciousness, at the same time as saving the specificity of different impairments [Sherrington, 1906].

To investigate the brain mechanisms of consciousness loss, a newly introduced measure called Intrinsic Ignition, together with general anesthesia, offer a unique opportunity to quantify the hierarchy of neural activity and its disruptions [Deco and Kringelbach, 2017]. On the one hand, this is possible through the massive modulation of both arousal and conscious access (i.e. awareness) by elective pharmacological drugs, called anesthetic agents. Different anesthetics, with different pharmacological and molecular pathways, generate comparable dynamical disruption [Uhrig et al., 2018]. On the other hand, Intrinsic Ignition quantifies the

neural propagation activity in space and time, from one region to other areas of the brain [Deco and Kringelbach, 2017].

Intrinsic Ignition combines simplified versions of integration from the ”integrated information theory (IIT)” [Deco et al., 2015, Tononi and Koch, 2015], and broadcasting from ”the global neuronal workspace theory (GNW)” [Dehaene and Changeux, 2011]. Moreover, Intrinsic Ignition is complementary to the concept of ignition/broadcasting from the GNW theory. The former being considered intrinsic (due to internal interactions under resting state) and the later extrinsic (due to external stimuli) [Van Vugt et al., 2018], however, in its original formulation, ignition was also related to spontaneous neural activity [Dehaene and Changeux, 2005]. Intrinsic ignition might become a middle concept to unify ideas about phenomenal consciousness and access consciousness. Broadcasting, as the first stage of neural processing and integration as the second stage, are combined into one measure of brain activity [Deco and Kringelbach, 2017]. Intrinsic Ignition uses the graph theory to define integration. Integration is the accumulative and averaged value of the maximal path in a network at different states, computing the value among spatial areas and time evolution. This measure quantifies different modes of consciousness and estimates the type of hierarchical organization for different conditions. At spontaneous waking brain activity, analyses using Intrinsic Ignition suggest that the brain organization is maximally hierarchical, but not uniformly graded [Deco and Kringelbach, 2017, Deco et al., 2017b]. Intrinsic Ignition has been applied to compare awake versus sleep conditions [Deco and Kringelbach, 2017, Deco et al., 2017b] and normal subjects versus meditators [Escrichs et al., 2019].

Here, we measured Intrinsic Ignition of cortical areas from awake and anesthetized macaques, with unique access to six experimental conditions (awake, ketamine, light/deep propofol, light/deep sevoflurane anesthesia) [Barttfeld et al., 2015, Uhrig et al., 2018]. Intrinsic Ignition of cortical areas assessed with fMRI reveals spatial-temporal hierarchical differences and allows clustering anaesthetics. Finally, Intrinsic Ignition, as a unifying concept across theoretical frameworks of consciousness, quantifies the dynamical disruption in terms of hierarchical organization arrangement and defines a multidimensional signature of consciousness.

4.2 Results

4.2.1 Dynamical differences among Anesthetics

The analyzed data here corresponds to 119 runs. An example of the time series for one subject in the awake condition is plotted in Figure 1a. The probability density (density distribution) of fMRI values is plotted after a normalization procedure (using z-score). These plots suggest the use of non-parametric statistical tests since the center of the data distributions present close to zero mean, but seemingly different variance. The Kolmogorov-Smirnov test finds statistical differences among all conditions ($p < 0.001$, confidence intervals (CI) reported in Captions Figure 4.1b). The FC matrix plotted as the average among subjects also supports dynamical differences among conditions (Figure 4.1c).

These differences also appear when plotting the dynamical functional connectivity (dFC) (Figure 4.1d). Awake (followed by light propofol and light sevoflurane sedation) seems to present more correlated activity among functional matrices across time than the deep anesthesia conditions. Deep propofol anesthesia is slightly more activated than ketamine and deep sevoflurane. To quantify these qualitative differences a Pearson correlation between the FC per subject and the structural connectivity (SC, CoCoMac) is performed (Figure 4.1e). It reveals that the awake state has a lower correlation value, indicating that functional activity is farther from SC, than other conditions (Kolmogorov-Smirnov test, $p < 0.001$). The violin plots in Figure 1e also demonstrate differences in the type of distributions. Light propofol sedation has the highest mean correlation value, however, anesthetics are not differentiated in terms of statistical tests ($p > 0.01$). Quantifying the dynamical variability through metastability, as the standard deviation of the Kuramoto's order parameter (synchrony), shows the awake condition with higher values of metastability (Figure 4.1f, Kolmogorov-Smirnov test, $p < 0.001$), light propofol and light sevoflurane sedation slightly higher than ketamine and deep sedations, but not statistically significant. Deep propofol anesthesia has a slightly lower value than ketamine anesthesia (Kolmogorov-Smirnov test, $p = 0.01$), as well as light propofol ($p < 0.001$) and light sevoflurane anesthesia ($p = 0.0011$). Other conditions do not present major statistical differences in terms of metastability ($p > 0.01$).

These dynamical analyses support and replicate previous results suggesting disruption of the dynamical functional organization under anesthesia [Barttfeld et al., 2015, Uhrig et al.,

2018]. However, they do not distinguish among anesthetics or quantify the degree of disruption in terms of hierarchical organization. Therefore, extra analyses may offer a complementary picture.

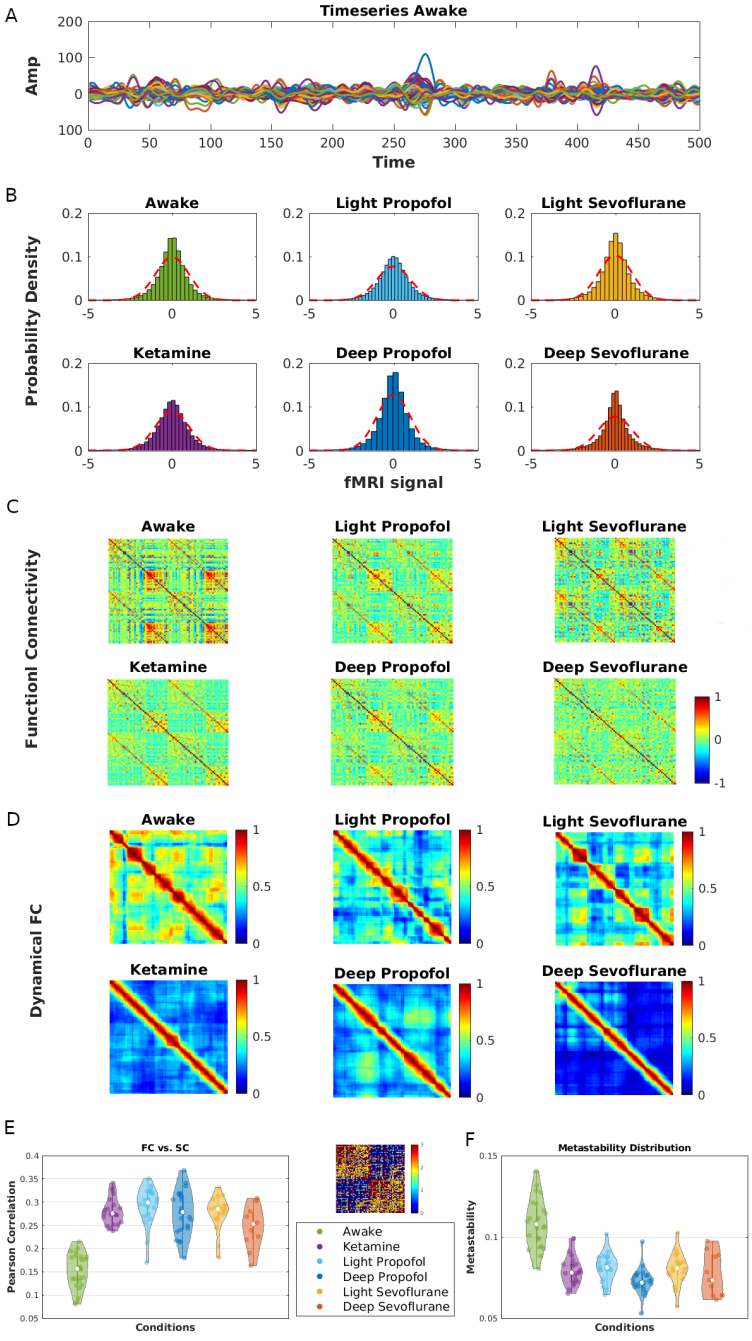


Figure 4.1: Dynamical Analysis. (**A**) Example of time series for one monkey, awake condition, 500-time points with a 2400 ms repetition time (TR). (**B**) Distribution plots of the fMRI signal for each condition. All conditions are significantly different (Kolmogorov-Smirnov test $p < 0.001$, awake CI , $[0.0062 - 0.0332]$, ketamine $[0.0027, -0.0021]$, light propofol CI $[0.0025, -0.0026]$, deep propofol CI $[0.0033, -0.0029]$, light sevoflurane CI $[0.0032, -0.0058]$, deep sevoflurane anesthesia CI $[0.0061, -0.0057]$). (**C**) Functional connectivity matrices (FC) for each condition, CoCoMac, 82 cortical regions. (**D**) Example of dynamical functional connectivity (dFC) for one subject in each condition. (**E**) Pearson Correlation between FC and Structural connectivity (SC) for each subject and plotted as a violin plot, see methods, and [Hintze and Nelson, 1998]. SC matrix is plotted alongside for reference. Awake condition is significantly different than the other conditions (Kolmogorov-Smirnov test $p < 0.001$, awake CI $[0.1671, 0.1369]$, ketamine CI $[0.2926, 0.2687]$, light propofol CI $[0.3093, 0.2738]$, deep propofol CI $[0.2957, 0.2568]$, light sevoflurane CI $[0.2943, 0.2601]$, deep sevoflurane CI $[0.2725, 0.2196]$), while other conditions are not statistically different ($p > 0.01$). (**F**) Metastability is higher in the awake state than in the anesthesia conditions (Kolmogorov-Smirnov test $p < 0.001$, awake CI $[0.1143, 0.1024]$, ketamine CI $[0.0835, 0.0759]$, light propofol CI $[0.0849, 0.0775]$, deep propofol CI $[0.0762, 0.0697]$, light sevoflurane CI $[0.0853, 0.0759]$, deep sevoflurane CI $[0.0844, 0.0687]$).

4.2.2 Intrinsic Ignition and global Hierarchical Organization

The computation of Intrinsic Ignition is based on network theory and binarization techniques (Methods and (Figure 4.2a). An example of the raster plot generated for each subject is shown in Figure 4.2b. In this case, the raster plot is calculated for one subject in six conditions. The ignition capability generates two measures: Intrinsic Ignition and Ignition Variability.

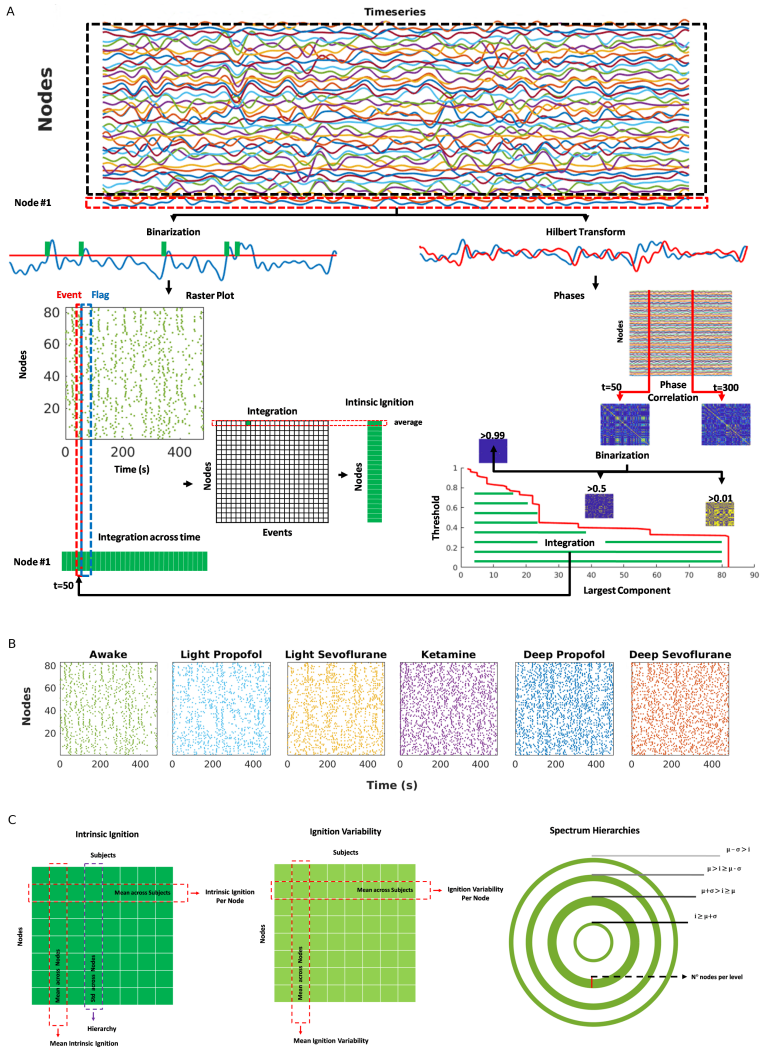


Figure 4.2: Intrinsic Ignition Measures. (**A**) From a binarization (Methods), a raster plot is generated for all nodes. In parallel, a Hilbert transform is applied to each time series, from which a phase correlation or pairwise phase synchronization between regions is defined. For each matrix generated (e.g, $t = 50$ and $t = 300$) another binarization process is applied. With the remaining connections, the largest component is computed as the level of integration given by the length of the connected component of that undirected graph. It creates a curve (lower figure) and the area below the curve (green) corresponds to the integration value for that node at that time-point. Running the same procedure for all time-points results in a table of integration across time for each node. Finally, for each event, the total integration is computed as the average across the time window, the event and a flag of 4TR, building a matrix of integration across events. The average integration across events corresponds to the Intrinsic Ignition, while the standard deviation integration across the same events is the Ignition Variability. (**B**) Example of raster plot for one subject in six conditions, 82 nodes. (**C**) The previous procedure is repeated through subjects and conditions, creating a data matrix D_{ij} for each condition, such that i corresponds to the nodes and j to the subjects. The mean across subjects (j) is the Intrinsic Ignition per node and Ignition Variability per node respectively. The mean across nodes (i) defines the Mean Intrinsic Ignition and the Mean Ignition Variability. The standard deviation of the Mean Intrinsic Ignition (intrinsic ignition across nodes) returns a quantification value of the shape of the Intrinsic Ignition curve, which is defined as the Hierarchy. Ultimately, a Spectrum Hierarchy is a circle plot with different levels. Each level corresponds to a threshold of the Intrinsic Ignition and/or Ignition Variability curve (see Methods and Supplementary Figure B.2). The thickness of each level line is the number of nodes on that level (e.g. red line marks the thickness of level two).

Plotting the Intrinsic Ignition per node values from highest to lower creates a sorted curve. The shape of that curve informs about the possible types of hierarchical organization in a network. Qualitatively, the Intrinsic Ignition per node curve for the awake condition seems to correspond to a graded non-uniform hierarchy, as previously reported [Deco and Kringelbach, 2017, Deco et al., 2017b]. While the anesthesia curves transit from graded non-uniform hierarchy to less pronounced curve slope, suggesting spatial modifications towards weak non-hierarchies (Figure 4.3a). However, a zoom on these curves indicates that the graded non-uniform nature is maintained, and what changes are the degrees of this non-uniformity (Supplementary Figure B.2). The higher values of Intrinsic Ignition are found in the awake condition, followed by light sevoflurane and light propofol, and later deep propofol, deep sevoflurane, and ketamine anesthesia. Ketamine anesthesia has a similar effect than deep propofol and deep sevoflurane anesthesia. The curves seem to differentiate at least two groups: awake and anesthesia.

Taking both the Intrinsic Ignition value and the Ignition Variability generate a scatter plot with one point per node (Figure 4.3b). Two clusters are now clearly separated; one corresponds to the awake condition (green dots) and the others to sedation conditions. The box plots, on the right side, shows the distribution of Intrinsic Ignition values across nodes in each condition. Awake values are significantly higher than other anesthetic conditions (Kolmogorov-Smirnov test $p < 0.001$, CI reported in captions Figure 4.3b), supporting the idea of maximal hierarchical organization. Values for ketamine anesthesia differentiate from light propofol, light and deep sevoflurane ($p < 0.001$) but not deep propofol anesthesia ($p > 0.01$). Light propofol anesthesia presents slightly higher values of Intrinsic Ignition per node, becoming statistically differentiated from deep propofol and deep sevoflurane anesthesia ($p < 0.001$) but not from light sevoflurane anesthesia ($p > 0.01$) which is the third-highest value after the awake condition and light propofol sedation. Intrinsic Ignition in deep propofol anesthesia is significantly lower than in light sevoflurane anesthesia ($p < 0.001$) and slightly higher than in deep sevoflurane anesthesia ($p = 0.0013$). Finally, light and deep sevoflurane anesthesia also present statistical differences as shown in the box plot. All the Intrinsic Ignition per node values across different subjects (full distribution) lead to similar conclusions (Supplementary Figure B.3a). Effect size analyses, performed to quantify these differences, also support these findings (Supplementary Figure B.4a and Supplementary Table C.1). These results support the idea that the spatial dimension regarding the hierarchical organization is disrupted differently among conditions. This disruption is classified in at least three clusters: Awake, Light and Deep sedation effects.

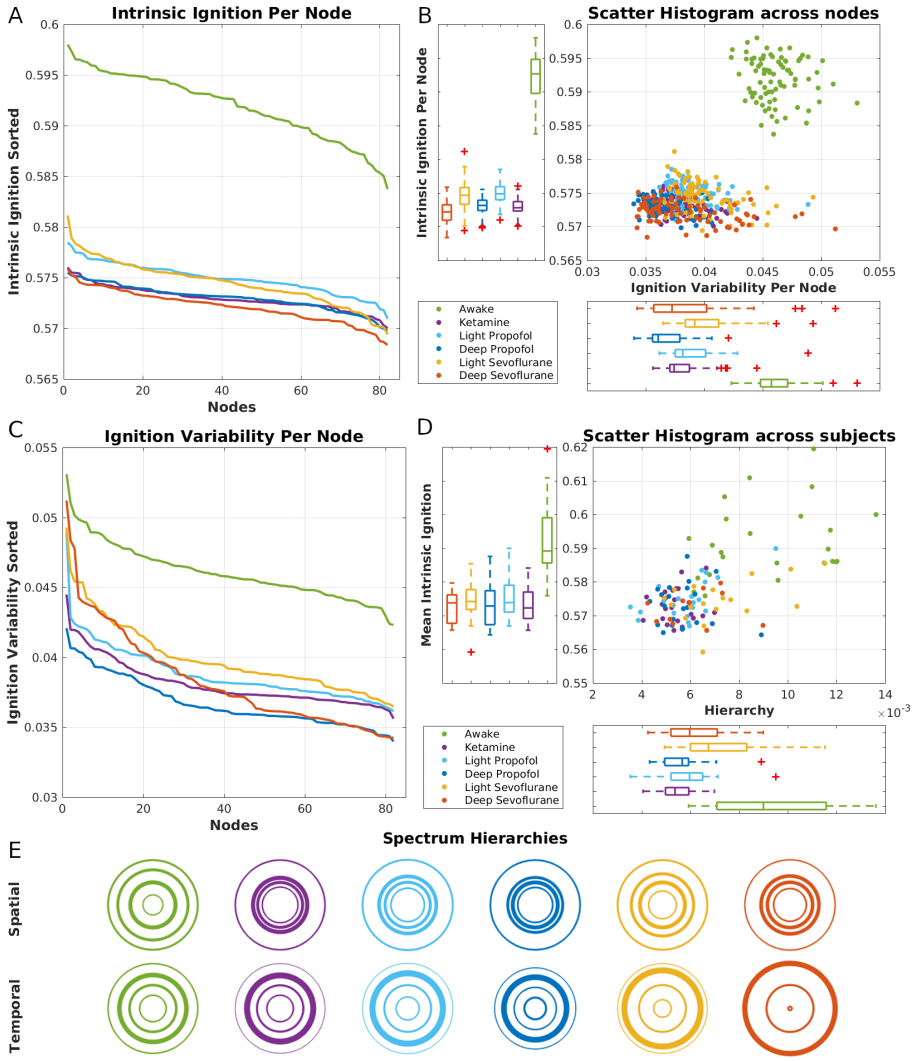


Figure 4.3: Intrinsic Ignition reveals hierarchical disruption. (**A**) The shape of ignition curves changes slightly across conditions, suggesting spatial modifications towards weak non-hierarchies. Nodes are indexed in descendent order. (**B**) Scattering plot shows two different groups, awake and anesthesia. Upper right, awake group is significantly different than other conditions (Kolmogorov-Smirnov test $p < 0.001$, awake CI $[0.5926, 0.5912]$, ketamine CI $[0.5732, 0.5727]$, light propofol CI $[0.5752, 0.5746]$, deep propofol CI $[0.5733, 0.5727]$, light sevoflurane CI $[0.5750, 0.5740]$, deep sevoflurane CI $[0.5725, 0.5718]$). Bottom left, awake is again differentiated from the others conditions (Kolmogorov-Smirnov test $p < 0.001$, awake CI $[0.0465, 0.0456]$, ketamine CI $[0.0384, 0.0377]$, light propofol CI $[0.0392, 0.0384]$, deep propofol CI $[0.0371, 0.0363]$, light sevoflurane CI $[0.0405, 0.0394]$, deep sevoflurane CI $[0.03910.0376]$). (**C**) Ignition Variability curves suggest more intricate ways to disrupt the temporal organization. (**D**) Scatter histogram. Awake condition is significantly different than sedations (upper right, Kolmogorov-Smirnov test $p < 0.001$, awake CI $[0.5963, 0.5875]$, ketamine CI $[0.5750, 0.5708]$, light propofol CI $[0.5774, 0.5724]$, deep propofol CI $[0.5756, 0.5705]$, light sevoflurane CI $[0.5773, 0.5716]$, deep sevoflurane CI $[0.5750, 0.5693]$), as well as across the Hierarchy (bottom left, Kolmogorov-Smirnov test $p < 0.005$, awake CI $[0.0102, 0.0083]$, ketamine CI $[0.0058, 0.0052]$, light propofol CI $[0.0064, 0.0053]$, deep propofol CI $[0.0061, 0.0052]$, deep sevoflurane CI $[0.0070, 0.0053]$), with the exception of light sevoflurane ($p = 0.08$, light sevoflurane CI $[0.0081, 0.0064]$). (**E**) Spectrum Hierarchies, spatial and temporal. Different changes suggest that disturbing one of the two dimensions of the hierarchical organization may be enough to cause loss of consciousness. For each Box Plot, the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The whiskers extend the extreme values without outliers, while outliers are marked as a red cross. The center indicates the median.

In terms of how these hierarchical disruptions affect the temporal dimension, the distribution of Ignition Variability per nodes across conditions is also plotted in the lower part of Figure 3b. Statistical tests suggest that the Ignition Variability values are more sensitive to the effect of each anesthetic. All the conditions are differentiated (Kolmogorov-Smirnov test $p < 0.001$) between them, with the exception of deep propofol and deep sevoflurane anesthe-

sia ($p = 0.02$). The awake condition, once again, presents the highest Ignition Variability per node value, followed by light sevoflurane, light propofol, ketamine, deep sevoflurane and deep propofol anesthesia (Figure 4.3b, c, and Supplementary Figure B.3b). The effect size analysis is also in agreement with these results (Supplementary Figure B.4b and Supplementary Table C.1). It indicates that disruption effects are bigger among the temporal dimension of hierarchical organization, especially at the moment of differentiating conditions. To look for these effects, the Ignition Variability curves are plotted in Figure 4.3c (values are sorted from highest to lowest). The shapes of the Ignition Variability curves seem to capture more complex relations. One example is the case of deep sevoflurane anesthesia, higher values are close to the values of light sevoflurane, while lower values are near deep propofol anesthesia. These results suggest more intricate ways to disrupt the temporal organization than the observed spatial network hierarchy.

One form to quantify the hierarchical disruption is by using the standard deviation of the Intrinsic Ignition curve across nodes, generating one value for each subject (Method and Figure 4.2c). This, together with the mean across nodes, defined as the Mean Intrinsic Ignition value per each subject, can characterize the spatial hierarchy for each subject. Using both values per subjects, the mean of Intrinsic Ignition and Hierarchy produces another scatter plot in Figure 4.3d. The clusters are not as evident as before, instead, the points tendency shows a correlation between Mean Intrinsic Ignition and Hierarchy values (Pearson coefficient 0.6, CI $[0.47, 0.70]$). Furthermore, in terms of the Mean Intrinsic Ignition, the awake condition is significantly different from the anesthetics (upper box plot, Kolmogorov-Smirnov test $p < 0.001$, CI reported in captions Figure 4.3d), while no anesthetics are differentiated between them (Kolmogorov-Smirnov test, $p > 0.01$). Similar analyses on the mean of Ignition Variability are found in Supplementary Figure B.5. In terms of hierarchy, the awake condition is also differentiated (bottom box plot, Kolmogorov-Smirnov test $p < 0.001$, CI reported in captions Figure 4.3d), with the exception of light sevoflurane anesthesia ($p = 0.08$, light sevoflurane). Hence, light sevoflurane anesthesia is the only anesthetic condition which seems to be distinguished from the others ($p < 0.05$), with the exception of deep sevoflurane anesthesia ($p = 0.39$).

Another form to characterize the hierarchical structure is the spectrum hierarchies of each condition (Figure 4.3e and Supplementary Figure B.2). This is a plot that takes the mean and standard deviation of the Intrinsic Ignition per node curve (Spatial, Figure 4.3e upper) and Ignition Variability per node curve (Temporal, Figure 4.3e bottom) to describe levels and number

of areas for each level. The spatial graphs show how the spectrum changes from a graded non-uniform in the awake condition to a different type of graded and non-uniformity under anesthesia. Among anesthetics, each spectrum presents non-evident visual changes from one spectrum to the other, with only slight changes on light and deep sevoflurane anesthesia (as Hierarchy tests confirmed above), suggesting again that the spatial dimension of the hierarchical organization does not change dramatically across anesthetics. Quite intriguing, if the temporal spectrum is now observed among conditions, no huge differences are perceived from awake state, ketamine, and light propofol anesthesia, but some differences appear in comparison of the awake state, deep propofol, light, and deep sevoflurane anesthesia. This seems to be the opposite tendency from the spatial spectrum. These two different types of changes on the spectrum hierarchies support the idea that disturbing one of the two dimensions (spatial or temporal) may be enough to cause loss of consciousness. It suggests more complex structural qualitative differences concerning spectrum hierarchies that need to be solved in terms of the local number of nodes per level. For example, different regions seem to take different dynamical roles across different conditions, in order to preserve part of the hierarchical organization (Supplementary Tables 2 and 3 in doi:10.1016/j.neuroimage.2020.117618). Everything together indicates that there is a hierarchical disruption between the awake condition and the anesthetics conditions, while under anesthesia the global values of hierarchy seem to correspond to a similar organization. To understand what is changing and what is not, a local analysis of the differences among nodes is needed.

4.2.3 Local hierarchical Organization

Following the results in Figure 4.3, each value of Intrinsic Ignition and Ignition Variability per node was plotted in Figures 4.4 and 4.5 to show how local differences evolve across conditions. Moreover, a local analysis is performed to isolate regions that may have a bigger impact on the global changes of spatial and temporal hierarchy disruption (Figure 4.6). To find those regions, the global tendency was defined in terms of three logical propositions from previous analyses (Methods, Figure 4.2). This global tendency corresponds to higher values of Intrinsic Ignition and Ignition Variability per node during awake ($nodes \geq \mu$), middle values in light sedation ($\mu + \sigma \geq nodes$) and lower levels during deep sedation ($\mu \geq nodes$). Each node was classified in terms of the spectrum hierarchies levels (Supplementary Tables 2 and 3 in doi:10.1016/j.neuroimage.2020.117618). Nodes satisfying the logical proposition per condi-

tion were defined as a potential area to follow/drive the global tendency observed (red indexes in Supplementary Tables 2 and 3 in doi:10.1016/j.neuroimage.2020.117618). To visualize the areas and their changes, all values were rescaled/normalized, taking the lower value as zero and the highest as 1 (other normalizations present similar visualizations, data not shown). In Figures 4.4 and 4.5, results for Intrinsic Ignition per node and Ignition Variability are plotted respectively. According to this intuitive classification/level method, regions that follow the global tendency are distinguished from others. In the case of the Intrinsic Ignition, these nodes are the right subgenual cingulate cortex, the right posterior cingulate cortex, the right inferior parietal cortex, the right intraparietal cortex, the right frontal eye field, the left parahippocampal cortex, the left subgenual cingulate cortex, the left primary somatosensory cortex, the left intraparietal cortex and the left superior parietal cortex. For Intrinsic Variability, the regions are the right temporal polar, the right central temporal cortex, the right subgenual cingulate cortex, and the left dorsomedial prefrontal cortex. In Figures 4.4 and 4.5, these regions are signaled by red arrows (names indexes reported in Supplementary Table C.2).

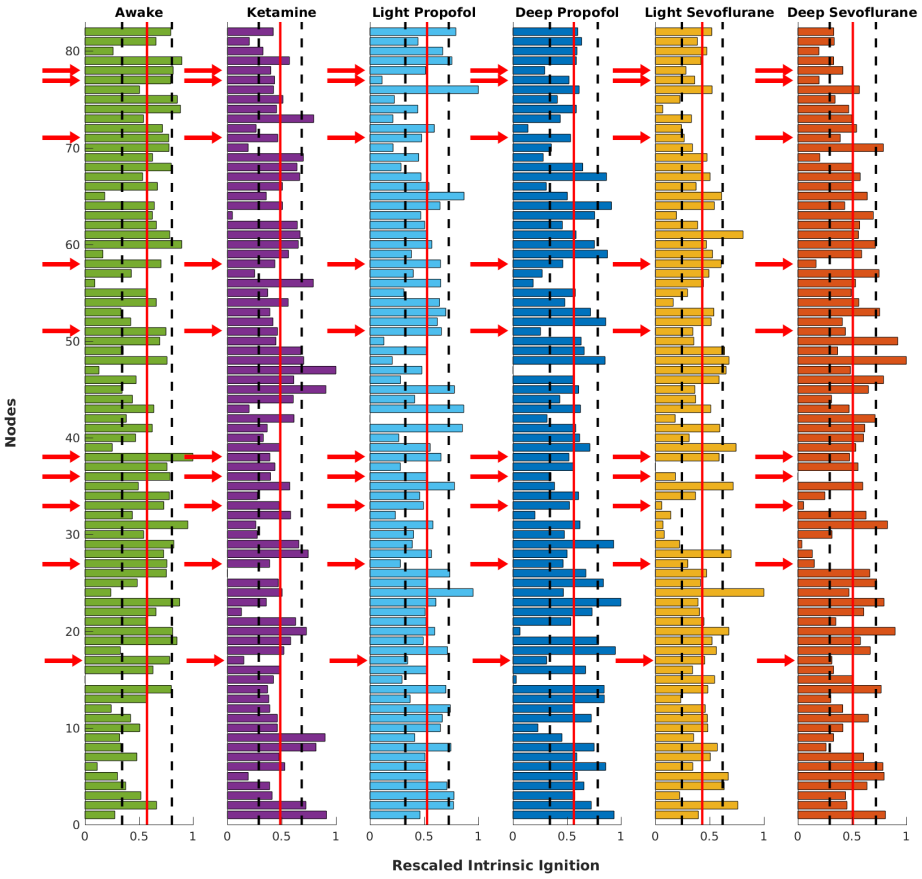


Figure 4.4: Local tendency of Intrinsic Ignition among conditions. The absolute value of Intrinsic Ignition per node was rescaled from zero (minimal value) to one (max value), allowing a visual comparison across nodes and conditions. The red vertical line corresponds to the mean value of the distribution (μ) and black vertical lines are the limits given by $\mu + \sigma$, with σ the standard deviation for each condition. The global tendency was translated into three logical propositions: $nodes \geq \mu$ for awake, $\mu + \sigma \geq nodes$ for light conditions and nodes $\mu \geq nodes$ for deep conditions. Only 10 regions (Subgenual cingulate cortex right, Posterior cingulate cortex right, Inferior parietal cortex right, Intraparietal cortex right, Frontal eye field right, Parahippocampal cortex left, Subgenual cingulate cortex left, Primary somatosensory cortex left, Intraparietal cortex left, Superior parietal cortex left) satisfied the three propositions simultaneously (Supplementary Table 2 in doi:10.1016/j.neuroimage.2020.117618),becoming candidates for areas which follow the global tendency of Intrinsic Ignition changes. These regions are signaled by red arrows and correspond to the indexes 17, 27, 33, 36, 38, 51, 58, 71, 77, and 78 respectively (Supplementary Table 3 in doi:10.1016/j.neuroimage.2020.117618).

Supplementary analyses were performed using the median instead of the mean as a cut off on the logical propositions (plots available by request). It slightly changes the results for Intrinsic Ignition. In the case of Ignition Variability, using the median instead of the mean gives no results. Finally, to ensure consistency, the same analyses were performed subject by subject (see Methods), generating a histogram of occurrences. Nodes that appear to satisfy the logical propositions per conditions above the 60% of the time were identified. These regions are the right subgenual cingulate cortex and the right intraparietal cortex for Intrinsic Ignition (Supplementary Figure B.6b), and only the right central temporal cortex, for Ignition Variability (Supplementary Figure B.6b).

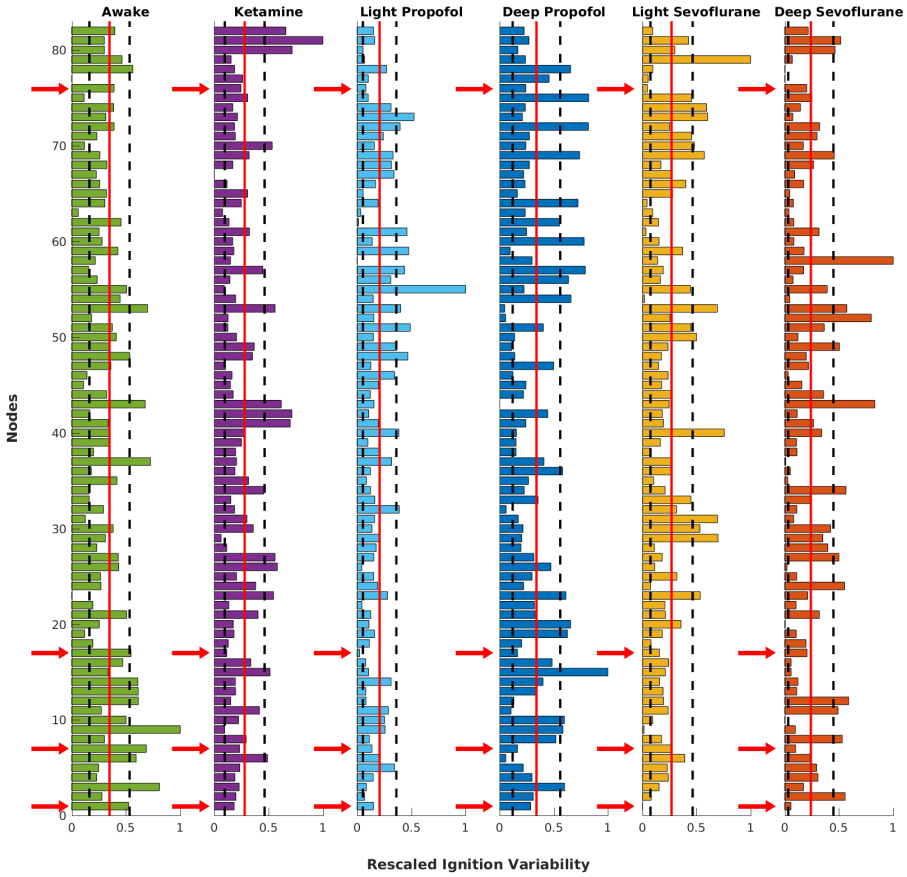


Figure 4.5: Local tendency Ignition Variability among conditions. The red vertical line corresponds to the mean value of the distribution (μ) and black vertical lines are the limits given by $\mu + \sigma$, with σ the standard deviation for each condition. The global tendency was translated into three logical propositions: $nodes \geq \mu$ for awake, $\mu + \sigma \geq nodes$ for light conditions and nodes $\mu \geq nodes$ for deep conditions. Only 4 regions (Tempolar polar right, Central temporal cortex right, Subgenual cingulate cortex right, Dorsomedial prefrontal cortex left) survived the three propositions simultaneously (Supplementary Tables 3 in doi:10.1016/j.neuroimage.2020.117618), becoming candidates for areas which follow the global tendency of ignition variability changes. These regions are signaled by red arrows and correspond to the indexes 1, 7, 17, and 76 respectively (Supplementary Table 3 in doi:10.1016/j.neuroimage.2020.117618).

4.3 Discussion

In this study, we analyzed resting-state fMRI acquired from non-human primates in the awake state and during anesthesia-induced loss of consciousness using distinct pharmacological agents. By applying Intrinsic Ignition measurements, we demonstrate that loss of consciousness is paralleled by a disruption of brain hierarchy, making it both a new signature of consciousness and consciousness loss.

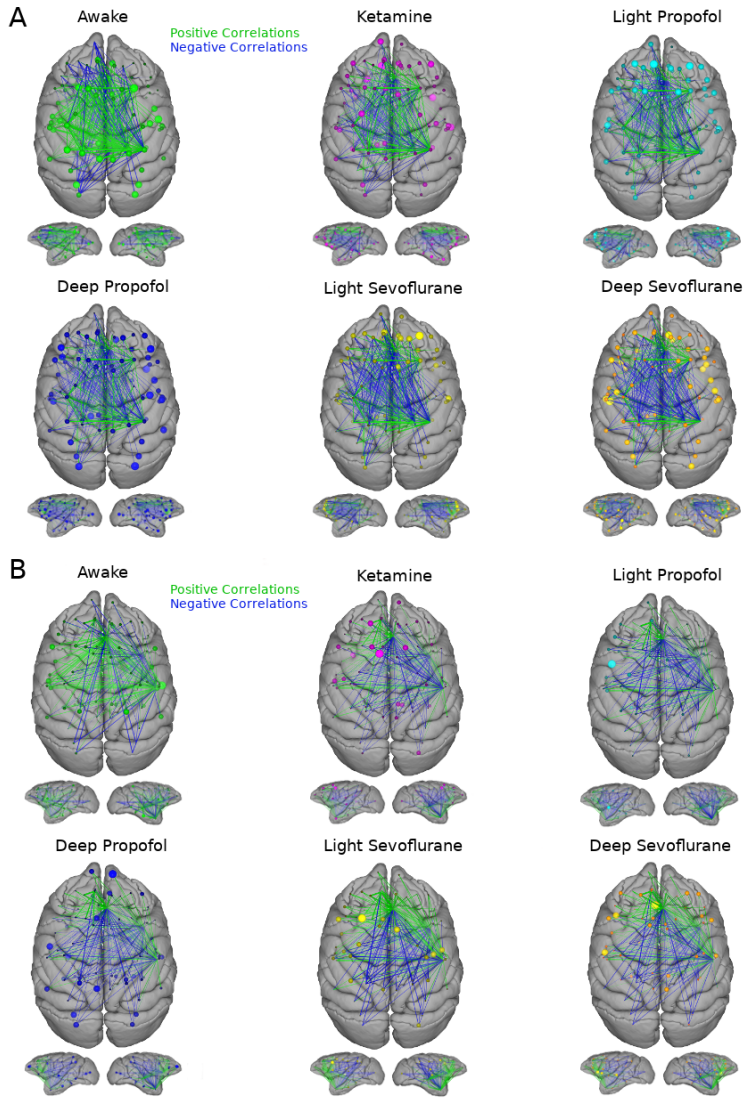


Figure 4.6: Global and local aspects of the brain functional network. A graphical way to observe how local and global disruption are intrinsically connected but probably differently driven under dissimilar anesthetics. (**A**) Intrinsic Ignition driven network. For each condition and node, the values of Intrinsic Ignition are plotted as the size of the node, while the links correspond to the functional correlation only for the nodes identified as relevant from Figure 4.4. Changes among both sizes as well as the number of links at different anesthetics are observed, suggesting interdependency between the local (nodes) and global (network) changes. (**B**) Similar plots for Ignition Variability driven network with the functional correlation only for the relevant nodes from Figure 4.5. In this case, changes among node sizes seem more relevant than among the number of links, suggesting that Ignition Variability may be more sensitive to these local (node) changes.

At the neuronal level, anesthetic agents act through different molecular pathways [Hudetz and Mashour, 2016], such as γ-aminobutyric acid (GABA) receptors, N-methyl-D-aspartate (NMDA) receptors and K+ channels [Franks, 2008] to induce loss of consciousness. At the system level, anesthetics strongly affect brain networks that are involved in arousal and awareness. Propofol [Bonhomme et al., 2001], sevoflurane [Kaisti et al., 2003] and xenon [Laitio et al., 2007] inhibit midbrain areas, associated with the ascending reticular system, the thalamus and cortical areas such as the precuneus, posterior cingulate cortex, and the prefrontal cortex. Moreover, studies with ketamine, propofol, and sevoflurane [Lee et al., 2013, Uhrig

et al., 2018] reported disruption of frontoparietal activity despite their distinct neurophysiology. Ketamine and sevoflurane have opposite effects on thalamocortical connectivity. Under ketamine, an NMDA receptors antagonist [Anis et al., 1983], thalamocortical functional correlations seems to be preserved [Bonhomme et al., 2016], while sevoflurane, a γ-aminobutyric acid receptor type A agonist and NMDA antagonist [Wu et al., 1996], induces a decrease in thalamocortical functional correlations, but preserving functional correlations between thalamus and sensory cortex [Huang et al., 2014, Ranft et al., 2016]. Propofol, targeting γ-aminobutyric acid-mediated neurotransmission [Peduto et al., 1991], induces disconnections in thalamocortical circuits. Under anesthesia, the brain patterns are not uniform, some areas seem more deactivated than others, while most anesthetics, with the exception of ketamine [Langsjo et al., 2005], cause a global reduction of cerebral blood flow [Franks, 2008]. Taking this evidence, finding a common neurophysiological pathway for this different anesthetics is challenging, but might involve the disruption of thalamocortical connectivity and/or frontoparietal connectivity.

Our results indicate that Intrinsic Ignition can discriminate between the awake condition and different anesthetics, as well as between different levels of sedation (i.e. light versus deep anesthesia). Moreover, higher values of ignition are associated with richer brain dynamics, while lower values relate to structural driven dynamics, as shown in previous reports [Barttfeld et al., 2015, Uhrig et al., 2018, Demertzi et al., 2019]. Our results (Figure 4.1) are in line with these previous studies, and additionally indicate that the spatial and temporal dimensions of hierarchical organization change under anesthesia, while keeping different types of the same graded non-uniform hierarchy through all conditions. Different regions occupy different levels on those hierarchies, indicating that different anesthetic paths may indeed act differently, however, generating the same hierarchical disruption. Everything together supports the hierarchical breakdown hypothesis and places the local and global disruption of the hierarchy as a possible common signature to reconcile seemingly different types of loss of consciousness.

Reports from non-REM sleep also provide evidence for a partial breakdown of the hierarchical organization of large scale networks [Boly et al., 2012b]. In order to quantify these disruptions, Intrinsic Ignition (Figure 4.3)), was applied in two hierarchical dimensions: spatial and temporal. The awake condition in monkeys presents maximal hierarchical organization (Figure 4.3), while different concentrations of anesthetics range from middle to lower values of Intrinsic Ignition, under light and deep anesthesia (Figure 4.3b and d). Changes observed on the spatial dimension (Intrinsic Ignition per node), recognized three groups: awake, light

(propofol and sevoflurane) and deep (propofol, sevoflurane, and ketamine). On the other hand, changes observed on the temporal dimension (Intrinsic Variability per node) seems more sensitive, distinguishing among all the conditions (Figure 4.3b and c) and suggesting more complex hierarchical disruptions across time. Interestingly, light conditions presented middle values on both Intrinsic Ignition (spatial) and Ignition Variability (temporal), while ketamine was closer to deep conditions under spatial dimension but near to light conditions when compared across the temporal aspect. This suggests that the previous evidence on the dynamical disruptions is related to more complex multidimensional disruptions given by at least two dimensions: structural organization (spatial) and dynamical organization (temporal). It supports the idea that spatial and temporal hierarchies are affected differently among conditions.

The degree of Hierarchy seems disturbed among all the different sedations (with the exception of light sevoflurane). It may imply that hierarchical disruptions are indeed more complexly related to possible local mechanisms. The Hierarchy quantifier recognizes that light and deep sevoflurane anesthesia might be affecting similarly but in different degrees the hierarchical organization, in turn that estimates that light sevoflurane anesthesia may have weaker effects on that disruption and still be closer to an awake organization. If this interpretation is right, ketamine, light, and deep propofol anesthesia would break similarly the hierarchical organization, since they present similar distributions in terms of Hierarchy. Moreover, spectrum hierarchies reveal similar spatial disruptions between ketamine, light, and deep propofol anesthesia (Figure 4.3e upper) and partially similar ones in the temporal dimension (Figure 4.3e lower). It indicates that different mechanisms disrupt one or another aspect of hierarchical organizations: the spatial dimension conveys the impression that some anesthetics are distinguished but not others, while in the temporal dimension it happens among different anesthetics. Plot curves (Figure 4.3a and c) and spectrum plots (Figure 4.3e, Supplementary Figure B.2) show that graded non-uniform type of hierarchy is maintained across conditions. Disturbing one of the two dimensions may be enough to cause loss of consciousness and different anesthetics target differently these two aspects of the neural organization: ketamine and propofol would target spatial aspects, while sevoflurane would disturb temporal aspects of that configuration.

To deal with these structural differences, local analyses were performed. According to our results, organizational disruptions cannot be reduced to only global effects but also local differences (Figure 4.4, 4.5,4.6 and Supplementary Tables 2 and 3 in doi:10.1016/j.neuroimage.2020.117618). Globally, some anesthetics may act similarly in terms of concentration, such as light propofol

and light sevoflurane anesthesia, however, locally, different anesthetics may also present differences given mainly in terms of Ignition Variability curves (Figure 4.3). This is clearly noticed from the different distributions of values for Intrinsic Ignition and Ignition Variability per node presented in Figures 4.4 and 4.5. Once observed which areas followed the global tendency as candidates for driving these changes, Subgenual cingulate cortex and Intraparietal cortex presented a consistent occurrence in terms of Intrinsic Ignition, and Central temporal cortex in terms of Ignition Variability. Additionally, Posterior cingulate cortex, Inferior and Superior parietal cortex, Parahippocampal cortex, and somatosensory cortex among other prefrontal regions appeared as relevant areas for spatial aspects of ignition. Most of these regions have been previously associated with GNW [Uhrig et al., 2014, Uhrig et al., 2016, Uhrig et al., 2018]. In terms of the temporal component, Temporal polar cortex, Subgenual cingulate cortex, and Dorsomedial prefrontal cortex also seem to play a role in the global disruptions among the six conditions.

These local findings are in line with previous reports. For instance, studies on anesthesia and early stages of sleep have identified that the precuneus and lateral temporoparietal components of DMN persisted under anesthesia [Vincent et al., 2007], but the connectivity of the posterior cingulate cortex (PCC) is reduced during sedation [Kaisti et al., 2002, Greicius et al., 2008, Uhrig et al., 2018]. Moreover, studies using light propofol point out changes among PCC connectivity with other areas, such as somatomotor cortex, the anterior thalamic nuclei and the reticular activating system [Stamatakis et al., 2010]. According to our results, the PCC is following the global tendency among conditions and therefore, is a candidate for globally driving the hierarchical changes observed. An fMRI meta-analysis of resting-state activity in disorders of consciousness concluded that a reduction of activity in midline cortical (PCC, precuneus, medial temporal lobe, middle frontal lobe) and subcortical sites (bilateral medial dorsal nuclei of the thalamus) is associated with conscious impairments [Hannawi et al., 2015], with a more pronounced reduction in the vegetative state than in the minimally conscious state. Moreover, medial parietal cortex, PCC and precuneus are the first regions to reactivate when patients recover [Laureys et al., 2006]. Due to the current parcellation, our method cannot target all these areas, nevertheless, according to intrinsic ignition, some of them may be related to the global changes observed.

Our results, however, are not free of limitations. The data analyzed was with a parcellation of only 82 cortical regions of interest (CoComac) and therefore does not allow to infer the

disruption of the whole brain hierarchical organization. A subcortical parcellation and better definition for SC (although not mostly used here) are desirable for these effects [Kennedy et al., 2013]. The preprocessing pipeline can be improved [Tasserie et al., 2019] in order to avoid the extra cleaning procedure. Although our results indicate similar global hierarchical disruptions as a common signature driven by locally different re-organizations, these results need modelling and simulations to give a full answer about the casual driving disruptions. For example in [Chaudhuri et al., 2015] and [Joglekar et al., 2018], the ignition capability was explored as an inter-areal balanced amplification signal through large scale circuits, supporting ignition models of consciousness [Joglekar et al., 2018]. Moreover, temporal hierarchies naturally emerged from the heterogeneity of local networks [Chaudhuri et al., 2015], with slower prefrontal and temporal regions having a strong impact on global brain dynamic. Therefore, in order to link intrinsic ignition and mechanistic models, large scale models [Breakspear, 2017] are expected as future steps to give light on part of the neuronal mechanisms involved. It may help to connect our results with other studies on sleep [Jobst et al., 2017] and disorders of consciousness, as well as the simulation and exploration of manipulated brain states using deep brain stimulation [Saenger et al., 2017].

Another relevant question is about our understanding of hierarchy [Hilgetag and Goulas, 2020]. Hierarchical brain structures might refer to i) topological sequence of projections, ii) a gradient of structural or functional cortical features, iii) a progression of scales or iv) a laminar projection pattern. These hierarchies can be nested or non-nested [Feinberg, 2011, Northoff et al., 2011]. In the first case, the low levels are incorporated into the higher ones, forming intertwined relationships [Feinberg, 2011]. As such, higher levels on the hierarchy weakly constrain low levels. It defines an embodied system lacking a clear ”control zone”. Contrary, non-nested hierarchies convey strong control hubs. Some examples of non-nested hierarchies involve the anatomical hierarchical organization of a rich club where the brain region’s topological connectivity exceeds the connectivity given by random networks [Honey et al., 2010, van den Heuvel and Sporns, 2011, van den Heuvel and Sporns, 2013]. Another example is the dynamical core, a hypothetical functional cluster in which the regions inside are connected between them much more than with the rest of the brain under conscious processing [de Pasquale et al., 2018, Tononi and Edelman, 1998]. This dynamical core also resembles the idea of a workspace of neurons allowing conscious access. In both cases, structural or functional, hierarchy is quantified with topological measures such as degree of centrality. In our case, Intrinsic Ignition measure quantifies a type of function hierarchy. However, hierarchy is not defined by the anatomical or

functional connectivity, as in the case of the rich club and dynamical core respectively. Here, the hierarchical structure is defined by dynamical signal processing, where upper regions may trigger activity on lower ones more frequently than in the opposite direction. In this line, our findings suggest that the brain hierarchical organization is not uniformly graded [Deco and Kringelbach, 2017], contrary to dynamical core and GNW implies, but probably a more complex combination of functionally nested and anatomically non-nested hierarchies [Feinberg, 2011, Northoff et al., 2011].

Finally, the intrinsic character of our measurement raises the question of embodied and interoceptive signals [Tanabe et al., 2020, Signorelli and Meling, 2021]. Physiological fluctuations may have subtle but important cognitive effects. In this case, it is important to highlight two types of contributions, the ”first-order” and the ”second-order” contribution of physiological signals [Breakspear, 2017]. The first order (e.g. respiration and heart rate) contribute to many cofounds in the blood oxygen level-dependent (BOLD) signal. In our analyses, these contributions were avoided by our regression method and the preprocessing of the data. Regarding the second-order effects, for example, heart rate variability, it is known that some of these effects co-vary with a specific activity in the interoceptive cortex, such as the insula. Some of these spontaneous fluctuations locked to heartbeats seemed to predict and shape conscious visual detection [Park et al., 2014]. Other authors demonstrated that the interoceptive activity of heartbeats sent to the insula has a systemic effect on conscious perception, modulating exteroceptive awareness [Nguyen et al., 2016, Salomon et al., 2016]. Explicit cardiac perception also influences activity in regions such as the posterior and anterior insula, dorsal anterior cingulate, somatomotor cortices, among others, supporting interoceptive awareness [Critchley et al., 2004]. If all these effects are removed, the nonstationarities observed in neural data are also removed [Laumann et al., 2017], i.e. it might also remove true neuronal fluctuations. As such, we did not intend to remove those effects. The remaining contributions of the second-order effects might be also intrinsic to the true neuronal fluctuations and nonstationarities of a brain embodied in a physiological body. While a very interesting question, in this paper we do not intend to answer the question about all the sources of these intrinsic fluctuations, but only identify the main regions that are participating in those intrinsic interactions.

4.4 Conclusion

In conclusion, the global values of hierarchical organization indicate similar global organization (disruptions) under anesthesia, while local analyses on which areas habit hierarchical levels inform on the different ways that anesthetics affect spatial and temporal aspects of that organization. Our study provides a common brain signature of anesthesia-induced loss of consciousness beyond molecular pharmacology, also called "common anesthetic endpoint" [Hudetz and Mashour, 2016]. This is in line with the idea that disruptions in long-distance network dynamics are a common signature of anesthesia-induced loss of consciousness, but adding the breakdown of hierarchical organization and its two dimensions, space and time. The hierarchical organization is characterized by internal ignition activity, reconciling the observed common global changes (hierarchies) with different local changes (ignition power by node).

Our local results suggest that areas proposed by GNW, such as fronto-parieto-cingular networks, which underpin conscious access [Dehaene and Changeux, 2011], and regions considered by IIT as the parietal-posterior cortical zones (supporting phenomenal or subjective experience [Siclari et al., 2017]), are both participating in changes of hierarchical organization. These hierarchical changes find their common ground in cingulate and parietal regions. Under other theoretical frameworks such as temporo-spatial theory of consciousness [Northoff and Huang, 2017], embodiment theories [Varela et al., 2016] and interoceptive approaches [Tallon-Baudry et al., 2018], our findings might be discussed in light of intrinsic space and time dimensions of neural activity. In those cases, spontaneous activity given by intrinsic local activity triggering global patterns may become a common signature for consciousness. It may imply, that under the hierarchical hypothesis, these theories are complementary to each other, and approaching compatible aspects of the same conscious phenomenon [Aru et al., 2012, Block, 2005, Dehaene et al., 2014, Tagliazucchi, 2017, Northoff and Lamme, 2020, Mashour et al., 2020]. This solves in part the requirement of a desirable common and global signature to explain how brain dynamics are similarly affected under anesthetics, at the same time than recovering the specificity of affecting and modulating correlations and couplings of brain regions.

4.5 Methods

4.5.1 Animals

The acquisition of this data set is previously reported in [Barttfeld et al., 2015] and [Uhrig et al., 2018], http://links.lww.com/ALN/B756). Five rhesus macaques were included for analyses (Macaca mulatta, one male, monkey J, and four females, monkeys A, K, Ki, and R, $5-8$ kg, $8-12$ yr of age), in a total of six different arousal conditions: Awake state, ketamine, light propofol, deep propofol, light sevoflurane, and deep sevoflurane anesthesia. Three monkeys were used for each condition: Awake (monkeys A, K, and J), Ketamine (monkeys K, R and Ki), Propofol (monkeys K, R, and J), Sevoflurane (monkeys Ki, R, and J). Each Monkey had fMRI resting-state acquisitions on different days and several monkeys were scanned in more than one experimental condition. Only one monkey, monkey A was scanned in one experimental condition, the awake state. This experimental design ensures (under the limitations of current regulation), that our data set is as representative as possible. All procedures are in agreement with the European Convention for the Protection of Vertebrate Animals used for Experimental and Other Scientific Purposes (Directive 2010/63/EU) and the National Institutes of Health's Guide for the Care and Use of Laboratory Animals. Animal studies were approved by the institutional Ethical Committee (Commissariat á l'Énergie atomique et aux Énergies alternatives; Fontenay aux Roses, France; protocols CETEA $\#10-003$ and $12-086$).

4.5.2 Anesthesia Protocols

The anesthesia protocol is thoroughly described in previous studies [Barttfeld et al., 2015, Uhrig et al., 2018]. Monkeys received anesthesia either with ketamine [Uhrig et al., 2018], propofol [Barttfeld et al., 2015] or sevoflurane [Uhrig et al., 2018], with two different levels of anesthesia depth for propofol and sevoflurane anesthesia (Light and Deep). These levels were defined according to the monkey sedation scale, based on spontaneous movements and the response to external stimuli (presentation, shaking or prodding, toe pinch), and corneal reflex. For each scanning session, the clinical score was determined at the beginning and end of each scanning session, together with continuous electroencephalography monitoring [Uhrig et al., 2016].

During ketamine, deep propofol anesthesia, and deep sevoflurane anesthesia, monkeys stopped responding to all stimuli, reaching a state of general anesthesia. Monkeys were intubated and ventilated as previously described [Barttfeld et al., 2015, Uhrig et al., 2018]. Heart rate, noninvasive blood pressure, oxygen saturation, respiratory rate, end-tidal carbon dioxide, and cutaneous temperature were monitored (Maglife, Schiller, France) and recorded online (Schiller).

Ketamine was applied by intramuscular injection ($20\ mg/kg$; Virbac, France) for induction of anesthesia, followed by a continuous intravenous infusion of ketamine (15 to 16 $mg * kg - 1 * h - 1$) to maintain anesthesia. Atropine ($0.02\ mg/kg$ intramuscularly; Aguettant, France) was injected 10 min before induction, to reduce salivary and bronchial secretions. For propofol, monkeys were trained to be injected an intravenous propofol bolus (5 to 7.5 mg/kg; Fresenius Kabi, France), followed by a target-controlled infusion (Alaris PK Syringe pump, CareFusion, USA) of propofol (light propofol sedation, 3.7 to 4.0 $\mu g/ml$; deep propofol anesthesia, 5.6 to 7.2 $\mu g/ml$) based on the "Paedfusor" pharmacokinetic model (Absalom & Kenny, 2005). During sevoflurane anesthesia, monkeys received first an intramuscular injection of ketamine (20 mg/kg; Virbac) for induction, followed by sevoflurane anesthesia (light sevoflurane, sevoflurane inspiratory/expiratory, $2.2/2.1$ volume percent; deep sevoflurane, sevoflurane inspiratory/expiratory, $4.4/4.0$ volume percent; Abbott, France). Only 80 minutes after the induction the scanning sessions started to get a washout of the initial ketamine injection [Schroeder et al., 2016]. To avoid artefacts related to potential movements throughout magnetic resonance imaging acquisition, a muscle-blocking agent was coadministered (cisatracurium, $0.15\ mg/kg$ bolus intravenously, followed by continuous intravenous infusion at a rate of $0.18\ mg * kg - 1 * h - 1$; GlaxoSmithKline, France) during the ketamine and light propofol sedation sessions.

4.5.3 Functional Magnetic Resonance Imaging Data Acquisition

For the awake condition, monkeys were implanted with a magnetic resonance-compatible head post and trained to sit in the sphinx position in a primate chair [Uhrig et al., 2014]. For the awake scanning sessions, monkeys sat inside the dark magnetic resonance imaging scanner without any task and the eye position was monitored at $120Hz$ (Iscan Inc., USA). The eye-tracking was performed to make sure that the monkeys were awake during the whole scanning session and not sleeping. The eye movements were not regressed out from rfMRI data. For

the anesthesia sessions, animals were positioned in a sphinx position, mechanically ventilated, and their physiologic parameters were monitored. No eye-tracking was performed in anesthetic conditions. Before each scanning session, a contrast agent, monocrystalline iron oxide nanoparticle (Feraheme, AMAG Pharmaceuticals, USA; 10 mg/kg, intravenous), was injected into the monkey's saphenous vein [Vanduffel et al., 2001]. Monkeys were scanned at rest on a 3-Tesla horizontal scanner (Siemens Tim Trio, Germany) with a single transmit-receive surface coil customized to monkeys. Each functional scan consisted of gradient-echoplanar whole-brain images (repetition time = $2,400ms$; echo time = $20ms$; $1.5mm^3$ voxel size; 500 brain volumes per run). Monkeys were scanned with F-W phase encoding direction to avoid the major axis of distortion since no visual task was performed.

4.5.4 Functional Magnetic Resonance Imaging Preprocessing

A total of 157 functional magnetic imaging runs were acquired [Barttfeld et al., 2015, Uhrig et al., 2018]: Awake, 31 runs (monkey A, 4 runs; monkey J, 18 runs; monkey K, 9 runs), Ketamine, 25 runs (monkey K, 8 runs; monkey Ki, 7 runs; monkey R, 10 runs), Light Propofol, 25 runs (monkey J, 2 runs; monkey K, 10 runs; monkey R, 12 runs), Deep Propofol, 31 runs (monkey J, 9 runs; monkey K, 10 runs; monkey R, 12 runs), Light Sevoflurane, 25 runs (monkey J, 5 runs; monkey Ki, 10 runs; monkey R, 10 runs), Deep Sevoflurane anesthesia, 20 runs (monkey J, 2 runs; monkey Ki, 8 runs; monkey R, 11 runs). For details, check the supplementary tables for [Uhrig et al., 2018] (http://links.lww.com/ALN/B756).

Functional images were reoriented, realigned, and rigidly coregistered to the anatomical template of the monkey Montreal Neurologic Institute (Montreal, Canada) space with the use of Python programming language and Oxford Centre Functional Magnetic Resonance Imaging of the Brain Software Library software (United Kingdom, http://www.fmrib.ox.ac.uk/fsl/; accessed February 4, 2018) [Uhrig et al., 2014]. From the images, the global signal was regressed out to remove any confounding effect due to physiologic changes (e.g., respiratory or cardiac changes). If this regression is not performed, the functional connectivity artificially increases, and preliminary results do not recover the previously observed SC vs FC correlations changes under anesthetics. Voxel time series were filtered with a low-pass (0.05-Hz cutoff) and high-pass (0.0025-Hz cutoff) filters and a zero-phase fast-Fourier notch filter (0.03 Hz) to remove an artifactual pure frequency present in all the data [Barttfeld et al., 2015, Uhrig et al., 2018].

Furthermore, an extra cleaning procedure was performed to ensure the quality of the data after time-series extraction (Supplementary Figure B.1). The procedure was based on a visual inspection of the time series for all the nodes, the Fourier transform of each signal, the functional connectivity for each subject and the dynamical connectivity computed with phase correlation. Trials were kept when the row signal did not present signs of artifactual activity, functional connectivity was coherent with the average and dynamical connectivity presented consistent patterns across time.

Finally, a total of 119 runs are analyzed in subsequent sections: Awake state 24 runs, ketamine anesthesia 22 runs, light propofol anesthesia 21 runs, deep propofol anesthesia 23 runs, light sevoflurane anesthesia 18 runs, deep sevoflurane anesthesia 11 runs.

4.5.5 Anatomical Parcellation and Structural Connectivity

Anatomical (structural) data were derived from the CoCoMac 2.0 [Bakker et al., 2012] database (cocomac.g-node.org) of axonal tract-tracing studies using the Regional Map parcellation [Kötter and Wanke, 2005]. This parcellation comprises 82 cortical ROIs (41 per hemisphere; Supplementary Table C.2). Structural (i.e., anatomical) connectivity data are expressed as matrices in which the 82 cortical regions of interest are displayed in x-axis and y-axis. Each cell of the matrix represents the strength of the anatomical connection between any pair of cortical areas. The CoCoMac connectivity matrix classifies the strength of the anatomical connections as weak, moderate, or strong, codified as 1, 2, and 3, respectively [Barttfeld et al., 2015].

4.5.6 Dynamic Analyzes

Functional connectivity matrices (FC) for each condition are first computed for each subject using Pearson correlation and then averaged across subjects. Each FC has 82 cortical regions. The dynamical functional connectivity (dFC) is computed using a sliding window technique (50 TR correlation window and 5 TR sliding size). The chosen TR intervals are better suited for visualization purposes. This procedure results in 90-time partitions and for each one of them, one FC is computed. Then, the dFC is the correlation among these 90 FCs. Correlations between FC and Structural connectivity (SC) for each subject are computed with Pearson correlation and plotted as a violin plot. For violin plots, the shape describes the distribution

density, the white dot corresponds to the median, the thick inner line is the first quartile (down), and the third quartile (up). The borders are the upper and lower adjacent values [Hintze and Nelson, 1998]. Finally, metastability is computed as the standard deviation of the Kuramoto's order parameter (synchrony).

4.5.7 Intrinsic Ignition Analyzes

The ignition capability can be defined in terms of its spatial and temporal components, generating two measures: Intrinsic Ignition and Ignition Variability. This procedure generates one value for each node and subject that is later averaged to form the Intrinsic Ignition per node and Ignition Variability per node value. Therefore, Intrinsic Ignition tells us about the spatial diversity of a network, while the Ignition Variability, about the diversity across time.

Intrinsic Ignition is a novel technique based on graph and network theory (Figure 4.2a). For any node, its inner ignition capability is fully characterized by the Intrinsic Ignition as a measure of its spatial diversity, and its variability as a measure of its diversity across time. To compute both aspects, any continuous signal can be binarized using a threshold θ such that the binary sequence $\sigma_i(t) = 1$ if $z_i(t) > \theta$, crossing the threshold from below, and $\sigma_i(t) = 0$, otherwise [Tagliazucchi et al., 2012]. This simple method generates a raster plot with a discrete sequence of events, which is more efficient in terms of complex computations (see raster plots examples for each condition in Figure 4.2b).

Moreover, a Hilbert transform is performed to the continuous signal, defining the phases for each time point and node. Using these phases, a phase correlation or pairwise phase synchronization between regions j and k is defined as $P_{jk}(t) = e^{-3(\varphi_j - \varphi_k)}$. For each of these matrices, another binarization process is applied for a given absolute threshold between 0 and 1 (scanning the whole range), and therefore the symmetric phase lock matrix $P_{jk}(t)$ can be binarized such as 0 if $P_{jk}(t) < \theta$, 1 otherwise. Then the length of the largest component is computed, generating a curve with this value for each binarized phase lock matrix. The area below the curve is defined as the integration value for that node (ROI) at that time point. Running the same procedure for all time points creates a table of integration across time for each node. For each event defined from the first binarization procedure, and a flag window (commonly 4 TR), the total integration is computed as the average across the delta time defined by the event and the flag. It builds a matrix of integration across events. The average integration across events is

defined as the Intrinsic Ignition, while the standard deviation integration across the same events corresponds to the Ignition Variability.

Intrinsic Ignition and Intrinsic Variability produce one value for each of the 82 nodes. The procedure is repeated through subjects and conditions, creating a data matrix D_{ij} for each condition such that i corresponds to the nodes and j to the subjects (Figure 4.2c). The mean of these values across subjects (j) corresponds to the Intrinsic Ignition per node and Ignition Variability per node, respectively. The mean across nodes (i) is defined as the Mean Intrinsic Ignition and the Mean Ignition Variability. The standard deviation of the Mean Intrinsic Ignition (intrinsic ignition across nodes) returns a quantification value of the shape of the Intrinsic Ignition curve (considering all the values sorted from higher to lower), which here is defined as the Hierarchy, a quantifier for each subject.

To complement these analyses, a Spectrum Hierarchy plot is specified as a circle plot with different levels. Each level corresponds to a threshold of the Intrinsic Ignition (Spatial) and/or Ignition Variability (Temporal) curve: $i \geq \mu + \sigma$; $\mu + \sigma > i \geq \mu$; $\mu > i \geq \mu - \sigma$; $\mu - \sigma > i$. i refers to the index node, the mean value, and the standard deviation of the curve. The distance between level is given by the thickness of the level and the value of $\mu + \sigma$ for level 1, for level 2, $\mu - \sigma$ for level 3, $\mu - \sigma - min(nodes)$ for level 4 (Supplementary Figure B.2). The thickness of each level line is the number of nodes on that level; a thicker line means more nodes than thinner lines (e.g. red line in the figure marks the thickness of level two). The uniformity of the spectrum hierarchy for one condition characterizes the ignition curves in terms of the hierarchical organization across nodes [Deco and Kringelbach, 2017].

To explore if some nodes would be locally driving the global changes observed, a local analysis was performed on Intrinsic Ignition and Ignition Variability per node. This test consisted of finding which nodes are following the global tendency measured in the spatial and temporal aspects of ignition. As will be discussed in Results, the global tendency was the highest values of Intrinsic Ignition and Ignition Variability in awake, medium values in light sedation, and lower values in deep conditions. These tendencies were translated to logical propositions in order to find the nodes that satisfied all the propositions across conditions. These propositions were: $nodes \geq \mu$ for awake, $\mu + \sigma \geq nodes$ for light conditions and $\mu \geq nodes$ for deep conditions. More restricted logical propositions do not produce results. Additionally, instead of the mean (μ), the median was also used as cut off in order to search for consistent results. The number of regions obtained from one or another method changes slightly for Intrinsic Ignition,

but there is a clear overlap of regions, confirming part of the consistency expected. Finally, a subject by subject analysis was also performed as a supplementary test. In this case, the same propositions above were run in each subject to later generate a histogram of occurrence. A threshold of 60% of occurrence was imposed to find the regions above the threshold.

4.5.8 Statistical Analyzes

The main statistical test used in this work was the non-parametric Kolmogorov-Smirnov test (unless another test is explicitly stated). It is due to the characteristics of the data and their distributions [Rosner, 2012], for example in order to compensate for the unbalanced number of trials per condition. Therefore, the independence of measures and conditions is a statistical assumption commonly accepted for monkey data)[Uhrig et al., 2018], together with the continuous nature of our measures [Rosner, 2012]. Confidence intervals (CI) at 95% were computed as $\mu \pm 1.96 * \sigma/n$. As above, μ the mean, σ the standard deviation and n the length of the data points.

Additionally, effect size analyses were performed to quantify the differences given by statistical tests on Intrinsic Ignition and Ignition Variability per node. The method used was the rank-biserial correlation analysis for independent samples (other effect size techniques such as mean difference, AUROC and Cohen U1 did not present major differences with the results of rank-biserial correlation). In this test, ranks between -1 to +1 correspond to maximal effects and 0 means no effect. To compute a confidence interval for effect size analyzes, 10.000 bootstrapping iterations were performed (more details [Hentschke and Stüttgen, 2011]).

Chapter 5

TOWARDS NEW CONCEPTS

In the search for a sound model of consciousness, we aim at introducing new concepts: closure, compositionality, biobranes and autobranes. This is important to overcome reductionism and to bring life back into the neuroscience of consciousness. Using these definitions, we conjecture that consciousness co-arises with the non-trivial composition of biological closure in the form of biobranes and autobranes: conscious processes generate closed activity at various levels and are, in turn, themselves, supported by biobranes and autobranes. This approach leads to a non-reductionist biological and simultaneously phenomenological theory of conscious experience, giving new perspectives for a science of consciousness. Future works will implement experimental definitions and computational simulations to characterize these dynamical biobranes interacting.

5.1 Introduction

Neuroscience needs new concepts to approach the brain and its cognitive functions [Stern, 2017]. Many current influential concepts are based on computer metaphors [Daugman, 2001, McCulloch and Pitts, 1943, Ashby, 1957, von Neumann, 1958, Wiener, 1985, Piccinini, 2004, Miłkowski, 2018]. For example, information processing, integration, codification, and communication reduce the brain complexity to physical computations. These computations leave us without biology, but with new open questions. Some of these questions range from simple to complex, such as how to define brain regions [Stern, 2017], how consciousness emerges from physical processes [Chalmers, 1995a, Thomas Nagel, 1974], and whether computers might become conscious or not [Dehaene et al., 2017, Signorelli, 2018b].

In the field of neuroscience of consciousness that need is evident. For instance, the two most influential theories of consciousness are computational theories [Dehaene et al., 2014, Mashour et al., 2020, Tononi et al., 2016]. Their language reduces consciousness to electrochemical neural interactions, without mentioning what is unique in cells and neurons. However, cells and neurons are not only electrochemical, or even more general biophysical mechanisms. Instead, there is something unique in the intrinsic *organization* of cells and neurons which makes them alive. Are these unique and irreducible qualities of life somehow related to the irreducible features of conscious experience?

In this article, we introduce some novel and reintroduce few old concepts to suggest that life is at the core of any sound explanation of consciousness. Instead of treating cells and neurons as performing sophisticated coding and decoding, a better metaphor is the living cell itself: cells and neurons are living beings interacting in order to get food and energy that keep them safe and alive. As such, two neurons do not send or communicate through intricate signals, but may just get and send biological resources. This systemic closure is understood as an operational closure, a more elaborated form of biological autonomy that we will introduce across these pages. We claim that this biological circularity is at the core of the conscious experience, composing a further living closure between multilevel and multidimensional brain-body systems and the animal's environment.

5.2 Philosophical and Experimental Perspective

Despite recent progress in the neuroscience of consciousness [Seth, 2018], signatures of conscious experience convey isolated experiments about disparate neural correlates [Aru et al., 2012]. These different correlates suggest different aspects of the conscious experience, e.g. the phenomenal consciousness and access consciousness [Block, 2005], among others [Aru et al., 2012, Bachmann and Hudetz, 2014, Tsuchiya et al., 2015, Storm et al., 2017]. Unfortunately, these aspects and their neural signatures also lack an integrative explanation [Bachmann and Hudetz, 2014, Bayne et al., 2016], as well as a direct link to the phenomenology of consciousness.

We suggest that a sound model of consciousness requires a more promising point of departure: i) A *radical embodiment reformulation* [Thompson, 2004, Thompson, 2007], and ii) the

integration of brain-body signatures of consciousness in a multilevel organization to reconcile different signatures of conscious experience.

5.2.1 Radical Embodiment

At the core of scientific studies of consciousness lies the hard problem of consciousness. The hard problem of consciousness is a consequence of reducing the mental ontology to the physical ontology. The mental corresponds to unverifiable claims and subjective modes of existence, such as pain or the "redness" experienced only by the subject (subjectivity). Contrary, the physical corresponds to verifiable claims and objects existing independently of others (objectivity). The reduction of the former to the later conveys the question illustrated by Thomas Nagel: "If mental processes are physical processes, then there is something it is like, intrinsically, to undergo certain physical processes. What it is for such a thing to be the case remains a mystery" ([Thomas Nagel, 1974], pp. 445-446). This way of *formulating* the problem implies that "the mental" and "the physical" are two opposed reified substance-ontologies, i.e. two different substances having constant properties and existing each one by itself. On the one side is consciousness (qualia), on the other side, the physical body (with its structure, functions, and mechanisms).

We can avoid this problem by changing our ontologies. Instead of invariant and independent substance-ontologies, we consider variant and interdependent process-ontologies [Rescher, 2012]. In this case, the existence is only given by interdependent transformations, and the mental and the physical body become related to each other: they are two different modes of the same existence [Signorelli et al., 2021d]. Then, in Nagel's formulation, one can replace the term "physical" by "bodily" and reformulate the above question in the following way [Thompson, 2004, Thompson, 2007]: if mental processes are *bodily* processes, then there is something it is like, intrinsically to undergo certain *bodily* processes. In other words, what is it for a physical living body (*Körper/ leiblicher Körper*) to be also a lived body (*Leib/ körperlicher Leid)*)? Critically, the explanatory gap is now between *two types* within *one typology* of embodiment: The living body (*Körper*) and the lived body (*Leib*) are the two modes of appearance of one and the same body. This is called the radical embodiment reformulation.

The radical embodiment reformulation demands two important conditions for a sound model of consciousness: i) consciousness requires a living body, and ii) consciousness cannot be re-

duced to only neural states. Although the first condition seems evident for any biologist, what makes cells and neurons alive is rarely considered relevant regarding consciousness. The second condition follows the first: because the living body and the lived body (consciousness) are two modes of the same body, it is wrong to assume that the lived body emerges from the living body (including the brain).

Therefore, a sound model of consciousness must account for what makes cells and neurons living entities, including the co-dependence between the living body and the lived conscious body. The radical embodiment reformulation urges us to account for the various biological processes that relate to consciousness without reducing it to neural systems. To tackle these processes, we need to ask: what are the relevant *brain-body* signatures from scientific studies that can inform such a sound model of consciousness?

5.2.2 Brain-Body Signatures

Instead of focusing on the necessary and sufficient neural events for conscious experience, we rather ask about the necessary and sufficient kind of *organization* for that conscious experience to occur. In other words, instead of describing one-dimensional interactions only at the level of electrochemical components (cells and neurons), we propose a shift to the relevant interactions at the level of organization [Gershenson, 2013a, Gershenson, 2013b, Mazzocchi, 2008] *between* various kinds of biophysical components (systems).

To this end, brain-body signatures of conscious experience suggest a multilevel organization, as well as various aspects of conscious experience that need theoretical reconciliation.

One interesting example of brain signatures is the activity of the conscious resting-state brain and its connections with brain-body activity. At resting state, studies of functional magnetic resonance imaging (fMRI) show that brains present intricate anticorrelated activity [Biswal et al., 1995, Fransson, 2006, Fox et al., 2005, Pessoa, 2014]. In many cases, this activity is simulated by dynamical systems at the edge of criticality [Deco et al., 2008, Breakspear, 2017]. These models need to adjust different parameters, among those, noise plays an important role [Ghosh et al., 2008, Deco et al., 2009]. Noise is associated with the intrinsic noisy cellular and neural activity [Faisal et al., 2008], and it also relates to the physiological coupling between the brain and the rest of the body as nonstationarities reveal [Thompson and Varela, 2001, Lau-

mann et al., 2017, Nguyen et al., 2016]. The existence of these nonstationarities suggests that brain and body systems are interconnected. Two examples are the interconnected brain-stem system, which regulates homeostasis, and nuclei that regulate sleep and wakefulness [Thompson and Varela, 2001]. These couplings change during sleep [Bashan et al., 2012, Bartsch et al., 2015, Ivanov et al., 2017] and under different anaesthetics [Stankovski et al., 2016]. Even more fascinating, part of this bodily activity seems to influence conscious perception [Park et al., 2014] and confidence [Allen et al., 2016].

The molecular environment and metabolism also regulate these brain-body couplings [Haydon and Carmignoto, 2006, Petit and Magistretti, 2016, Jha and Morrison, 2018]. In relation to conscious activity, evidence suggest that glial cells and their energetic production is involved on these regulations [Bélanger et al., 2011, Ramadasan-Nair et al., 2019, Perouansky et al., 2019, Velazquez, 2020]. Another example corresponds to the minimal energetic requirement which is necessary to recover consciousness from chronic impairments [Shulman et al., 2009, Stender et al., 2016, Di Perri et al., 2016]. It implies a relevant metabolic coupling. Furthermore, if one compares disorders of consciousness with normal awake subjects, dynamical changes are observed in the form of a reduction of the brain repertoire [Demertzi et al., 2019]. This intricate dynamics of healthy brains is partially recovered using deep brain stimulation in different zones of the chronic impaired brain [Schiff, 2010, Schiff, 2013, Koubeissi et al., 2014, Corazzol et al., 2017]. Consciousness loss during applications of anaesthetics also show similar dynamical signatures [Barttfeld et al., 2015, Uhrig et al., 2018], but differently, anaesthesia presents two types of emergence modes from sedation. One of them is a very graded and gradual emergence, whereas the other generate abrupt arousal, typically followed by disorientation and sudden movements [Canet et al., 2003, Lepousé et al., 2006]. Under anaesthesia, the induction and emergence also present asymmetries [Lee et al., 2011, Chander et al., 2014, Warnaby et al., 2017]. All together, remind us about the importance of biochemical and molecular interactions, mainly between endocrine, immune systems and neural systems [Thompson and Varela, 2001].

Finally, these brain-body activities generate different brain signatures of consciousness associated with different aspects of consciousness. For example, during sleep states, electroencephalogram (EEG) activity and body rhythms show clear physiological changes and transitions [Simon and Emmons, 1956, Brown et al., 2012]. Unlike anaesthesia, those changes are natural and gradual. During dreams states, EEG measurements reveal brain activity mostly in

the parietal-occipital cortex [Siclari et al., 2017]. It is associated with phenomenal consciousness. Contrary, in awake conditions, experiments on conscious perception in humans and other primates, convey evidence about a frontoparietal-cingulate network and ignition activity from the frontal cortex to the rest of the brain [Van Vugt et al., 2018]. It is called access consciousness [Block, 2005]. In this mode, the trajectories of brain states seem to accelerate when someone perceives a stimulus, compared with the opposite situation [Baria et al., 2017]. It suggests transient dynamics of access consciousness. Other distinctions indicate two different cognitive systems [Shea and Frith, 2016, Herzog et al., 2016, Dehaene et al., 2017, Signorelli, 2018a], associated with two conscious processes: the awareness of content (awareness) and the awareness of the processing on these contents (self-reference or self-monitoring).

Are these signatures and distinctions conflicting evidence about the neural correlates of consciousness? [Boly et al., 2017]. Taking them in isolation, probably yes. However, taking them as a whole, these signatures and modes of consciousness may correspond to different brain-body couplings and dynamical phase transitions.

The evidence above supports a multilevel organization, where the molecular environment, cellular organization and neural systems interact to ensure conscious experience [Thompson and Varela, 2001, Prentner, 2017, Kringelbach et al., 2020]. These multilevel cycles and processes between brain and body underpin the integrity of the organism as a whole [Thompson and Varela, 2001]. The body activity relevant to brain activity may define a subjective frame supporting subsequent conscious experiences through interactions of neural and cellular responses [Velazquez, 2020] to external but also visceral stimuli [Critchley et al., 2004, Seth, 2013, Park and Tallon-Baudry, 2014]. The body activity signalled to the brain, and this subjective frame may represent different brain-body systems interacting. The various transitions, asymmetries observed during sleep and anaesthesia, and diverse network signatures of phenomenal consciousness and access consciousness, may reflect the degrees of couplings of these different systems and the dynamical phase transitions triggered by them [Werner, 2012, Werner, 2013].

As a consequence of this discussion, we propose that a sound biological model of consciousness must integrate conscious experience in its irreducibility, in order to constitute a comprehensive framework.

5.3 Concepts for a biological model of consciousness

According to the previous discussions, the aforementioned conditions impose further requirements. On the one hand, the definition of the living body needs to capture the uniqueness of living beings in contrast to non-living things. On the other hand, we need a principle to explain the mutual relationship between the living body and the lived body, i.e. their co-dependence at different scales.

In the following, our framework provides a concrete implementation and extends the original embodiment conjecture [Merleau-Ponty, 2005, Thompson and Varela, 2001]: consciousness relies on how brain dynamics are embedded in the somatic and environmental context of the animal's life.

5.3.1 Closure and Biological Autonomy

The starting point of our model is the living system. One way to distinguish living systems, such as cells, neurons, and bacteria, from non-living systems, draws on a living system's distinct network of internal productions [Varela et al., 1974, Maturana and Varela, 1998, Ruiz-Mirazo and Moreno, 2004, Maturana, 2011]. In theoretical biology, this internal cellular organization is referred to as closure [Varela et al., 1974, Letelier et al., 2006, Letelier et al., 2011, Cárdenas et al., 2010]. There are, however, different notions of this closure, and one form to refer to them is called metabolic closure (Figure 5.1A). Metabolic closure means that all the catalysts needed to stay alive are produced by the organism [Letelier et al., 2011]: "molecules that define the metabolic network of a cell, whether metabolites or enzymes, are produced by processes which are themselves mediated by other molecules produced by the very same metabolic network". Those biochemical reactions constitute metabolism from which enzymes and other proteins participate in those reactions as well as are the product (metabolites) of those reactions [Letelier et al., 2011]. One example of a metabolic reaction is the glucose metabolism and its different profiles in neurons and astrocytes [Magistretti and Allaman, 2015]. In the case of the glucose metabolism the reaction is catalysed by the enzyme glucokinase: $Glucose + ATP \xrightarrow{G} Glucose6phosphate + ADP$. This reaction can be seen as the action of an operator G transforming the input molecules into the output molecules. The internal set of participating molecules, enzymes and proteins signify the closure: Sometimes they operate as

catalysers and at other times as inputs or outputs (Figure 5.1A). Hence, the organism becomes distinct from its environment through its dynamics of production. The product of this metabolic process of production is the producer itself. In other words, living systems exhibit a particular form of *closure* (see [Letelier et al., 2011] for a detailed discussion).

Definition 8 *Living systems exhibit closure: They are sustained as a network of processes that are recursively dependent on each other.*

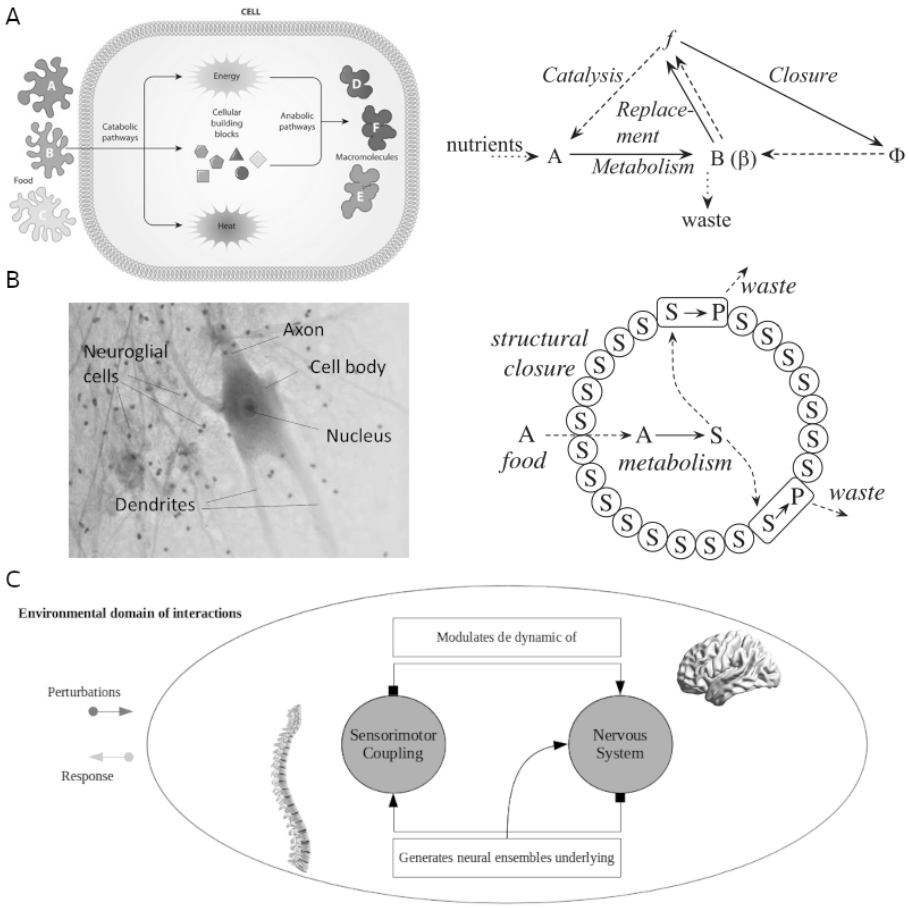


Figure 5.1: Different types of closure. (**A**) Metabolic closure refers to the assumption that "all catalysts needed for metabolism are themselves products of metabolism" [Cárdenas et al., 2010]. Cellular metabolism corresponds to the set of chemical reactions to maintain cellular life. The left diagram summarizes one way to represent the closure of these reactions. Dashed arrows represent catalysis, and continuous arrows represent transformations of matter by chemical reactions [Letelier et al., 2011]. Metabolism is a set of chemical transformations $A \rightarrow B$, catalysed by enzymes f. Replacement corresponds to the re-synthesis of f by a replacement system ϕ. Enzymes are synthesized from the products of metabolism, requiring other catalysts so that $\phi \rightarrow B \rightarrow f$. Then, closure becomes the continuous replacement of any catalyst, such that the diagram is closed [Letelier et al., 2011]. Figures adapted from [O'Connor and Adams, 2010] and [Letelier et al., 2011]. (**B**) Living systems are also structurally closed. This applies to cells, neurons and glias. The concept of autopoiesis signifies this closure as described by the left diagram. In that diagram, dashed arrows are physical movement while solid arrows represent chemical reactions. In this sense, autopoietic systems become "encapsulated systems" [Letelier et al., 2011]. The metabolic reactive network produces molecular components that determine the bounded system that generates the metabolic reactive network [Thompson, 2007]. Diagram adapted from [Letelier et al., 2011]. (**C**) The nervous system is one example of biological autonomy and organizational closure. In this case, the closure is at the level of patterns of activity. Sensorimotor coupling modulates the nervous system that reciprocally generates patterns of activity shaping the sensorimotor system. In other words, the nervous system is immersed in a loop of activity where sensory input defines motor output and vice versa. Diagram adapted from [Thompson, 2007].

Closure makes cells and neurons unique. One form to specify that closure corresponds to

the concept of *autopoiesis* as structural closure [Cárdenas et al., 2010]. An autopoietic system is itself a network of biological recursive and intertwined actions of components and production [Varela et al., 1974, Maturana and Varela, 1998, Ruiz-Mirazo and Moreno, 2004, Maturana, 2011]. These biological actions imply a context of spatial-topological boundaries or "membranes" (Figure 5.1B), which also require and define the component production network [Ruiz-Mirazo and Moreno, 2004, Thompson, 2007]. These interactions continuously regenerate their own network of processes and enact a concrete topological unit in space [Varela et al., 1974]. Therefore, cells and neurons are closed systems: Instead of being static entities, they only exist as arising and temporally sustained networks of recursively interacting processes.

Definition 9 *Biological closed systems internally produce what then constitutes their operation of production.*

Autopoiesis is considered one of the possible minimal set of requirements to define what is living and what is not [Varela et al., 1974, Letelier et al., 2011]. It conveys three minimal criteria that any autopoietic organization needs to satisfy: a) Semipermeable boundary: Does the system have a boundary that allows us to distinguish between inside and outside in relation to its relevant components?; b) Reaction network: Are the components being produced by a network of reactions inside the boundary?; c) Interdependency: Are conditions a) and b) interdependent? Are the components of the boundary being produced by the internal network of reactions as well as this network is regenerated by conditions from the boundary itself? If a system meets these three criteria, then the system is an autopoietic organization (Table 5.1).

Entity	**Boundary**	**Network**	**Interdependent**	**Is autopoietic?**
Virus	Yes	No	No	No
Crystal	Yes	No	No	No
Bacterium	Yes	Yes	Yes	Yes
Amoeba	Yes	Yes	Yes	Yes
Mitochondria	Yes	Yes	No	No
DNA section	No	No	No	No
Autocatalytic set	No	Yes	No	No

Table 5.1: Autopoietic systems according to three classification criteria. Table from [Thompson, 2007].

The paradigm example for autopoiesis is the living cell. In a living cell, the constitutive processes are chemical. Those chemical metabolic reactions recursively depend on each other. This means that, in order to occur, one chemical metabolic reaction requires the products of other chemical metabolic reactions. Those reactions mutually depend on each other. By this,

their whole network of relations constitutes the living cell as a unity in the biochemical domain. Interestingly, this constitution of the living cell as a unity takes a special form as spatial boundary [Thompson, 2007]. This spatial boundary is realized through the living cell's membrane that enables metabolic reactions while the metabolic reactions bring forth the cell's membrane. This mutual dependence is at the core of autopoiesis [Letelier et al., 2011].

Autopoietic systems are a specific kind of *autonomous systems*. As living systems are metabolically and structurally closed, they are open systems in terms of thermodynamics. In other words, they are connected with the environment to obtain the energy that its metabolism requires. In Figure 5.1A-B, this is represented by nutrients, food and the irreversible reaction of producing waste. However, external causes do not modify the internal organization but may contribute or modulate reactions to them (Figure 5.1C). The inputs and outputs of biologically closed organizations come from and go to the environment [Thompson, 2007]. The closure property is not about the exchange of energy or materials, it is about how this exchange is regulated. In a biological system the flow of energy that keeps the system away from the thermodynamic equilibrium is regulated by the organization of the system itself (endogenous self-organization), while in the case of a physical system, it is controlled by external mechanisms. The first condition defines an autonomous system, whereas the second condition defines a heteronomous system [Thompson, 2007]. The former develops internal, local, and global processes to stay away from thermodynamic equilibrium [Ruiz-Mirazo and Moreno, 2004], keeping its intrinsic dynamic, while the latter is determined by external mechanisms [Thompson, 2007].

Definition 10 *Biological autonomy is the closure between the internal productions (metabolism) and the external extensions of this internal organization as actions in the environment (agency) [Ruiz-Mirazo and Moreno, 2004].*

Biological autonomy makes use of a more general instance of closure. In this case, closed interactions define only virtual boundaries (nonphysical/non-material membranes), i.e the closure is operational [Varela, 1979, Varela, 1997]. In other words, what is now being regenerated is the internal topology, not the components. This is called *operational closure*, where all the dynamic processes to keep the organization of the system are maintained or sustained by the system: They construct and reproduce their own internal topology. Examples of such systems are microbial communities, immune system, the nervous system, neural assemblies,

multi-cellular organisms, but also insect colony, or animal society, among others.

In contrast to the aforementioned autopoiesis, the realization of operational closure in, for example, multi-cellular organisms and neural assemblies does not involve a *spatial* boundary. Rather, they bring forth an *identity* constituted through the recursive network of relational processes without a fixed physical membrane. In the case of an autonomous social network such as an insect colony, the boundary is social and territorial, not material. While metabolic or autopoietic closure enacts a minimal bodily unity at the metabolic level, another example, the sensorimotor closure as in the case of the nervous system (Figure 5.1C), brings forth a sensorimotor unity at the perception-action level [Thompson, 2007, Varela, 1997].

In summary, we characterized a living system by its metabolic and operational closure. This closure presupposes the notion of biological autonomy: "[e]very autonomous system is organizationally closed" ([Varela, 1979], p. 58), i.e. operationally closed [1]. This makes closure and autonomy deeply interdependent concepts which emphasize the system's dependence on a network of recursively interacting processes.

5.3.2 Compositionality and Co-arising

If the relationship between the living body and lived body is co-dependent, how does this co-dependence work? Dynamically, one can understand this co-dependence as local *bodily* processes giving rise to novel global *consciousness* processes that have "their own features, lifetimes, and domains of interaction" ([Thompson and Varela, 2001] p. 419). Simultaneously, those global characteristics of a system's conscious activity constrain the local interactions on the body level [Thompson and Varela, 2001, Rodríguez, 2008]. In other words, none of the two is reduced to the respective other, they co-arise.

Definition 11 *The co-arising of the living body and the lived body implies that they reciprocally depend on each other. None of the two can be reduced to the respective other.*

In our framework, co-arising becomes a principle of *compositionality*: the parts and the whole are mutually defined. In Evan Thompson's words, "part and whole are completely in-

[1] Varela in his writings does not distinguish between organizational and operational closure. He uses the two terms interchangeably.

terdependent: an emergent whole is produced by a continuous interaction of its parts, but these parts cannot be characterized independently from the whole" ([Thompson, 2004], p. 391). This interdependence of parts and whole is called *dynamic co-emergence*. It reflects the notion that (i) the parts give rise to the whole, (ii) the whole gives rise to the parts, and that (iii) none of the two can be reduced to the respective other, they co-emerge. Therefore, dynamic co-emergence refers to the idea that both propositions apply simultaneously. Following the examples above, the inside ("sensorimotor self") and the outside ("environment of otherness") co-emerge through nervous system's operational closure (sensorimotor closure) at the level of perception and action (context). This notion of dynamic co-emergence shows important parallels to the idea of *compositionality*.

Here, compositionality formalizes this interdependent relationship. A new whole is a composition of its parts only if the whole has the properties of its parts and vice versa (Figure 5.2). In other words, the system is non-trivially decomposable [Coecke, 2011, Coecke et al., 2016]. Importantly, compositionality is not the same as composing. Composition highlights the parts, while categorical readings of compositionality accentuate the whole, such that the parts need to be defined by the properties we want to describe in the whole. In other words, any compositional division is contextual to the whole property described [Atmanspacher and Rotter, 2008]. For example, if we want to recover whole-brain oscillatory activity, the minimal section in our system analysis becomes a group of oscillators interacting. This partition is independent of the physical partition of the brain organization, i.e. it is operational: we divide the brain according to the operation we want to describe. In category theory, a branch of mathematics, the composition of two morphisms (processes) $f : x \to y$ and $g : y \to z$ in a category, needs to produce another morphism, such as $g \circ f : x \to z$. The new morphism $g \circ f$ is called the composite of f and g. Compositionality forces us to define the parts and the whole simultaneously, demanding a principle which is neither reductionism nor holism (Figure 5.2).

Definition 12 *The whole is constituted by the relations of the parts, and the parts are constituted by the relations they bear to one another in the whole [Thompson, 2004].*

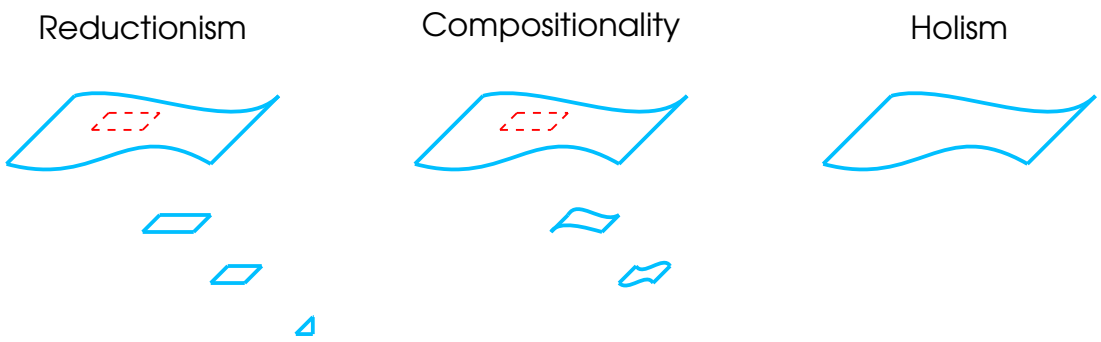


Figure 5.2: Compositionality. Take the example of oscillatory membranes. In order to analyze them, reductionism will divide the membrane into small pieces (e.g. lipids, proteins, ion channels, etc). In the process, the global properties of the membrane are usually lost, in this case the whole oscillation. In general, a reductive approach allow us to understand the physical components, but it is unable to recover the whole (usually emergent properties) by the mere description of the parts. Contrary, a compositional approach divides the membrane considering the property of the whole being described. In this example, its capability to oscillate. The smallest component is indeed a group of oscillators. As such, the whole property is always present in the relationship of their parts, and the composition of their parts is much more than the sum of them. Finally, a holistic approach would negate the possibility to explain the membrane oscillation by the mere sum of their parts. Like compositionality, holism claim that the parts of a whole and the whole are interdependent, but differently than compositional approaches, these parts cannot exist without the whole.

The radical embodiment also implies a contextual relationship between the whole and parts. The output of any biological experiment will depend on the feature observed in the biological system, but the same observed system, regarding a different context, might bring other conclusions. In the neuroscience of consciousness, this contextual relationship appears on the multiple neural signatures of consciousness. Different experimental conditions point out to different neural correlates of consciousness [Aru et al., 2012], e.g. prefrontal-parietal networks are relevant during conscious perception [Van Vugt et al., 2018], but posterior cortical regions are dominant under conscious dreams [Siclari et al., 2017]. According to the embodiment conjecture, we cannot escape to this contextual behaviour of neural activity, because this activity is immersed in a body with complex intertwined relationships. As a consequence, the definition of the relevant brain-body organizations for conscious experience is compositional and contextual to the whole experience.

5.3.3 Biobranes and Autobranes

Biology is all about autonomy, and biological membranes seem to signify this autonomy. Biological membranes play the role of boundaries between the living system and the environment. They regulate the exchange of resources, protect the internal system, among other important functions. It is not surprising that the core idea behind the concepts of autopoiesis and metabolic closure is indeed a formalization of the intuitive notion of biological membranes. The external membrane in cells is not just part of the network of internal production that then

become elements of the membrane itself, but it is a topological closed system: there is no starting nor ending point in the cell membrane. Then, operational closure becomes a generalization of biological membranes aiming to incorporate virtual systemic boundaries, such as the non-material boundaries that define the intertwined social relationships in a group of animals. In this case, the boundary is organizational.

In physics, an acronym for membrane is *brane*. A brane is a n-multidimensional dynamic object that posses energy in form of tension over its volume. This energy becomes the energetic source for certain interactions, while the observable universe comprises the internal volume of that brane. Mathematically, the dynamical evolution of a brane is a map $\varphi : W \rightarrow M$, where W is a reference manifold with n+1 dimension and M represents the "spacetime" thought the brane propagates [Moore, 2005]. In this nomenclature, $\varphi(W)$ is called the *worldvolume*. Moreover, branes wiggle and bend through oscillations. These oscillations are sections of the normal sheaf to a subset $\varphi(W) \subset M$. Then, different mathematical structures [2] are added to W and M, in order to study different phenomena. One example of such branes is the surface of the ocean [Moore, 2005], while the best-known example of branes is associated with string theory and theories of gravity (D-branes). In this last example, a brane corresponds to local boundary conditions preserving multidimensional invariance (formally, conformal invariance) and the tension T becomes a key feature to define different types of particles and cosmological scenarios.

The concept of brane is relevant to our discussions because we can extend and formalize our biological intuitions without the need of reducing the brain to mere computations. Apart from the mathematical structure that defines branes, what also makes branes different than other theoretical descriptions, such as dynamical systems theory, is their capability to recover other physical systems, i.e. topological branes may become primary structure. For example, dynamical systems describe organizations usually evolving in time, and therefore making time the independent variable. In this case, the group of independent variables take the role of a fundamental dimension of description in which the system evolves. In the case of branes, they might be treated as physical self-sustained systems, i.e. their energetic interactions depend only on the brane volume. As such, the intrinsic structure is what defines the worldvolume, and we can further interpret this worldvolume in the context of biological entities. In this context, an interesting conceptualization is the D-branes as primary or the fundamental organization

[2] A Mathematical structure is the set of mathematical objects and their relationships that satisfied certain axioms.

from which spacetime and other dimensions emerge. As such, spacetime might arise from purely topological branes. In biology and cognition, these brane-structures may represent the different brain-body organizations that give rise to the cognitive space and time from mutual constraints between the environment and the biological agent [Signorelli et al., 2020a]. In other words, the biological space and time becomes embodied. Therefore, instead of reducing our biological membranes to physical branes, we conjecture that biological branes related to conscious experience would be as fundamental as the fundamental branes in theoretical physics.

With this conjecture in mind, we now introduce an extension of biological autonomy: the concept of *biobrane*. Closure and biological autonomy lead to organizational invariant [Letelier et al., 2006], self-organized, and self-regulated systems (the other way around does not always apply). These systems are called biological autonomous systems and their organizational invariant may take the form of either concrete membranes (e.g. cells and neurons) or virtual boundaries (e.g. immune system). Thus, we define a biobrane as all the possible closed biological membranes/boundaries at the meso-scale of a whole biological organism that self-sustain their interactions in relationship with conscious experience.

Definition 13 *A biobrane is a multidimensional dynamical description of biological autonomous system forming a unity (operational closure and self-regulation), in the form of concrete or virtual meso-scale membranes of an organism.*

Using compositionality, a particular biobrane is then the autobrane. If a biobrane is both operationally closed and composed of units with metabolic closure (specially autopoietic units), they are called autobranes and entail a double closure composition.

Definition 14 *An autobrane is a biobrane operationally closed and self-regulated, composed by elements that are operationally closed and self-regulated as well.*

We propose that biobranes and autobranes are a more powerful conceptual framework and its mathematical machinery may describe the notions of closure better than dynamical systems.

Biobranes and autobranes are extensions of biological networks, cells and neurons, such as D-branes are generalizations of close and open strings. This formal analogy is, however, just an

analogy. We do not claim that biobranes are built in the same way than cosmological branes, but that their mathematical structure is similar [3]. In other words, we can model biobranes using the mathematical machinery of branes, up to certain distinctions (e.g. we might not need quantum branes at the Plank scale). This approach follows the common pragmatic use of the same type of differential equations to model an endless number of different physical and biological phenomena. For example, the *cable equation* is a useful equation modelling the propagation of electromagnetic signals in a cable, as well as a useful approximation of the propagation of action potential in pyramidal neurons.

In this context, the radical embodiment and its connection with dynamical system theory is more practical than essential [Thompson, 2007]. This link looks for a mathematical formalization of the biological autonomy in neural systems that "actively generates and maintains its own coherent and meaningful patterns of activity, according to its operation as a circular and reentrant network of interacting neurons" [Thompson, 2007]. Biobranes and autobranes may provide that formalization. In our case, the biological closure discussed above might be modelled by the different topologies and biobrane volume, its T tension as intrinsic biological energy and its worldline as the dynamical evolution. These applications are left for future works, while we focus here in its conceptual introduction.

An important remark is that modelling does not imply a reductive or ontological metaphor. The metaphor to conceptualize these biobranes as a living organization is to understand their interactions in a similar way that two independent living beings interact, e.g. two amoebas. We understand that "our organism is a meshwork of "selfless selves," and we are and live this meshwork" [Varela, 1991]. Therefore, biobranes would act and behave as independent functional organisms, while autobranes as functional and anatomical organic units.

Furthermore, biobranes and autobranes may generalize previous neurophysiological divisions of the brain anatomy and its function. For example, the simplest way to analyse the brain is to parcel it in regions of interest (ROIs) and average the physiological activity in each of these regions (Figure 5.3A). More recent efforts are focusing on multidimensional activity (Figure 5.3B). In this case, the brain region is not reduced to one dimension of physiological activity but becomes a three dimensional or bigger dimensional object [Basti et al., 2020]. Dimension are usually taken from principal components analyses of electrophysiological signals (EEG,

[3] Claiming a similar mathematical structure means claim similar relationships. This theoretical hypothesis can be tested using the machinery of Category theory.

fMRI, among others). Other attempts define extended anatomical and functional regions as the minimal unit of analysis, mainly cortical layers (Figure 5.3C). This is called layer-approach and focus on layer-fMRI analyses [Huber et al., 2020]. In this line, autobranes correspond to multidimensional layers with structural boundaries, while biobranes are multidimensional layers that not only incorporate neural systems, but also more general brain-body systems. In figure 5.3D we give a hypothetical example. The anatomical and functional parcellation is translated to virtual multidimensional membranes as abstractions of biological autonomous systems. The dimensions of biobrane activity include electrophysiological, metabolic, kinetic, among any other relevant physiological activity that characterizes, ensure the unity and the survival of the biobrane.

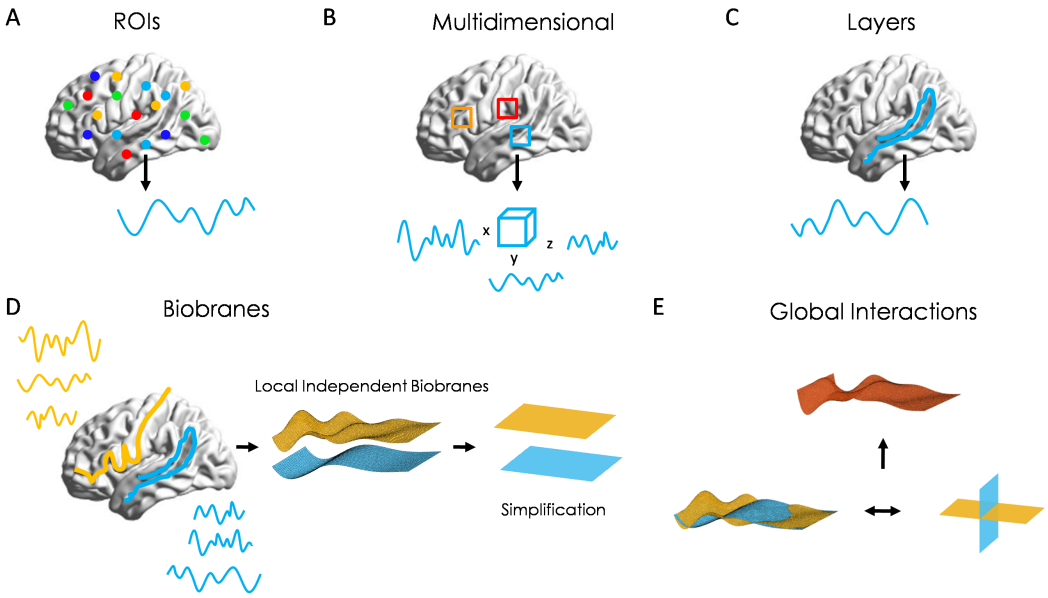


Figure 5.3: Biobranes. (**A**) In neuroscience, the brain parcellation across regions of interest (ROIs) is a common assumption. These regions form anatomical or functional brain networks, and its physiological activity is normally averaged to get one single time-series activity per node (light blue line). (**B**) Recent brain parcellation techniques define multidimensional time-series, using principal component analyses or other multivariate methods. The goal is to describe more complex multidimensional connections between brain regions. (**C**) Additional efforts focus on the physiological activity of different spatially extended regions, mainly cortical layers. In this case the anatomy and function define a whole-unite (the layer). (**D**) Our discussion extends previous assumptions to a brain-body organization, where each colour corresponds to one family of multidimensional biobranes. For visualization purposes, we plot two toy examples of hypothetical three dimensional branes. Each dimension may correspond to physiological signals (functional activity), like in multidimensional approaches, as well as kinetic changes of the whole (anatomy), metabolic exchanges, among others. In the general case, biobranes are not restricted to three or four dimension. To reason about the connectivity between biobranes, we can represent them as colour layers, a visual simplification. (**E**) By definition, these biobranes are independent organizations under unconscious conditions. Then, the interaction among biobranes generates new dynamical conditions due to breaks of symmetries within the biobranes. This is visualized by the overlap of their activities (bottom right) or by rotation layers (bottom left). This overlap composes a new "many colours" biobrane (top).

Biological realizations of biobranes may range from some cortical layers in the brain to the immune system in the body. The most relevant autobranes might convey neural layers and cell-glia layers, as we will discuss later [Velazquez, 2020].

In short, biobranes and autobranes are new and relevant concepts for brain functioning in

the context of conscious experience. First, the conceptual introduction of topological biobranes may lead to the mathematization of operational closure. From that, other closure compositions can be described. Secondly, we can implement multidimensional physiological brain-body signals using multidimensional approaches in biobranes. These dimensions may include anatomical, functional and metabolic interactions, among others. Thirdly, we can treat spatial and temporal dimensions as embodied in the biological agent, instead of treating them as independent variables external to the agent. In other words, physical space-time is no longer the theatre in which experience appears, but cognitive space-time might arise from experience, as biobranes and autobranes interact. Finally, biobranes and their ability to oscillate in multiple dimensions may serve to explore co-dependent configurations of conscious experience and brain-body signals, as they trigger each other. We discuss this final approach in the next section.

5.4 Closed Biobranes Composition

Following the main definitions above, we now introduce the main hypotheses regarding brain functioning and consciousness interaction.

5.4.1 Brain-body Architecture

Our first hypothesis generalizes the brain division to a membrane division of the brain and the body. In short, we propose a biobrane structure and *multibrane structure* (Figure 5.3D-E). Please notice that the common practice of region parcellation in cognitive neuroscience is, in fact, a weak form of our postulate:

Proposition 1 *The brain and the body allow for a multibrane structure.*

The multibrane structure is a group of multiple dynamical biobranes, here represented by layers. Each layer stands for an independent biobrane with its particular type of internal interaction and/or components. The brain multilevel structure now becomes a multidimensional group of biobranes interacting.

A biobrane of neurons may be defined by their main interactions through action potentials, while a molecular biobrane may interact by chemical gradients. These internal interactions represent local interactions across the biobrane. Then, interactions between biobranes correspond to branes acting as bridges between other branes and illustrated by rotation layers (Figure 5.3E). These types of interactions become global interactions and break the local dynamic, generating new dynamics and symmetries in the more general multibrane structure (Figure 5.3E). For example, a neural biobrane interacting with a molecular biobrane might make available more neurotransmitters to the neural biobrane, changing the neural biobrane dynamic, as well as this new activity changes the concentration of these chemicals and the dynamics of the molecular biobrane. The sustained interaction interferes with the biological stability of the original branes, triggering different responses in each other: a) breaking observed default synchrony as local integration across regions inside biobranes and b) spreading new activity through them.

An important observation is that the brane rotation is not material (Figure 5.3E), i.e. regions in the brain-body do not rotate as their mathematical representations. This point is evident in three-dimensional models of aggression (fear, behaviour and rage dimensions), where the animal in question is not physically moving in that topological manifold. The scientist models the mood of the animal [Zeeman, 1976]. In our case, rotations of layers are graphical characterizations of the membrane oscillatory activity: With increasing activity, overlaps with other biobranes increase, and therefore new influences in their intrinsic dynamic come into play.

5.4.2 Compositional Consciousness

We hypotetize that, poetically speaking, conscious experience is the biological universe of cosmological constellations supported by biobranes and autobranes interacting. Conscious experience is co-defined by the close coupling and compositional interactions between metabolism, autonomy and the animal's environment such that any conscious action returns to the animal in a meaningful way to that animal.

To understand this idea, we first define compositionality for brane structures. The interbrane interactions may become compositional, only if the new global system compounded by the branes is also a brane, i.e. the new system is also closed and self-regulated.

Proposition 2 *If two or more branes and their interactions generate a new global system*

of branes which is operationally closed and self-regulated, it is a composition of the former branes.

These brane compositions extend to biobranes and autobrane compositions. Autobranes compound other autobranes as long as the properties of autobranes, i.e. closure of the system and closure of their unites, still hold.

Following previous discussions, life is characterized by the closure of intertwined component productions, while biological autonomy is signified by the operational or organizational closure of these components (Table 5.2). In the embodiment framework, a cognitive agent corresponds to the coupling between the agent and the environment so that recurrent sensorimotor patterns of perception and action appear [Thompson, 2007] (Figure 5.1C). These patterns modulate but do not determine the endogenous activity, while the endogenous activity informs sensorimotor coupling. The internal realm is not a representation of the external. However, their mutual relationship is enacted by the living agent and the coupling mode with the environment. This recursive action, the closure of living systems and its environment, creates meaning and involves a *minimal lived experience*. This is the sense-making dimension that becomes the "intentionality in its minimal and original biological form" [Thompson, 2004, Varela, 1997].

Kind of Closure	**Characterize**	**Supporting Process**	**Description**
Compositional Closure	Conscious Experience	Multibrane	Closure among different levels of closure
Operational Closure	Identity	Biobranes	Closure of the internal topological organization
Structural Closure	Life	Autopoietic organization	Closure of component production

Table 5.2: Closure composition and co-arising of conscious experience.

We can extend these intuitions and claim that different living organizations convey different types of such experiences, and one particular form of these experiences is what we call *conscious experience*. In other words, we propose that consciousness is a non-trivial composition at different levels of the biological closure defined above. Therefore, the organization of consciousness requires the organization of life. It responds to the increasingly influential view that crucial processes for consciousness cut across brain-body-world divisions, rather than being mere brain-bound neural events (cf.[Thompson and Varela, 2001]).

The lived experience is to be seen as irreducible, since any perception of a world is enacted *co-dependently* through the system's biological organization [Thompson, 2007]: Our biological

organization shapes the world we experience. In this context, conscious experience becomes a composition of living experiences, such that they are mutually defined and form two modes of the same process of closure.

Conscious experience entails a close loop of brain functional activity [Llinás and Paré, 1991, Llinás, 2003], but also a closure at the metabolic and structural level, creating a sense of the interaction in question [Thompson, 2007]. The experience becomes a conscious experience when the activity of the system returns as meaningful benefits to the whole system, i.e. the system not only enacts the environment (autonomy) but projects ”intuitions” that in a short temporal scale reward the internal organization and at long term convey future additional benefits to the system.

The whole multibrane and the parts are mutually defined. In other words, the operational and metabolic closure is partially inherited from their components, but importantly, although the closure is the same operation, the objects of that closure are different. The closure of cells and neurons is at the level of molecular components (self-production of components), while the closure of biobranes and autobranes is at the level of dynamical organization (systems that self-reproduce their organizational complexity). The composition between these two closures generates a biological autonomous system which is both operationally closed and structurally coupled with its environment [Thompson, 2007]. The new compounded system is autonomous in a new form. The whole multibrane system compounded by autonomous biobranes and autobranes, self-sustains and self-produces its structural and organizational dynamic through closed interactions at various levels.

5.4.3 Aspects of Consciousness

We suggest that these special types of compositions, illustrated via compositional interactions of autobranes and biobranes, are involved in the co-arising of different aspects of conscious experience as a new closed system: Conscious experience co-arises with global interactions of biobranes and autobranes.

Proposition 3 *Compositional interactions of autobranes, biobranes and branes co-arise with aspects of conscious experience.*

The multibrane organization for living systems is called into play with the main motivation of unifying phenomenal and access consciousness in one single structure. For example, interactive biobranes and autobranes create new topological paths trough inter-interactions (Figure 5.4A), generating dynamical phase transitions [Werner, 2012, Werner, 2013]. Then, the global unconscious experience defined in humans may relate to the first configuration of non-interacting autobranes, followed by a first transition where two or more autobranes start to interact. As soon as more biobranes and autobranes get involved, a second transition is defined, until all possible biobranes and autobranes under interaction form a global and consistent topological new multistructure (Figure 5.4A). These transitions are triggered by biobranes interactions and co-arise with phenomenal and access aspects of experience.

These biobranes dynamically evolve and interact, like two overlapping amoebas or cellular membranes. Some of these interactions trigger transitions that may correspond to relevant biological processes for the living organism. Others may correspond to aspects of consciousness such as wakefulness, phenomenal consciousness, subjectivity, access consciousness (i.e. knowing about the content of experience) and metacognition (i.e. the capacity to inform about the processes on the contents of experience, knowing that I know). Once these reciprocal actions have emerged, each biobrane monitors the others without any biological dominance among them. It implies that if one process disappears, the awareness associated with that process in the whole system also disappears.

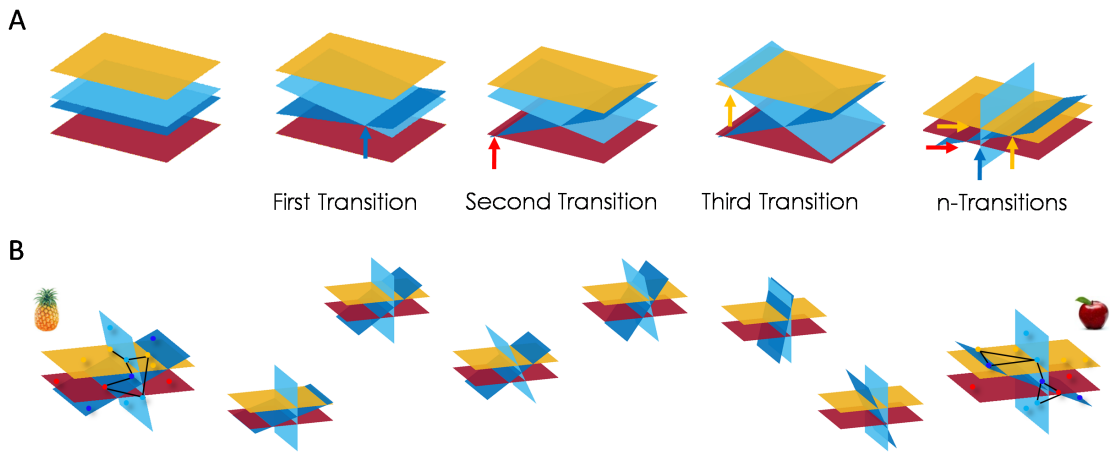


Figure 5.4: Consciousness interaction and phenomenology. (**A**) Different biobranes and autobranes compositional interactions generates different transitions and new local-global systems. First transitions may correspond to wakefulness, followed by transitions representing phenomenal experience, awareness, and other transitions related to more complex phenomenological experience, access consciousness and self-reference. (**B**) Different dynamical biobranes configurations, number and types of biobranes involved, degrees of interaction, regions of an intersection, types of oscillation, among others, would correspond to the phenomenology of conscious experience. Here, an example of autobranes as network configurations from the content of pineapple to apple, and their dynamical changes. The in-between configurations represent the dynamical evolution of these layers: the blue layers change the position while the others remain fixed. These changes inform about other states.

5.4.4 Phenomenology of Consciousness

Following the living metaphor, biobranes care about its processing as part of its biological requirements. Since biobranes are living structure, their interactions may express the dispositions and preferences of the whole agent [Cleeremans, 2011], like any animal interaction exhibits their dispositions and preferences. To satisfy their biological requirements, biobranes care about some states more than others, such as one cares about some temperatures that correspond to original survival living preferences. These preferences result from learned biobrane interactions and the respective co-emerging experience, such that each biobrane also cares about the processing of those other biobranes which may directly affect them. It is analogue to any living being acting and reacting to different stimuli and contexts to ensure its survival. The whole system and its parts try to balance out their coupling and decoupling, as part of their biological demands.

It generates a notion of phenomenal experience, expressed in the next proposition:

Proposition 4 *The types of biobranes, autobranes and the degree to which they interact with each other co-arise with the structure and content of experience.*

In other words, depending on the (i) degree and types of interactions, (ii) number and types of interacting biobranes, (iii) dynamic zones of an intersection, and (iv) oscillatory mechanisms involved, a certain experience co-arises (Figure 5.4B): the whole system feels one or another feeling, the experience evolves in one or another form, the content is about one or another element. This experience is not, however, unidirectionally dependent on these biobranes interactions, but both, experience and biobranes co-determine each other. The structure and content of experiences depend on the degree, types etc. of biobranes and autobrane interactions, but also the structure and content of experience affect the degree, types, etc. of the interactions on which they depend.

Other approaches, such as harmonic modes to conscious states [Atasoy et al., 2017], also suggest that oscillations giving by their harmonic structure are related to specific phenomenal experiences, but this approach ends up reducing that experience to brain interactions only, instead of emphasising its co-dependence character. In our case, emergent properties are as important as sub-emergent properties of systems (”downward causation”). In other words,

different oscillatory modes may correlate with phenomenal experience, e.g. resonance between different branes may involve access consciousness, while dissonance, lack of that access, or even further, representing different moods of the conscious animal. However, experience also triggers those oscillations in a co-determined balancing of branes interactions and experiences, creating a new form of oscillatory plasticity where sub-emergent properties play a crucial role [Rodríguez, 2008].

5.5 Implications and Predictions

In our closed biobranes composition framework (CBC), conscious experience is understood as a process which mainly interferes with internal brane integration in favour of global flux of activities and influences among biobranes. First, the activity of independent molecular, cellular, glia, and neural biobranes would correlate with an *unconscious* stage. Second, the conscious stage co-arises with the activity of now compositional interacting biobranes as the non-trivial composition of a new whole system. Both systems as a composition of closed biobranes form a new operationally closed whole multibrane. Therefore, conscious processes are related to the closed activity supported in particular by cellular self-generated activities [Llinás and Paré, 1991, Llinás, 2003] (in autobranes) and in general by biological autonomous systems (in biobranes).

5.5.1 Experimental Implications

The model predicts the uniqueness of the multibrane division (Figure 5.5A). A mathematical approximation of autobranes is the use of layers in a multilayer network [Signorelli and Joaquin Diaz Boils, 2020]. The conditions of coupling and splitting layers mathematically require a unique set of layers (see [Signorelli and Joaquin Diaz Boils, 2020] for details). This layer division implies a unique criterion, in the context of conscious experience, to parcel the brain and body in terms of anatomical, metabolic, and functional biological membranes. Eventually, this approach may overcome limitations of current brain divisions [Arslan et al., 2018]. Some of the autobranes/layers may represent the dynamical activity of cortical layers, sub-cortical regions, different types of cell assembly, and probably also molecular gradients acting as autonomous systems.

Identify these autobranes is not easy, but it is possible. The task requires the isolation of different brain regions, different types of cells and brain-body systems considering anatomical, metabolic (energetic exchange), and functional aspects. One empirical approximation is functionally isolating regions using anaesthetics. For example, recent evidence identify the cortical pyramidal cells in layers V and its modulation of brains states associated with consciousness (see [Suzuki and Larkum, 2020]). Another alternative is to approximate autobranes with brain-waves incorporating body interactions. In this case, we need to isolate brain-waves and identified them with fix brain regions during deep sleep, like harmonic brain modes [Atasoy et al., 2016], but with further functional restrictions. Then, we can measure their intrinsic changes, couplings and splittings during awake and other conditions. A simpler functional approximation is to identify the intrinsic oscillation of brain regions during deep sleep together with their anatomical connections, using techniques of multilayer dynamical models [Cabral et al., 2017]. Then, we can group similar inner frequencies in layers and study their evolution across other conditions. More detailed methods may target other aspects of autobranes, and new equipment such as deep optical modulation (e.g. fast high-resolution two-photon microscopy [Zong et al., 2017]), higher fMRI resolution (e.g. 11.7 Teslas, [Nowogrodzki, 2018]) and layer-fMRI analyses [Huber et al., 2020] may help to find and define these empirical autobranes and their inner activity.

Another prediction of our framework is the direct relationship between the content of experience and number of biobranes and autobranes involved (Figure 5.5B). According to complex systems theory, more configurations of interactions implies more complexity. Therefore, in our model, richer or detailed perceptual experiences would involve more biobranes. Simultaneously, having more biobranes implies the possibility of richer experiences. As such, these biobranes and autobranes are restricted by, as well as restrict different aspects of the environment [Signorelli et al., 2020a]. Consider a hypothetical example: the human cortex is divided, by convention, into six anatomical layers. Some anomalies are related to the disruption of these layer's configuration [LoTurco and Booker, 2013] (but see [Guy and Staiger, 2017]). Also, their density and structural changes among those layers are associated with a distinctive marker of human cognition [DeFelipe, 2011]. If some of these layers correspond to the more complex autobranes defined above (not just to the anatomical layer division), it may support the common assumption that human experience conveys richer and detailed contents in comparison with other animals that show traces of less developed cortical layers [DeFelipe, 2011]. A testable hypothesis would state that some of these structural and functional cortical organizations may

correspond to the autonomous biobranes defined by our conceptual model.

The model may also inspire new measures of consciousness as the degree of biological autonomy on each biobrane A_l and the whole multibrane system A_w (Figure 5.5C). Living systems produce more of their complexity compared with what is produced by their environment. Biobrane autonomy A_{l_i}, with $i = [1, ...n]$, n number of biobranes, is then defined as the complexity of the biobrane C_{l_i} divided by the complexity of the environment C_w, that corresponds to the union of other biobranes forming the whole system. For that measure, a value greater than 1 would mean more autonomy [Fernández et al., 2014]. Similarly, the whole system autonomy is defined as $A_w = C_w/C_E$, where C_E is the complexity produced by the environment outside the biobranes union. These definitions assume a generalized version of autopoiesis and autonomy (operational closure) based on notions of information that allow us to consider systems that self-produce its organization instead of their components. Moreover, complexity is defined as $C_j = E_j * S_j$, where $j = [l, w]$. E_j represents emergent properties as a group of new chaotic patterns in the system, i.e. new properties of a system which are not present in their elements. S_j is the system self-organization, in the form of organized patterns that appear from local patterns interacting [López-Ruiz et al., 1995, Fernández et al., 2014, Gershenson, 2015]. Therefore, a high value of complexity requires a balance between emergence (chaos) and self-organization (order) [López-Ruiz et al., 1995]. This consequence is equivalent to the required balance of integration and segregation/differentiation in the early version of integrated information theory (IIT). However, in our model, complexity by itself is not enough to define the conscious capabilities, as well as these interactions, are not only neural events.

Consciousness would require biological autonomy, specifically a decrease of A_l while increase in A_w. This recovers the observed increase of complexity values on the whole system as a consequence of decreasing the biobrane autonomy $C_w = C_l/A_l$. Accordingly, A_w values increase as the biobranes start to interact and compound the multibrane structure. The usual interpretation in the form of correlated activity brings the prediction that interconnectivity between autobranes increases with awareness, i.e. biobranes become coupled and having different types of influences among their dynamics: correlations between autobranes and their nodes increase. On the contrary, inside autobranes, intraconnectivity would decrease as awareness increases. Network analyses in physiological data and brain-body coupling systems seem to support these predictions [Bashan et al., 2012, Stankovski et al., 2016]. Taking these biobranes and autobranes as independent living systems that interact to provide and obtain biological re-

sources, this conclusion is not surprising. During their interactions, both systems depend on each other (signifying conscious processes). After these interactions take place, they need to come back to their intrinsic independent non-interacting activity, and therefore recovering their autonomy (signifying sleep states).

Measures of autonomy and the multibrane/multilayer structure also support multidimensional measurements of consciousness [Bayne et al., 2016]. For example, differences between sleep stages and anaesthetics transitions would end into different values of that autonomy levels, among the different biobranes that are being affected. During deep sleep, it may be that biobranes and autobranes are naturally disconnected (Figure 5.5E) or the connection has a different structure. During REM, some biobranes become disconnected, but others may still interact (Figure 5.5F). In the case of different anaesthetics, some of them may act in part of the system's branes but not in others; generating various forms of splitting biobranes (Figure 5.5G-H). During non-conscious conditions, such as sleep stages and different anaesthetic, one biobrane may drive the interaction, breaking the balance requirement and reducing values of A_w. For instance, anaesthetized patients will present low values of A_w, while awake subjects will have maximal values according to a fix and normalized scale. Moreover, the sudden emergence in elderly subjects is linked here with an intrinsic change on the autobranes orientation that impacts on the functional distance between biobranes. It makes the biobranes suddenly interact once the effect of anaesthetics starts to decay, avoiding the usually smooth time transition. Asymmetries between anaesthetics induction and emergence (hysteresis) are explained by their actions on one or another biobrane with different intrinsic dynamical properties, that impact differently on the rest of the biobranes and trigger dynamical transitions in different order. Then, the global recovery is influenced by this biobrane recovery, while the others remain intact. Therefore, the coupling and splitting of that biobrane generate differences in the time and concentration effect of anaesthetics. So the type of brane, orientation and system disrupted play a crucial role in the model, and local and global autonomy values, as degrees of interaction between different biobranes, would form a global multidimensional consciousness quantifier [Bayne et al., 2016].

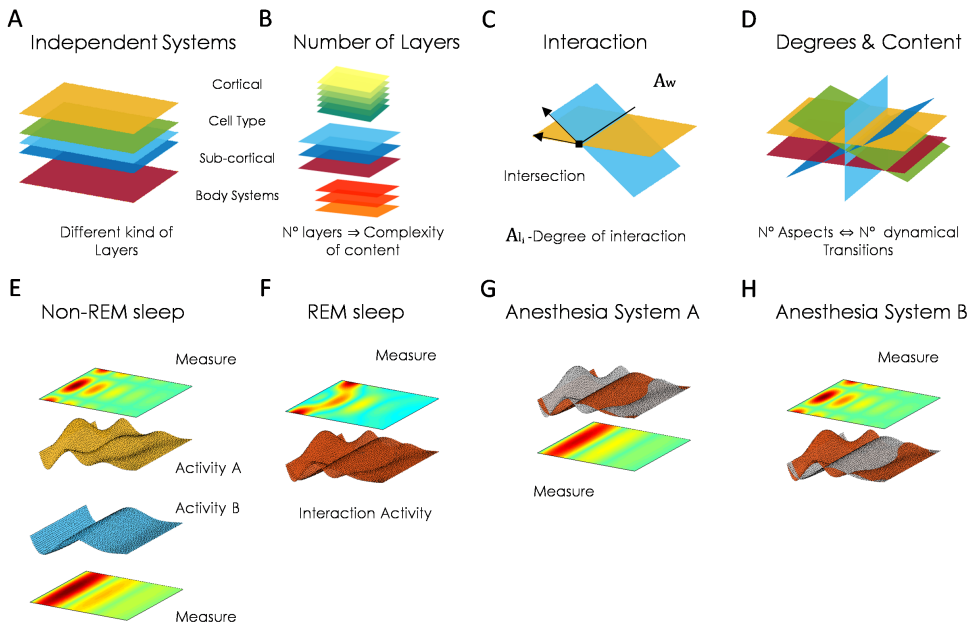


Figure 5.5: Implications and predictions. (**A**) The model predicts a unique multibrane division. In addition to neural assemblies, this division also includes other cell types and more general body systems. (**B**) More biobranes imply more richer and detailed content of experience. (**C**) Interaction between biobranes as a multidimensional measure of consciousness (local and global biobrane autonomy) in the context of disorders of consciousness, anaesthesia protocols, sleep and psychedelics. (**D**) Spatial-temporal interactions (place, biobranes involved) and mechanisms (resonance, superposition, among others) may account for the details of phenomenological experience and distinctive/unique subjective conscious experiences. (**E**) A two-brane-system configuration in non-REM sleep. (**F**) A two-brane-system configuration, interacting during REM sleep. (**G**) A two-brane-system configuration, A and B, anaesthesia affects only one of the systems (system A). (**H**) A two-brane-system configuration, A and B, anaesthesia affects the other system (system B). (**E**),(**F**),(**G**) and (**H**) are toy examples.

As a consequence, a breakdown of these biobranes and interacting structure implies a disruption of the usual conscious experience. Due to this organization, some sections of the biobranes may participate in more intersections than other regions. Therefore, their intrinsic functional activity being notoriously disrupted under conditions of global rearrangement such as chronic disorders of consciousness and/or sleep [Signorelli et al., 2021a]. It may generate the impression that some areas define a hierarchy of active regions or hubs, such as proposed by global neural workspace theory (GNW). Once biobranes naturally decouple, these hierarchies would appear as disrupted, but in reality, they go back to their intrinsic dynamic. For multibranes, it means that any dynamic impairment of a biobranes will lead to the awareness associated with that biobranes disappear. This impairment may also affect the global biobranes balance through decoupling some biobranes. Therefore, the local causal driven forces that generate similar global disruptions would correspond to disruptions on different biobranes. These disruptions become a common mechanism for loss of consciousness, at the same time that they save the specificity of different impairments. In other words, independently of the molecular pathways of different anaesthetics, stages of sleep or localization/types of brain injuries, the dynamical disruption correspond to a re-arrangement of biobrane organization distant from the

awake condition. In summary, a breakdown of seemly hierarchical organization indeed corresponds to break of natural balance from a conscious interacting multibrane structure to an unconscious partially non-interacting structure

5.5.2 Theoretical Implications

The multibrane framework forces us to specify: i) the type of organizational structure (types of networks, multinetwork, membranes, etc), ii) the components of these structures, iii) the interactions among components, and iv) the explicit mechanism supporting conscious experience [4]. This situates our model within the discourse of current models of consciousness. For instance, the commonalities and differences with IIT and GNW are summarized in Table 5.3.

Model	Structure	Components	Interaction	Consciousness
IIT	Monoplex-Time evolving	Physical Systems	Cause-effect interactions	Maximal causal integration
GNW	Multilevel-Time evolving	Neurons	Inter-area Action Potentials	Global broadcasting
CBC	Dynamical Multibrane	Biological Autonomous Systems	Multiple Interactions	Closed composition of Biobranes

Table 5.3: Comparison of three models of consciousness according to a multilevel interacting framework.

For example, the GNW and IIT are neither embodied nor compositional models, since they reduce consciousness to only properties of neural interactions. On the contrary, our discussion proposes that the relevant interactions to conscious experience are found on the multiple levels of the organizationally closed biological system. It entails multiple types of interactions. Contrary, GNW emphasises the role of broadcasting electrical activity from certain areas of the brain to other areas [Dehaene and Changeux, 2005, Van Vugt et al., 2018], and IIT focuses on intricate mechanisms of causal integration at the level of assemblies and single neurons [Oizumi et al., 2014]. A multibrane extended framework subsumes the network structure of these two models: while GNW seems more general than IIT, our model can be seen as an extension of GNW but avoiding functionalism and philosophical reductionism. Recently, new dynamical approaches also suggest that other new physical principles are playing a relevant role in the conscious activity. Among them are harmonic modes [Atasoy et al., 2017] and criticality [Werner, 2013, Tagliazucchi, 2017]. Following our general framework, we believe the concepts introduced here may integrate all these principles while emphasising the importance

[4]Mechanisms of consciousness applies for reductionist approaches. In our framework, experience and living are at the same level, therefore a reductive mechanism does not fully apply.

of formal biological definitions in models of consciousness [Signorelli et al., 2021a]. This is a crucial step forward to define a formal model of consciousness, namely, models which make explicit their theoretical and experimental assumptions [Kleiner, 2019].

5.5.3 Philosophical Implications

Our approach also has philosophical implications which make it a promising model against other scientific models of consciousness. First, it is a *nondualistic framework*. Second, it acknowledges the *primacy of embodiment*. Third, it acknowledges the *primacy of consciousness*. Fourth, it is *pragmatic*. In this subsection we briefly discuss why our framework is a relevant conceptual apparatus.

The Nondualist Framework

Our model starts from a *radical embodiment reformulation* of the mind-body problem [Thompson, 2007, Thompson, 2014]. The mind-body problem arises when theories of consciousness assume the primacy of substance-like ontological objectivity, i.e. elements having a uniform and constant properties. Therefore, the focus is on those physical parts, cells, neurons, regions of the brain, from which the experience is thought to emerge as a whole [Searle, 2000]. However, the existence of subjective ontologies seems not reducible to objective ones (e.g. the redness of the red is not yet explained by the light wavelength or any elaborated neural event). Then, the concept of qualia is coined to extend the same uniform and constant substance ontology, but this time for subjectivity. This strategy leads us to irreconcilable mind (subject) and matter (object) ontologies. Rather, our model makes use of process-ontologies [Rescher, 2012], i.e. what exists are processes of transformations (as opposed to constant substances). These ontologies seem better suited to explain the idea of the living body and the lived body (consciousness) as two modes of appearance of one and the same body: mind and body are not separate. Mind and body are different modes of the same process: closure. This closure leads us to the property that makes living differently than nonliving and consciousness different than unconsciousness. In consequence, any mathematical or computational structure/architecture supporting consciousness should constitute and realize consciousness, as well as its components, constitute and realize living being.

The Primacy of Embodiment

Our model builds on the idea of *primacy of embodiment*, implying two important aspects. First, we proposed that a sound and scientific model of consciousness must build on a principled definition of the living body. Current models treat biological processes (autonomy) as mere physical processes (heteronomy). Rather, in our model, biological processes always involve organizational closure, with all its implications for a proper paradigm of biological interactions [Maturana, 2011, Maturana and Varela, 1998]. Second, our framework builds on the notion of *embodiment* of neural activity. Current models reduce consciousness to brain states. Rather, in our approach, consciousness is related to multidimensional biological processes which go beyond mere neural events. Via the mathematical machinery for biobranes and autobrane, our model may allow future mathematical precise descriptions of biological autonomous processes across various types of cells of the living system.

This embodiment has strong implications for artificial life and consciousness. Materialistic approaches imply that replicating the *computational architecture* of neural systems convey artificial consciousness [Dehaene et al., 2017, Tononi and Koch, 2015]. We reject this conclusion unless the unique biological features of biological autonomous systems and the organizational multibrane structure are primarily replicated. If the biological autonomy is replicated as an essential prerequisite to achieve consciousness, life needs to be replicated as requisite of biological autonomy [Signorelli, 2018a, Signorelli, 2018b]. In this context, a thermostat is ”less conscious” than a rat not because of its information capacity, or because it misses all the relevant functional architecture and electrochemical mechanisms of a neural workspace. The thermostats or any non-living being is unable of awareness just because they are non-living systems and as such, they lack biological autonomy. At the end of the day, only living beings seem to have conscious experiences, hence the living organization is indeed important.

The Primacy of Consciousness

Our model builds on the idea of *primacy of consciousness*. There are two important aspects. First, consciousness is ineliminable and cannot be reductively explained in terms of brain states or even the living body. While current nondualist models reduce consciousness to electrochemical processes, our approach does not. Rather, biological processes and consciousness

dynamically co-arise. Therefore, they are co-dependent: Consciousness depends on biological processes, but it also affects the biological processes on which it depends. Second, many interpretations of the primacy of consciousness subscribe to dualism or panpsychism. Our model does not. Dualism says that "the physical" and "the mental" have distinct natures. Panpsychism says that every physical phenomenon, intrinsically, carries some measure of consciousness. However, our model is neither dualistic nor panpsychism. Rather, it builds on a non-dualistic framework in which physical being and experiential being imply each other [Thompson, 2014]. Mind and body are (i) neither separate, (ii) nor only mind, (iii) nor only body. Our approach is a non-dualistic framework that acknowledges the simultaneous primacy of embodiment and consciousnes.

Pragmatic Usefulness

The philosophical attitude supporting our model is strongly influenced by the modern approach to formal mathematics after Gödel's incompleteness theorems [Gödel, 1931]. It means that any set of axioms is a useful set of axioms in the context in which formal theories are based on, but the same set of axioms may not be useful in another theoretical contexts [Gershenson, 2013a]. It does not mean that these axioms are true or false, as well as Newton's laws are not true or false because they work at some scale but not at others (e.g. macro-scale and micro-scale). On the contrary, neuroscientific models of consciousness assume that something true, objective and invariable is said about consciousness and its mechanisms. However, the operation of observing distinguishes between what we, as observers, can say about any system that appears to us from what we say about what may occur in the internal operation of the observed system [Maturana, 2011]. There is nothing obscure or restricted to the micro or atomic world, we only observe transformations, and it becomes essential to any introspective operation of observation: a feature which is not objectively measurable. As pointed out by Box, "all models are wrong" [Box, 1976], and as emphasized later, "but some are useful". Hence, models are descriptions of the modelled phenomenon, and as such, models depend on the observer [Gershenson, 2013a, Maturana, 2011].

Therefore, we just intend to introduce a useful framework guiding further experimental hypotheses through useful axioms and definitions [Signorelli and Joaquin Diaz Boils, 2020]. Different than other models, our framework addresses the complexity of the conscious phenomenon, not through reducing the system to a certain group of ontological laws or compo-

nents, but composing abstractions and proving its power explaining and predicting new features of the phenomenon in question. In other words, our model is a pragmatic phenomenological approach, not an ontological one. It makes our conceptual apparatus a new promising approach to the biology of conscious experience.

5.6 Conclusions and further work

Across these pages, we introduced a conceptual apparatus to explore the biology of conscious experience. We invoke biological closure to rescue the living structure as an essential requirement of radical embodiment. Compositionality is called into play to discuss different levels of closure and its relationship to a multilevel structure in the brain-body. Biobranes and autobranes are conceptualized to model these compositional closures in the context of conscious experience. This framework centres on the living and its unique organizational structure: i) the co-dependency between the living body and the lived body, and ii) a multilevel organization to reconcile different brain-body signatures of experience. This multilevel organization accounts for the various biological processes, cell types and biological systems that relate to consciousness. As such, our approach is simple and eventually a mathematical theory inspired by enactive and embodiment approaches to conscious experience.

Moreover, our model subsumes some concepts from previous network models [Dehaene and Changeux, 2011, Tononi et al., 2016], but surpasses them by making a clear distinction regarding the biological definition of the elements that can form interacting biobranes and what it means for them to be compositional. It turns to be a clear advantage over other models: Different from previous models of consciousness, our framework states co-dependencies between brain-body and experience, avoiding reductionism.

In future attempts, we expect to develop the mathematical and empirical machinery to test the main propositions and predictions. It might consider biological autonomy and closure at different levels. Operational definitions of biobranes and autobranes are a crucial step forward to implement biological autonomy as a local and global measurement of the degree of brane interactions and therefore, of multidimensional signatures of consciousness. Moreover, phenomenological approaches such as neurophenomenology [Varela, 1996] and microphenomenology [Petitmengin et al., 2019] shall be at the centre of that testing, specifically to

test the relationship between biobranes interacting and the phenomenology of conscious experience following our last proposition. We are aware that, all together, it conveys an ambitious research program.

Finally, we expect that some of the concepts introduced across these pages inspire new theoretical and empirical models of consciousness. Importantly, these concepts offer potential answers to the motivational questions at the beginning of this article: i) biobranes may define relevant brain-body regions and interactions, ii) conscious experience does not emerge, but co-arises with compositional closed interactions in a living multibrane structure, and iii) machines are not conscious unless they replicate the compositions of closure, from living to consciousness.

We believe that the only way to solve the apparent gaps between body and mind is through integrative models and therefore through new metaphors for biological neuroscience of consciousness.

Chapter 6

CONCLUSIONS

Across this work, we have discussed a classification for models of consciousness based on their explanatory profiles. Considering the consequences of such classification, a multilayer framework is introduced to unify empirical evidence and suggest theoretical tools to solve models' controversies. Another consequence points out the need for integrative and pragmatic computational tools. One example using the measure of Intrinsic Ignition is implemented. The main result of that implementation is the organizational hierarchy (measured by Intrinsic Ignition) as a novel and simple form to quantify disruptions of consciousness. Finally, the main ideas of our approach are explored in light of the radical embodiment and their empirical but also philosophical consequences. In such framework consciousness co-arise with entangled mechanisms of brain-body-world interactions, which conveys a renovated theoretical and empirical research project for decades to come.

In light of the previous chapters, the empirical testing of models of consciousness is far from trivial. The testing of models of consciousness is not only limited by current scientific techniques, but also by the way we try to confront and/or unify those models. Conclusive experiments might need to incorporate the fundamental assumptions and dynamical processing of the whole system, instead of experimental focus at the level of neural mechanisms only.

For instance, no one would doubt the empirical and proved relevance of global signatures of ignition (GNW) and recurrent activity (RPT). However, ignition and recurrent activity by themselves are not enough to explain the differences between the experience of one colour or another. Perhaps more importantly, those mechanisms are not sufficient to recover conscious experience as we know it, for example, from chronic disorders of consciousness. There is a complex anatomical, functional and metabolic machinery that also needs to be recovered in

order to support further activities such as ignition, resonance and integration. This is recognized by almost all authors of the main theories reviewed in the first chapter, and still many of them insist on the reduction of consciousness to only one clear necessary and sufficient mechanism.

Another issue in the neuroscience of consciousness is that we tend to assume that lower levels of consciousness are associated with vegetative states (VS), minimally conscious states (MCS), sleep, etc, and higher levels with awake states [Bayne et al., 2016]. However, psychedelic states present a paradox [Bayne and Carter, 2018]. They seem to enhance brain activity and some aspects of conscious experience (such as the experience of unity, perception), but impair others (cognitive functions). The assumption that consciousness comes in degrees becomes a problem, we can no longer attempt to measure global changes in consciousness using one single dimension. One alternative is to invoke multidimensional approaches [Sergent et al., 2017].

Nevertheless, the concept of multiple dimensions of experience is a direct challenge to all the current models of consciousness [Bayne and Carter, 2018], with the slight exception of embodied, PP&I and TTC theories. In the first and second case, embody and interoceptive signals might become multidimensional aspects of consciousness, while in the third case, different types and frequency ranges of neural activity are claimed to specify at least two dimensions of consciousness: temporal and spatial.

This is, however, only the surface. The real criticism roots on the notions of objectivity versus subjectivity, quantity versus quality or third versus first-person accounts. Subjective experience seems to be not gradable. In words of Bayne and colleagues ”one person can be conscious of more objects and properties than another person, but to be conscious of more is not to be more conscious” [Bayne et al., 2016]. Most models of consciousness equalize this subjective experience with objective neural mechanisms and measures. The problem here is that the objective neural or more general body signals might not correspond to a ”one to one relationship” with the experience of the associated content [Noë and Thompson, 2004], neither the structure-function mapping in the brain supports that simplistic treatment [Pessoa, 2014].

In modern words, this problem is reformulated as representation and non-representational models. Representation theories such as GNW, HOT and RPT assume an outside objective world brought into the organism via transduction of signals and hypothetical neural coding. Unfortunately, that neural coding is never specified, it is extrinsic to the neural system (role

of the observer) and presents many philosophical but also empirical challenges [Brette, 2016, Brette, 2017]. On the other side, non-representational theories like IIT, TLC, ERE, and CAN, assume that perception comes from internal constructions, either of biological organization, internal physical structure or because they are fundamental (respectively). These constructions are only secondarily shaped, or modulated, by external stimulation. Those models convey a primacy of experience, i.e. the objective reality is a construction from the subjective one, but unfortunately, there is no way to measure the entire system to falsify their deeper postulates. And yet, other models such as TTC and PP&I sometimes adhere to representationalism, others times to non-representationalism, or even new versions such as "self-representations", whatever the real philosophical commitment that might imply.

In light of the previous chapters, a new relevant distinction might involve contextual and non-contextual models. In the first case, the measure of the phenomenon of study changes according to the type of measurement. Metaphorically, we say that a measure (context) influences the output of that measurement. In the second case, independently of the measurement context, the output is always the same. In physics, quantum particles behave as contextual entities while classical objects as non-contextual. Cognition and language also seem to present some degrees of contextuality.

Most models of consciousness assume that consciousness is not contextual (with exception of embodiment theories), however, scientists distinguish between phenomenal and access consciousness. Indeed, as reviewed above, some neural signatures are interpreted as evidence for one or another type of conscious processing [Block, 2005]. In order to save the non-contextual character of consciousness, these signatures are thought of as two different phenomena. Nevertheless, a more economical approach would be to recognize the contextual dimension of conscious experience and assume that phenomenal and access consciousness are two sides of the same coin: conscious processing behaves differently according to the way we measure it.

If it is correct, models such as GNW, IIT, among others, are incomplete and require further integration [de Barros et al., 2017, Northoff and Lamme, 2020]. This incompleteness is already recognized by the authors of these models, although with slightly different wording. For instance, in [Oizumi et al., 2014], the authors explicitly mention five limitations that make their mathematical account incomplete for a robust theory of consciousness, e.g. definition of the optimal Spatio-temporal level of description, implementation of continuous signals, etc. In [Dehaene and Naccache, 2001], the authors explicitly say "A full theory will require many more

bridging laws to explain how these neural events organize into larger-scale active circuits, how those circuits themselves support specific representations and forms of information processing, and how these processes are ultimately associated with conscious report". To these important considerations, we add the contextual limitation of the current model of consciousness.

Finally, consciousness cannot be divorced from its quality or content. A full-fledged theory of consciousness needs to explain both the quantity and quality of consciousness. Several models tend to neglect the problem of quality and focus solely on what differentiates conscious vs unconscious systems. The problem of quality in such cases is delegated to the external world (i.e. sensory cues feel the way they do solely because they *carry* their quality from the environment, or stimulate *correct*, labelled receptors). If proponents of the theories differ significantly in terms of their explanatory target without realization, they are more likely to talk past each other, instead of fostering a theoretical synthesis.

Our suggested theoretical and eventually empirical framework (chapters 3 and 5) targets both quality and quantity, without committing to a reductive account. It is neither causal nor functional since it supposes co-dependency and causal-effective influence of conscious experience on its physical substrate (co-arising). This integrative model conveys both mechanistic and unificationists types of explanation since locally, it suggests mechanistic layers interacting, and globally, it unifies not only signatures and theoretical ideas about the conscious experience but also uses relevant concepts from other fields such as mathematics and physics.

Appendix A

SUPPLEMENTARY QUOTES

Table A.1: Mode of explanations. Mechanistic versus Unification.

Model	Quotes	Classification
CK	"...all the different aspects of consciousness (pain, visual awareness, self-consciousness, and so on) employ a basic common mechanism or perhaps a few such mechanisms" [Crick and Koch, 1998]	Mechanistic
DC	"It allows us to pay attention to the hard enough problem, which is to formulate a global theory that provides neural bases for the general and special features of consciousness."[Edelman, 2003]	Unificationist
TCL	"... the assumption that the intrinsic electrical properties of neurons, and the dynamic events resulting from their connectivity, result in global resonant states, which we know as cognition [and consciousness]" [Llinás, 2003] "we may conclude that consciousness is a noncontinuous event determined by synchronous activity in the thalamocortical system" [Llinás, 2003] "We propose...that consciousness, like locomotion, might be more a case of intrinsic activity than of sensory drive" [Llinas et al., 1998]	Mechanistic
ERE	"We also propose that the processes crucial for consciousness cut across the brain-body-world divisions rather than being located simply in the head" [Thompson and Varela, 2001]	Unificationist
BE	"Combining these observations with our quantum mechanical analysis of bouton exocytosis, we present now the hypothesis that the mental intention (the volition) becomes neurally effective by momentarily increasing the probability of exocytosis in selected cortical areas such as the SMA neurons" [Beck and Eccles, 1992]	Mechanistic
OrchOR	"It is to be expected that the actual mechanisms underlying the production of consciousness in a human brain will be very much more sophisticated than any that we can put forward at the present time, and would be likely to differ in many important respects from any that we would be in a position to anticipate in our current proposals. Nevertheless, we do feel that the suggestions that we are putting forward here represent a serious attempt to grapple with the fundamental issues raised by the consciousness phenomenon, and it is in this spirit that we present them here." [Hameroff and Penrose, 2014]	Mechanistic

HOT	”We understand something only when we can explain it, and explaining a natural phenomenon typically if not always means locating it in its distinctive causal nexus.” [Rosenthal, 2008]	Mechanistic
IIT	”IIT provides a principled explanation for several seemingly disparate facts about the PSC. (...) why is consciousness generated by the corticothalamic system - or at least some parts of it, but not by the cerebellum, despite the latter having even more neurons? Why does consciousness fade early in sleep, although the brain remains active? Why is it lost during generalized seizures, when neural activity is intense and synchronous? And why is there no direct contribution to consciousness from neural activity within sensory and motor pathways, or within neural circuits looping out of the cortex into subcortical structures and back, despite their manifest ability to influence the content of experience? Explaining these facts in a parsimonious manner calls for a theory of consciousness.” [Tononi et al., 2016]	Unificationist
GNW	”tools of cognitive psychology and neuroscience may suffice to analyze consciousness. (...) Our view, however, is that conscious access is one of the few empirically tractable problems presently accessible to an authentic scientific investigation. We further hope that an understanding of the neural processes that lead to overt report will eventually result in a theory of covert acts of self- report, and thus may ultimately contribute to an explanation of the nature of our private phenomenal world.[Dehaene and Naccache, 2001]	Mechanistic
RPT	”That is the point of view of a science that goes beyond neural correlates of things we believe to exist introspectively or behaviorally. In this account, neuroscience is used to produce explanatory correlates (Seth, 2009) to arrive at a framework with maximal explanatory power regarding consciousness and its relation to other cognitive functions...Here, the behavioral and neural data are taken together to arrive at concepts that are better than the ones that can be arrived at by either psychology or neuroscience independently.” [Lamme, 2010]	Unificationist
PPI	”Even if a measure based on information length should turn out to be equivalent to existing measures of dynamical complexity, it would be advantageous, because it could unify existing approaches under a single overarching principle (i.e., the FEP). (...) Minimising variational free energy could count as a minimal unifying model: variational free energy must be minimised by every self-organising system that persists, and hence also by any conscious system.” [Wiese and Friston, 2020]	Unificationist

TTC	”While these views [other current models] presuppose and implicitly touch upon the brain’s own time and space, they do not consider time and space themselves - central dimensions of the brain’s neural activity - in an explicit way, that is, how the brain itself constructs time and space in its neural activity” [Northoff and Huang, 2017] ”The main and overarching aim of this review is to provide a unified hypothesis that directly links and thus integrates the different forms of neural activity with the different dimensions of consciousness” [Northoff and Huang, 2017] ”The TTC is primarily a neuroscientific theory of brain and consciousness, which may carry major philosophical implications in terms of a novel view of consciousness, and a paradigm shift from mind-body problem to world-brain problem” [Northoff and Huang, 2017]	Unificationist
CAN	”classical notion of an observer-independent ”objective” reality comprising spatially-bounded, time-persistent ”ordinary objects” and well-defined local causal processes must simply be abandoned (...) if we want to go beyond this ”applied science” and understand the true nature of the mind and the reality beyond it, we can’t look to neurons or brains...” [Fields et al., 2018]	Unificationist

Table A.2: Mechanism of explanations. Functional versus Causal.

Model	Quotes	Classification
CK	”A node, all by itself, cannot produce consciousness. Even if the neurons in that node were firing appropriately, this would produce little effect if their output synapses were inactivated. A node is a node, not a network. Thus a particular coalition is an active network, consisting of the relevant set of interacting nodes that temporarily sustains itself. (...) What could be special about this activity that reaches above the consciousness threshold? It might be some particular way of firing, such as a sustained high rate, some sort of synchronized firing or firing in bursts. Or it might be the firing of special types of neurons, such as those pyramidal cells that project to the front of the brain” [Crick and Koch, 2003].	Causal
DC	”A scientific view that assumes that consciousness arises from reentrant interactions among neural populations must therefore conclude that it is the neural activity of the dynamic core that is causal. If we call that activity C and the qualia it entails C, then it is C that is the cause of our actions and further C events” [Edelman, 2003]	Causal
TCL	”consciousness is an oneiric-like internal functional state modulated rather than generated, by the senses” [Llinas et al., 1998] ”consciousness (i.e. being awake and able to feel, judge and remember) is but one functional state of our brain” [Llinas et al., 1998] ”functional states such as wakefulness or REM sleep and other sleep stages are prominent examples of the breadth of variation that self-generated brain activity will yield” [Llinás, 2003]	Functional
ERE	”causal-explanatory relationship is one-way, from internal neural events to conscious experience. We propose a more expansive, two-way account...(1) that as a result of the generic feature of ’emergence’ in complex systems, one can expect there to be two-way or reciprocal relationships between neural events and conscious activity; and (2) that the processes crucial for consciousness cut across brain-body-world divisions, rather than being brain-bound neural events” [Thompson and Varela, 2001] ”First, there is local-to-global determination or ’upward causation’ as a result of which novel processes emerge that have their own features, lifetimes and domains of interaction. Second, there is global-to-local determination, often called ’downward causation’, whereby global characteristics of a system govern or constrain local interactions” [Thompson and Varela, 2001]	Neither causal nor functional
BE	This theory is based on dualism. The ”function-causal structure” distinction does not apply.	Neither causal nor functional

OrchOR	”Consciousness results from discrete physical events; such events have always existed in the universe as non- cognitive, proto-conscious events, these acting as part of precise physical laws not yet fully understood. Biology evolved a mechanism to orchestrate such events and to couple them to neuronal activity, resulting in meaningful, cognitive, conscious moments and thence also to causal control of behavior.” [Hameroff and Penrose, 2014]	Causal (on the basic level of proto-consciousness); functional with respect to higher forms of consciousness (via. orchestration)
HOT	”Causal connections are irrelevant here, since there need be no causal tie between a HOT and its target. Rather, HaTs result in conscious qualities because they make us conscious of ourselves as being in certain qualitative states, which results in the subjective impression of conscious mental qualities” [Rosenthal, 2002a]	Functional
IIT	”IIT then postulates that, for each essential property of experience, there must be a corresponding causal property of the PSC. The postulates of IIT state that the PSC must have intrinsic cause-effect power; its parts must also have cause-effect power within the PSC and they must specify a cause-effect structure that is specific, unitary and definite.” [Tononi et al., 2016]	Causal
GNW	”Many cognitive theories share the hypothesis that controlled processing requires a distinct functional architecture which goes beyond modularity and can establish flexible links amongst existing processors. (...) Here we synthesize those ideas by postulating that, besides specialized processors, the architecture of the human brain also comprises a distributed neural system or ’workspace’ with long-distance connectivity that can potentially interconnect multiple specialized brain areas in a coordinated, though variable manner (...) It would therefore be incorrect to identify the workspace, and therefore consciousness, with a fixed set of brain areas. Rather, many brain areas contain workspace neurons with the appropriate long-distance and widespread connectivity, and at any given time only a fraction of these neurons constitute the mobilized workspace.” [Dehaene and Naccache, 2001]	Functional
RPT	”Recurrent processing has high explanatory power in accounting for important features of conscious percepts, as there is a strong homology between the integrated structure of perception and the structure of recurrent processing.” [Lamme, 2010]	Causal

PPI	”Internal state transitions posited by a computational explanation must then not only be mapped to internal states of the system in question, but these internal states must also be causally connected to the sensory input and behaviour figuring in the description of the input-output patterns that are to be explained. (...) it is likely that some computing devices will never be conscious, regardless of which computations they perform (e.g., desktop PC). Such systems might simulate consciousness, but will never be conscious. (...) not all virtual machines that perform approximate Bayesian inference by encoding a probability distribution over external states, given blanket states, are realised by physical machines with the same Markov blanket partition. If we keep this mind, we can maintain a distinction between simulating and instantiating consciousness, but still retain the hypothesis that the right computational properties are sufficient for consciousness (Chalmers, 2011), if they are instantiated by the right kind of system.” [Wiese and Friston, 2020]	Causal
TTC	”There may thus be a ”neuronal-phenomenal correspondence” between phenomenal features and the temporo-spatially expanded stimulus-induced activity...while conceptually it may be described as ”isomorphism” [Northoff and Huang, 2017]	Causal
CAN	”the limited syntax of the CA formalism is sufficient to implement memory, predictive coding, active inference, attention, categorization and planning. These functions emerge naturally, moreover, from asking what structure an RCA [i,e. a ”reduced conscious agent”] must have in order for its perceptions to be useful for guiding action within the constraints imposed by ITP.” [Fields et al., 2018]	Neither causal nor functional. Consciousness is fundamental; all cognition arises as function of consciousness (but consciousness is not itself a function of anything)

Table A.3: Target of explanations. Quality versus Quantity of consciousness.

Model	Quotes	Classification
CK	”The primaries and the early events in an election would correspond roughly to the preliminary unconscious processing. The winning coalition associated with an object or event would correspond to the winning party, which would remain in power for some time and would attempt to influence and control future events.” [Crick and Koch, 2003]	Quantity
DC	”The framework position I have taken here is that consciousness consists of qualia, by which I mean not just isolated submodalities of red, warm, etc., but also complex scenes, memories, images, emotions; indeed, the entire rich panoply of subjective experience. If, as I have suggested, the neural systems underlying consciousness arose to enable high order discriminations in a multidimensional space of signals, qualia are those discriminations. Differences in qualia correlate with differences in the neural structure and dynamics that underlie them. Thus, for example, olfactory neurons and their circuits differ from retinal neurons and circuits, and such differences seem sufficient to account for differences in their respective qualia. ”[Edelman, 2003]	Quality and quantity
TCL	”Perhaps the most spectacular diference concerning global brain states is that between wakefulness and dreamless sleep...these two events must be electrical in nature given the large number of elements involved; electrical in the sense of the electrical activity of neurons and the synaptic input that initiate or terminate such activity.” [Llinas et al., 1998]	Quantity
ERE	”only a balanced and disciplined account of both the external and experiential side of an issue can make us move one step closer to bridging the biological mind-experiential mind gap”[Varela, 1996] ”Marked quantitative and qualitative differences are observed between the ’perception’(upright face) and ’no perception’ (upside-down face) conditions.” [Thompson and Varela, 2001] ”One strategy would be to precisely describe the ongoing cognitive activity by obtaining refined verbal reports from human subjects. These should reveal subtle changes in the subject’s experience (conditioned, for instance, by his/her cognitive strategy, attention level, and inner speech). This type of qualitative first-person data is usually omitted from brain-imaging studies. We show that if methodological precautions are taken when gathering first-person data, they can indeed be used to shed light on cognition via a joint analysis with quantitative measures of neural activity.” [Lutz et al., 2002]	Quality and quantity
BE	Cf. quote in Table A.1. (”increasing momentary states of probability”)	Quantity

OrchOR	”[I]n the Orch OR scheme, these events are taken to have a rudimentary subjective experience, which is undifferentiated and lacking in cognition, perhaps providing the con- stitutive ingredients of what philosophers call qualia. We term such unorchestrated, ubiquitous OR events, lacking information and cognition, ’proto-conscious’.[...] Such OR events would have to be ’orchestrated’ in an appropriate way (Orch OR), for genuine consciousness to arise.” [Hameroff and Penrose, 2014]	Quantity
HOT	”...HOT eliminates the cumbersome task faced by local theories of having to explain every distinct kind of phenomenology with a different first-order mechanism.” ”we need to understand how nonconscious HOTs can result in conscious qualities.” ...my strategy is to explain a state’s being a conscious state in terms of our being conscious of that state in some particular way. [Rosenthal, 2002a]	Quantity
IIT	”IIT is an attempt to characterize consciousness mathematically both in quantity and in quality (...) According to IIT, a MICS specifies the quality of an experience and integrated information Φ_{Max} its quantity.” [Oizumi et al., 2014]. ”this approach provides an initial account of phenomenal space, it may also serve as a starting point for investigating other aspects of the quality of experience and their physical correspondents (...) The one-to-one correspondence between phenomenal and physical properties (...) represents a first attempt to deploy the theory to account for the quality of experience” [Haun and Tononi, 2019].	Quality and quantity
GNW	”This dynamical constraint suggests the existence of two thresholds in human information processing, one that corresponds to the minimal stimulus duration needed to cause any differentiated neural activity at all, and another, the ’consciousness threshold’, which corresponds to the significantly longer duration needed for such a neural representation to be mobilized in the workspace through a self-sustained long-distance loop.” [Dehaene and Naccache, 2001] ”the following features of four unconscious states, that are causally very different from each other: deep sleep, coma/vegetative states, epileptic loss of consciousness, and general anesthesia under various agents. Surprisingly, despite their very different mechanisms they share major common features. These include: (i) widely synchronized slow waveforms that take the place of the fast and flexible interactions needed for conscious functions; (ii) frontoparietal hypometabolism; (iii) widely blocked functional connectivity, both corticocortical and thalamocortical; and (iv) behavioral unconsciousness, including unresponsiveness to normally conscious stimuli.” [Baars, 2005]	Quantity

RPT	”Conscious and unconscious processing and the feedforward-feedback distinction (...) Instead of attributing visual awareness to specific areas or pathways, it might be useful to relate the conscious-unconscious dichotomy to the distinction between feedforward and recurrent processing.” [Lamme, 2003]	Quantity
PPI	”Our use of the word ”sentience” here is in the sense of ”responsive to sensory impressions”. It is not used in the philosophy of mind sense; namely, the capacity to perceive or experience subjectively, i.e., phenomenal consciousness, or having ’qualia’. Sentience here, simply implies the existence of a non-empty subset of systemic states; namely, sensory states. In virtue of the conditional dependencies that define this subset (i.e., the Markov blanket partition), the internal states are necessarily ’responsive to’ sensory states and thus the dictionary definition is fulfilled. The deeper philosophical issue of sentience speaks to the hard problem of tying down quantitative experience or subjective experience within the information geometry afforded by the Markov blanket construction.” [Friston et al., 2020]	Quantity
TTC	”Based on various lines of empirical evidence, here we postulate the four dimensions of consciousness (level/state, content/form, phenomenal/ experience, cognitive/reporting) are mediated by four corresponding temporo-spatial neuronal mechanisms: (i) the neuronal mechanism of ”temporo-spatial nestedness” accounts for the level or state of consciousness; (ii) the neuronal mechanism of ”temporo-spatial alignment” accounts for selecting the content and constituting the form of con- sciousness; (iii) the neuronal mechanism of ”temporo-spatial expansion” accounts for the phenomenal dimension of consciousness, e.g. experience with qualia; (iv) the neuronal mechanism of ”temporo- spatial globalization” accounts for the cognitive dimension of con- sciousness, e.g. the reporting of its contents” [Northoff and Huang, 2017]	Quantity and quality
CAN	”The first is the combination of phenomenal experiences, i.e., of qualia. For instance, one’s taste experiences of salt, garlic, onion, basil and tomato are somehow combined into the novel taste experience of a delicious pasta sauce. What is the relationship between one’s experiences of the ingredients and one’s experience of the sauce? (...) To propose that we represent the possible qualia of a conscious agent by a probability space is to propose that qualia convey information, since probability and information are (as Shannon showed) transforms of each other. It is also to propose that qualia need not, in general, exhibit other structures, such as metrics or dimensions. Now certain qualia spaces, such as the space of phenomenal colors, do exhibit metrical and dimensional properties. These properties are not precluded. They are allowed but not required. All that is required is that we can meaningfully talk about the information content of qualia.” [Hoffman and Prakash, 2014].	Quality

Appendix B

SUPPLEMENTARY FIGURES

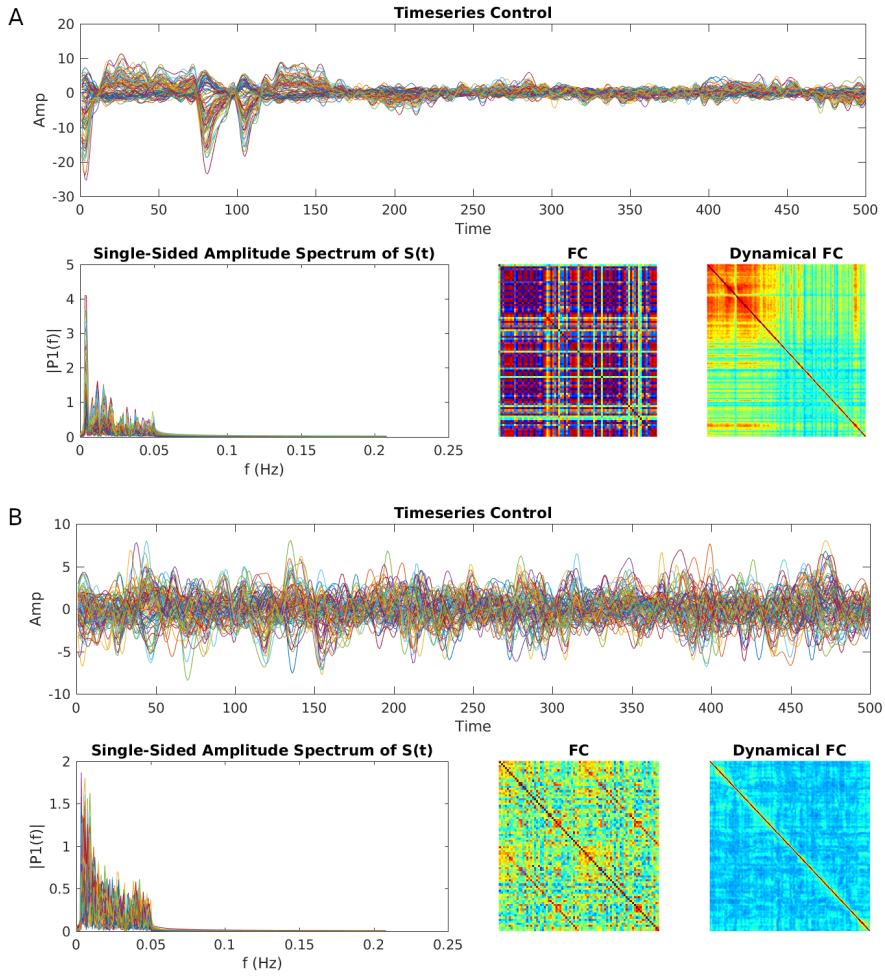


Figure B.1: (**A**) Example of discarded run for the deep sevoflurane condition. (**B**) Example of a correct run for deep sevoflurane condition. In both cases, time-series, the Fourier transform, the functional connectivity and the dynamical connectivity computed with phase correlation were plotted for visual inspection. An artifact is clearly recognized in (**A**), around 100 ms.

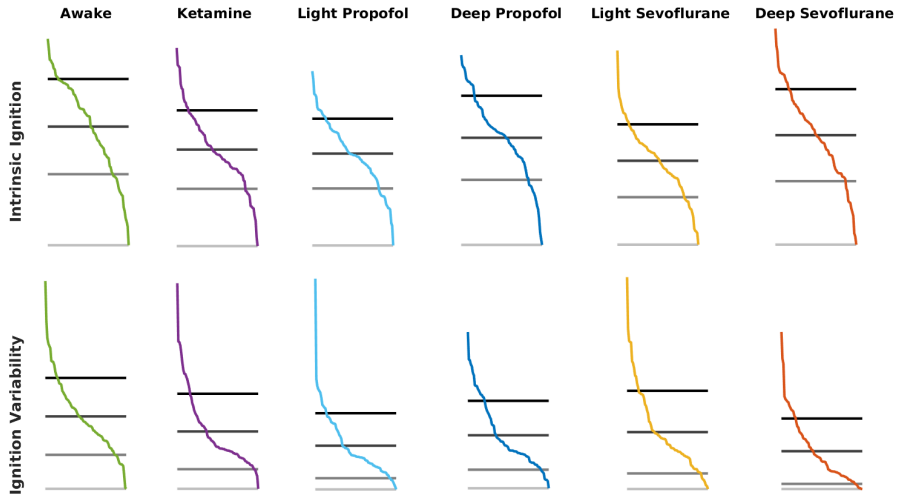


Figure B.2: Ignition curves zoomed. For each condition, the Intrinsic Ignition per node curve and the Ignition Variability per node curve are plotted in their own scales. Horizontal lines, from darker to lighter, correspond to $\mu + \sigma$; μ; $\mu - \sigma$; $\mu - \sigma - min(node)$; i is the index node, mean value and standard deviation for each curve. The min(node) is the minimum value of the curve. These curves suggest major differences in the Ignition Variability than the hierarchical structure given by the Intrinsic Ignition curve.

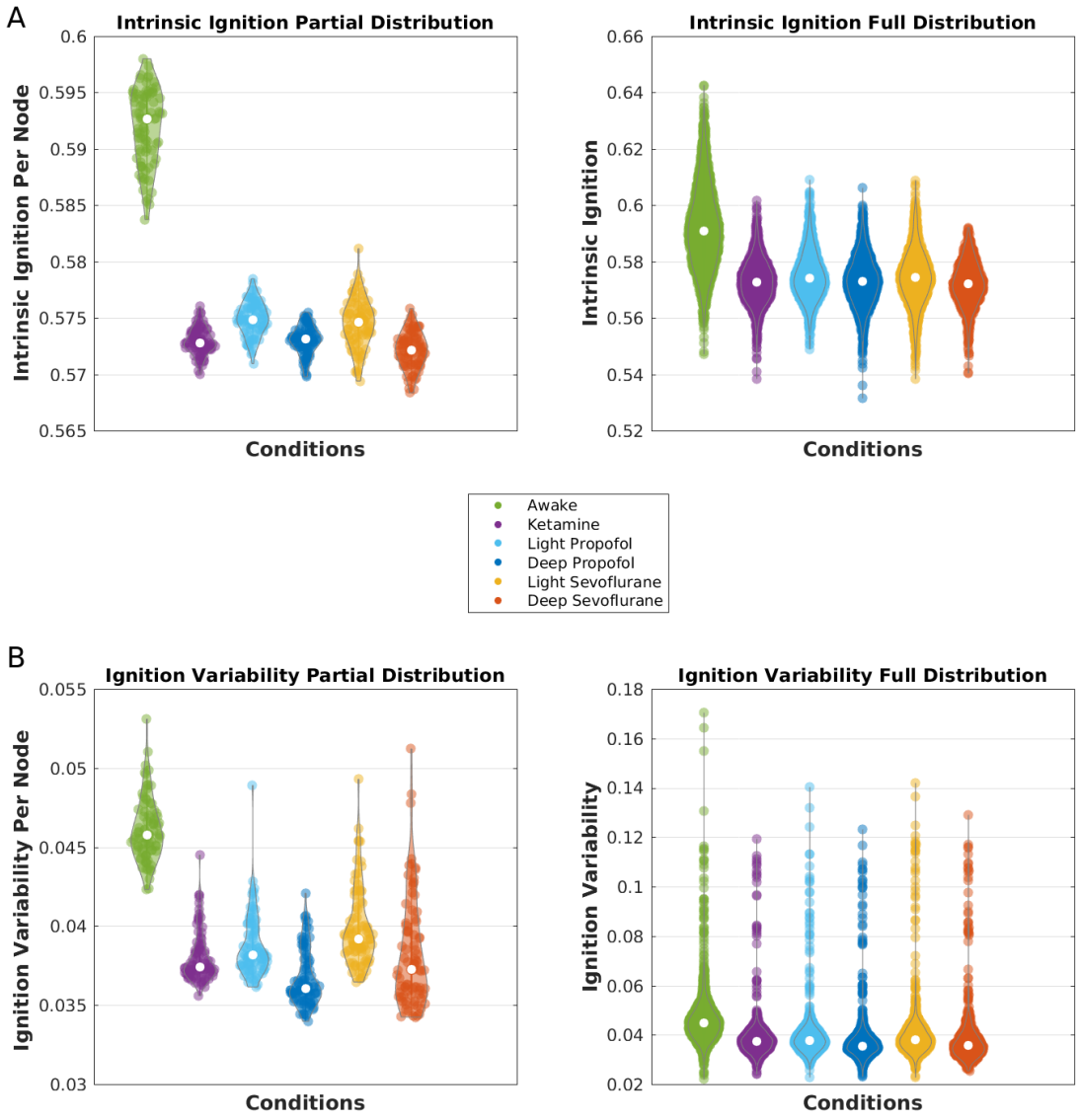


Figure B.3: Partial versus full Density Distributions on Intrinsic Ignition. Statistical analyses in both sets give similar results. For visualization purposes, we used the averaged data (Figure 4.3). (**A**) Partial distribution (ROIs per condition) and Full distribution (runs x ROIs per condition) for Intrinsic Ignition. Statistically, results are similar using the partial distribution or the full distribution. Using the full distribution, awake condition differentiates from all other conditions (KolmogorovâSmirnov test, $p < 0.001$), ketamine is not significantly different than deep propofol ($p > 0.01$) nor deep sevoflurane ($p > 0.01$), light propofol is not different from light sevoflurane ($p > 0.01$), deep propofol is not different from deep sevoflurane ($p > 0.01$). The student's t-test presents a similar tendency than Kolmogorov-Smirnov test with the exception of deep propofol and dep sevoflurane (t-test, $p = 0.009$). (**B**) Partial distribution (ROIs per condition) and Full distribution (runs x ROIs per condition) for Ignition Variability. Using the full distribution all conditions differentiate between them (Kolmogorov-Smirnov test, $p < 0.01$). Student's t-test showed a similar tendency than Kolmogorov-Smirnov test with the exception of ketamine versus deep sevoflurane (t-test, $p = 0.52$), light propofol versus deep sevoflurane (t-test, $p = 0.26$). Compare these results with results in Figure 4.3. The shape of Violin plots describes the distribution density, the white dot corresponds to the median, the thick inner line is the first quartile (down), and the third quartile (up). The borders are the upper and lower adjacent values [Hintze and Nelson, 1998].

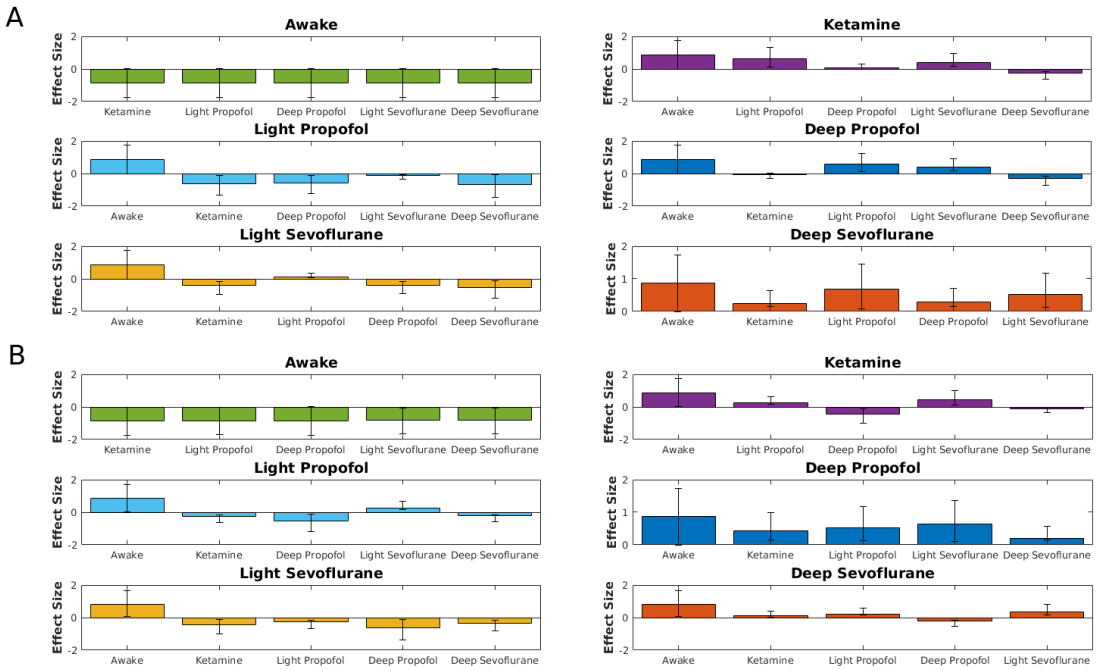


Figure B.4: Effect Size Analysis for Intrinsic Ignition. An effect size analysis was performed among conditions to quantify the apparent differences from the previous Figure 4.3 [Hentschke and Stüttgen, 2011]. Bar plots show the results from the rank-biserial correlation analysis (ranks between -1 to $+1$ with 0 no effect) for independent samples with 10.000 bootstrapping iterations to compute confidence intervals (other effect size techniques such as mean difference, AUROC and Cohen U1 did not present major differences with the results here). (**A**) The effect size analysis on the Intrinsic Ignition per node suggests bigger effects between awake and all the other sedation conditions (effect -0.866, CI $[-0.8661, -0.8661]$), while ketamine and deep propofol have small differences between them (effect 0.0659 CI $[-0.087, 0.222]$), as well as light propofol in comparison with light sevoflurane (effect -0.1 CI $[-0.259, 0.047]$. It supports previous analyses in Figure 4.3b. (**A**) The effect size analysis on the Ignition Variability per node suggests that all the conditions have a considerable effect size, with the exception of ketamine and deep sevoflurane (effect -0.1 CI $[-0.266, 0.063]$). Error bars are CI.

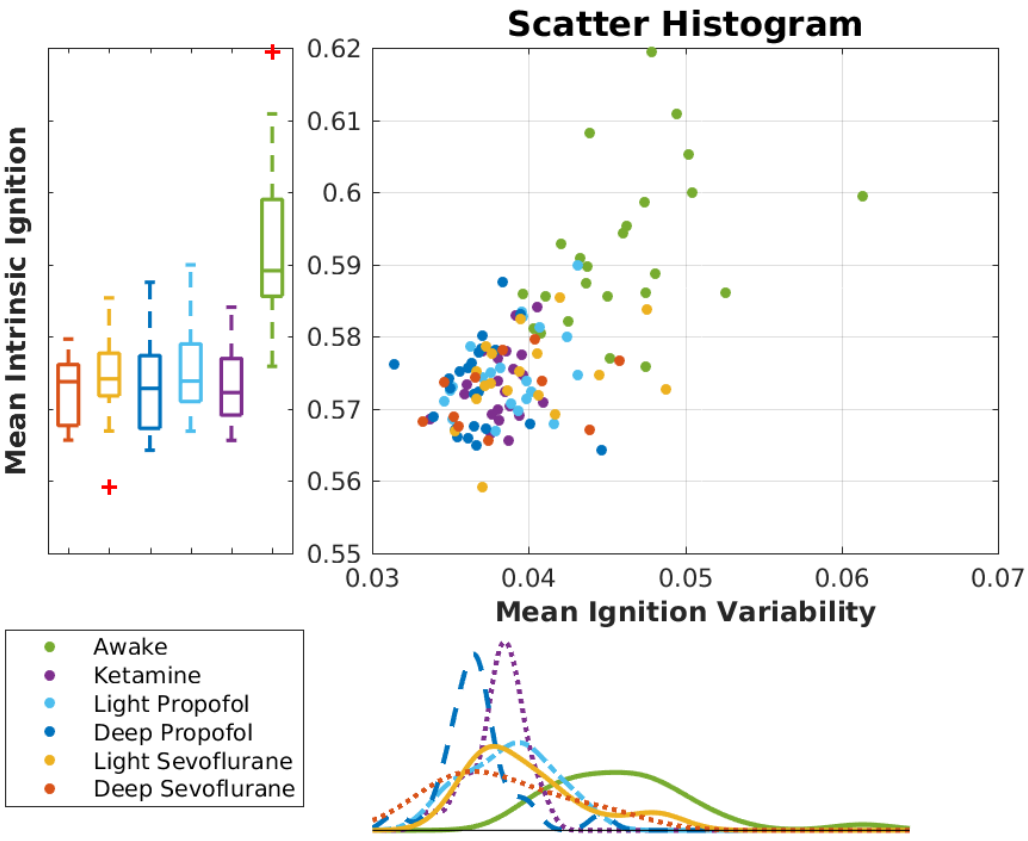


Figure B.5: Scatter plot. Scatter plot between Mean Intrinsic Ignition and Mean Ignition Variability. Two clusters are not so clearly observed as before, Figure 4.3b. However, statistically, the mean Intrinsic Ignition in the awake condition is different from anesthetics (values in captions Figure 4.3d), while for the mean Ignition Variability, the awake condition is also significantly higher than sedations. Deep propofol is statistically lower than ketamine ($p = 0.001$), light propofol ($p = 0.0053$), and light sevoflurane ($p = 0.004$). However other sedations are not statistically different between them (awake CI $[0.0480, 0.0442]$, ketamine CI $[0.0388, 0.0374]$, light propofol CI $[0.0399, 0.0377]$, deep propofol CI $[0.0377, 0.0357]$, light sevoflurane CI $[0.0417, 0.0382]$, deep sevoflurane CI $[0.0407, 0.0360]$).

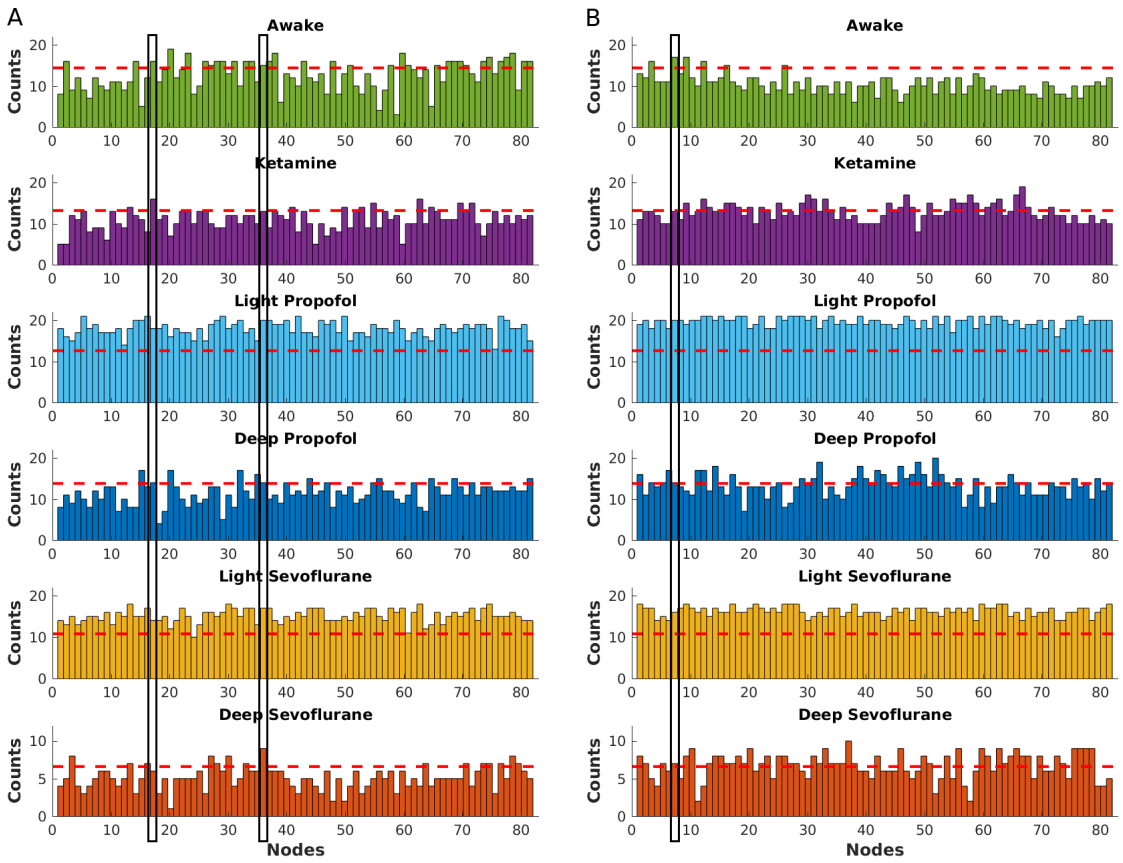


Figure B.6: Local tendency analysis across subjects. If the same analysis from Figures 4.4, 4.6, and Supplementary Tables 3-8 is now performed subject by subject across the six conditions, only a few nodes survived the local tendency analysis. For each subject in each condition, the logical propositions from Supplementary tables 3-8 were computed and each appearance saved to then plot the histograms for Intrinsic Ignition (**A**) and Ignition Variability (**A**). If a threshold of 60% of appearance is required (red dotted lines), only two nodes for Intrinsic Ignition and only one node for Ignition Variability satisfy the three logical propositions across the six conditions. In the first case, the regions are Subgenual cingulate cortex right and Intraparietal cortex right (index 17 and 36 respectively). In the second case, only the Central temporal cortex right (index 7) survived this restricted analysis. Black vertical squares highlight the areas mentioned.

Appendix C

SUPPLEMENTARY TABLES

Table C.1: Effect Size Values and Confidence intervals for Intrinsic Ignition per node and Intrinsic Variability. Effect size analyzes quantify how much different two distributions are, instead of only focusing on if two distributions are or not different. It is a complement of common statistical tests for Figure 4.3

Intrinsic Ignition	Awake		Ketamine		Light Propofol		Deep Propofol		Light Sevoflurane		Deep Sevoflurane	
	Effect	CI	Effect	CI	Effect	CI	Effect	CI	Effect	CI	Effect	CI
Awake	-	-	-0.87	[-0.87 -0.87]	-0.87	[-0.87 -0.87]	-0.87	[-0.87 -0.87]	-0.87	[-0.87 -0.87]	-0.87	[-0.87 -0.87]
Ketamine	0.87	[0.87 0.87]	-	-	0.61	[0.5 0.7]	0.66	[-0.09 -0.22]	0.41	[0.27 0.54]	-0.25	[-0.4 -0.11]
Light propofol	0.87	[0.87 0.87]	-0.61	[-0.7 -0.5]	-	-	-0.57	[-0.67 -0.46]	-0.11	[0.26 0.04]	-0.69	[-0.76–0.6]
Deep propofol	0.87	[0.87 0.87]	-0.07	[-0.22 0.09]	0.57	[0.46 0.67]	-	-	0.4	[0.25 0.53]	-0.29	[-0.42 -0.14]
Light sevoflurane	0.87	[0.87 0.87]	-0.41	[-0.55 -0.27]	-0.11	[-0.05 0.26]	-0.39	[-0.53 -0.25]	-	-	-0.53	[-0.64–0.4]
Deep sevoflurane	0.87	[0.87 0.87]	0.25	[0.1 0.4]	0.69	[0.61 0.76]	-0.29	[0.14 0.42]	0.53	[0.4 0.64]	-	-
Intrinsic Variability	Awake		Ketamine		Light Propofol		Deep Propofol		Light Sevoflurane		Deep Sevoflurane	
	Effect	CI	Effect	CI	Effect	CI	Effect	CI	Effect	CI	Effect	CI
Awake	-	-	-0.86	[-0.87 -0.85]	-0.85	[-0.87 -0.80]	-0.87	[-0.87 -0.87]	-0.80	[-0.84 -0.73]	-0.79	[-0.85 -0.72]
Ketamine	0.86	[0.85 0.87]	-	-	0.24	[0.09 0.38]	0.43	[-0.56 -0.29]	0.45	[0.32 0.57]	-0.1	[-0.27 0.06]
Light propofol	0.85	[0.8 0.87]	-0.24	[-0.38 -0.08]	-	-	-0.52	[-0.64 -0.4]	0.26	[0.11 0.4]	-0.2	[-0.36 -0.04]
Deep propofol	0.87	[0.87 0.87]	0.43	[0.29 0.56]	0.52	[0.4 0.64]	-	-	0.63	[0.52 0.72]	0.2	[0.05 0.36]
Light sevoflurane	0.8	[0.73 0.84]	-0.45	[-0.57 -0.32]	-0.26	[-0.4 -0.11]	-0.63	[-0.72 -0.52]	-	-	-0.33	[-0.47 -0.18]
Deep sevoflurane	0.79	[0.72 0.85]	0.1	[-0.06 0.27]	0.21	[0.05 0.37]	-0.2	[-0.35 -0.05]	0.33	[0.19 0.47]	-	-

Table C.2: Region of Interest. Names and indexes region of Interest in CoCoMac parcellation.

Index	Name	Hemisfere	Acronyms
1	Tempolar polar	Right	TCpol
2	Superior temporal cortex	Right	TCs
3	Amygdala	Right	Amyg
4	Orbitoinferior prefrontal cortex	Right	PFCoi
5	Anterior insula	Right	Ia
6	Orbitomedial prefrontal cortex	Right	PFCom
7	Central temporal cortex	Right	TCc
8	Orbitolateral prefrontal cortex	Right	PFCol
9	Inferior temporal	Right	TCi
10	Parahippocampal cortex	Right	PHC
11	Gustatory cortex	Right	G
12	Ventrolateral premotor cortex	Right	PMCvl
13	Anterior visual area (ventral)	Right	VACv
14	Posterior insula	Right	Ip
15	Prefrontal polar cortex	Right	PFCpol
16	Hippocampus	Right	HC
17	Subgenual cingulate cortex	Right	CCs
18	Ventrolateral prefrontal cortex	Right	PFCvl
19	Visual area 2	Right	V2
20	Medial prefrontal cortex	Right	PFCm
21	Ventral temporal cortex	Right	TCv
22	Anterior visual area (dorsal)	Right	VACd
23	Visual area 1	Right	V1
24	Centrolateral prefrontal cortex	Right	PFCcl
25	Secondary auditory cortex	Right	A2
26	Retrosplenial cingulate cortex	Right	CCr
27	Posterior cingulate cortex	Right	CCp
28	Anterior cingulate cortex	Right	CCa
29	Secondary somatosensory cortex	Right	S2
30	Primary somatosensory cortex	Right	S1
31	Primary auditory cortex	Right	A1
32	Primary motor cortex	Right	M1
33	Inferior parietal cortex	Right	PCi
34	Medial parietal cortex	Right	PCm
35	Dorsomedial prefrontal cortex	Right	PFCdm
36	Intraparietal cortex	Right	PCip
37	Superior parietal cortex	Right	PCs
38	Frontal eye field	Right	FEF
39	Dorsolateral prefrontal cortex	Right	PFCdl
40	Medial premotor cortex	Right	PMCm
41	Dorsolateral premotor cortex	Right	PMCdl
42	Tempolar polar	Left	TCpol
43	Superior temporal cortex	Left	TCs
44	Amygdala	Left	Amyg
45	Orbitoinferior prefrontal cortex	Left	PFCoi

46	Anterior insula	Left	Ia
47	Orbitomedial prefrontal cortex	Left	PFCom
48	Central temporal cortex	Left	TCc
49	Orbitolateral prefrontal cortex	Left	PFCol
50	Inferior temporal	Left	TCi
51	Parahippocampal cortex	Left	PHC
52	Gustatory cortex	Left	G
53	Ventrolateral premotor cortex	Left	PMCvl
54	Anterior visual area (ventral)	Left	VACv
55	Posterior insula	Left	Ip
56	Prefrontal polar cortex	Left	PFCpol
57	Hippocampus	Left	HC
58	Subgenual cingulate cortex	Left	CCs
59	Ventrolateral prefrontal cortex	Left	PFCvl
60	Visual area 2	Left	V2
61	Medial prefrontal cortex	Left	PFCm
62	Ventral temporal cortex	Left	TCv
63	Anterior visual area (dorsal)	Left	VACd
64	Visual area 1	Left	V1
65	Centrolateral prefrontal cortex	Left	PFCcl
66	Secondary auditory cortex	Left	A2
67	Retrosplenial cingulate cortex	Left	CCr
68	Posterior cingulate cortex	Left	CCp
69	Anterior cingulate cortex	Left	CCa
70	Secondary somatosensory cortex	Left	S2
71	Primary somatosensory cortex	Left	S1
72	Primary auditory cortex	Left	A1
73	Primary motor cortex	Left	M1
74	Inferior parietal cortex	Left	PCi
75	Medial parietal cortex	Left	PCm
76	Dorsomedial prefrontal cortex	Left	PFCdm
77	Intraparietal cortex	Left	PCip
78	Superior parietal cortex	Left	PCs
79	Frontal eye field	Left	FEF
80	Dorsolateral prefrontal cortex	Left	PFCdl
81	Medial premotor cortex	Left	PMCm
82	Dorsolateral premotor cortex	Left	PMCdl

Appendix D

LIST OF PUBLICATIONS

List of publication on Computational and Cognitive Neuroscience during the PhD. Unrelated work is not listed (for a complete list of publications during this period, see the author's CV).

Published Work. Published work that form part of the main chapters

1. **Chapter 4**: Hierarchical disruption in the cortex of anesthetized monkeys as a new signature of consciousness loss. Camilo Miguel Signorelli, Lynn Uhrig, Morten Kringelbach, Bechir Jarraya, Gustavo Deco. NeuroImage, 2021, 227.
2. **Chapter 5**: Towards new concepts for a biological neuroscience of consciousness. Camilo Miguel Signorelli and Daniel Meling. Cognitive Neurodynamics, 2021.

Submitted Work. Submitted work that form part of the main chapters.

1. **Chapter 2**: Explanatory profiles of models of consciousness - towards a systematic classification. Camilo Miguel Signorelli, Joanna Szczotka, Robert Prentner. Submitted to the neuroscience of consciousness journal, special issue.
2. **Chapter 3**: From brain-body function to consciousness interaction. Camilo Miguel Signorelli, Joaquin Diaz Boils, Enzo Tagliazucchi, Bechir Jarraya, Gustavo Deco. Submitted to Neuroscience & Biobehavioral Reviews.

Other Work. Peer-reviewed work that do not form part of the thesis.

1. Two paradigms of bistable plaid motion reveal independent mutual inhibition processes. Jean-Michel Hupé, Camilo Miguel Signorelli, David Alais. Journal of Vision, 2019, 19, 4.
2. Macroscopic quantities of collective brain activity during wakefulness and anesthesia. Adrián Ponce-Alvarez, Lynn Uhrig, Nikolas Deco, Camilo M. Signorelli,

Morten Kringelbach, Bechir Jarraya, Gustavo Deco. Submitted to Science Advances.

Bibliography

[Achermann et al., 1995] Achermann, P., Werth, E., Dijk, D. J., and Borbely, A. A. (1995). Time course of sleep inertia after nighttime and daytime sleep episodes.

[Aeschbach et al., 2001] Aeschbach, D., Postolache, T. T., Sher, L., Matthews, J. R., Jackson, M. A., and Wehr, T. A. (2001). Evidence from the waking electroencephalogram that short sleepers live under higher homeostatic sleep pressure than long sleepers. *Neuroscience*, 102(3):493–502.

[Alkire et al., 2008] Alkire, M. T., Hudetz, A. G., and Tononi, G. (2008). Consciousness and anesthesia. *Science*, 322(5903):876–880.

[Allen et al., 2014] Allen, E. a., Damaraju, E., Plis, S. M., Erhardt, E. B., Eichele, T., and Calhoun, V. D. (2014). Tracking whole-brain connectivity dynamics in the resting state. *Cerebral cortex*, 24(3):663–76.

[Allen et al., 2016] Allen, M., Frank, D., Samuel Schwarzkopf, D., Fardo, F., Winston, J. S., Hauser, T. U., and Rees, G. (2016). Unexpected arousal modulates the influence of sensory noise on confidence. *eLife*, 5(OCTOBER):1–17.

[Anderson and Chemero, 2013] Anderson, M. L. and Chemero, T. (2013). The problem with brain GUTs: Conflation of different senses of "prediction" threatens metaphysical disaster. *Behavioral and Brain Sciences*, 36(3):204–205.

[Anis et al., 1983] Anis, N. A., Berry, S. C., Burton, N. R., and Lodge, D. (1983). The dissociative anaesthetics, ketamine and phencyclidine, selectively reduce excitation of central mammalian neurones by N-methyl-aspartate. *British Journal of Pharmacology*, 79(2):565–575.

[Arslan et al., 2018] Arslan, S., Ktena, S. I., Makropoulos, A., Robinson, E. C., Rueckert, D., and Parisot, S. (2018). Human brain mapping: A systematic comparison of parcellation methods for the human cerebral cortex. *NeuroImage*, 170(April 2017):5–30.

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

[Aru et al., 2020] Aru, J., Suzuki, M., and Larkum, M. E. (2020). Cellular Mechanisms of Conscious Processing. *Trends in Cognitive Sciences*, 24(10):814–825.

[Ashby, 1957] Ashby, W. R. (1957). *An introduction to cybernetics.* Chapman & Hall, London, second edition.

[Atasoy et al., 2019] Atasoy, S., Deco, G., and Kringelbach, M. L. (2019). Playing at the Edge of Criticality: Expanded Whole-Brain Repertoire of Connectome-Harmonics. In Tomen, N., editor, *The Functional Role of Critical Dynamics in Neural Systems*, volume 11, pages 27–45. Springer International Publishing.

[Atasoy et al., 2017] Atasoy, S., Deco, G., Kringelbach, M. L., and Pearson, J. (2017). Harmonic Brain Modes: A Unifying Framework for Linking Space and Time in Brain Dynamics. *The Neuroscientist*, page 107385841772803.

[Atasoy et al., 2016] Atasoy, S., Donnelly, I., and Pearson, J. (2016). Human brain networks function in connectome-specific harmonic waves. *Nature Communications*, 7:10340.

[Atmanspacher, 2020] Atmanspacher, H. (2020). The Pauli-Jung Conjecture and Its Relatives: A Formally Augmented Outline. *Open Philosophy*, 3(1):527–549.

[Atmanspacher and Rotter, 2008] Atmanspacher, H. and Rotter, S. (2008). Interpreting neurodynamics: Concepts and facts. *Cognitive Neurodynamics*, 2(4):297–318.

[Atmaspacher, 2014] Atmaspacher, H. (2014). 20Th Century Variants of Dual-Aspect Thinking. *Mind and Matter*, 12(2):245–269.

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

[Bachmann and Hudetz, 2014] Bachmann, T. and Hudetz, A. G. (2014). It is time to combine the two main traditions in the research on the neural correlates of consciousness: C=LxD. *Frontiers in Psychology*, 5(AUG):1–13.

[Bachmann et al., 2004] Bachmann, T., Poder, E., and Luiga, I. (2004). Illusory reversal of temporal order: the bias to report a dimmer stimulus as the first. *Vision Research*, 44(3):241–6.

[Baez et al., 2018] Baez, J. C., Foley, J., Moeller, J., and Pollard, B. S. (2018). Network models.

[Bakker et al., 2012] Bakker, R., Wachtler, T., and Diesmann, M. (2012). CoCoMac 2.0 and the future of tract-tracing databases. *Frontiers in Neuroinformatics*, 6(DEC):1–6.

[Balestrini et al., 2016] Balestrini, S., Francione, S., Mai, R., Castana, L., Casaceli, G., Marino, D., Provinciali, L., Cardinale, F., and Tassi, L. (2016). Reply: The dorsal cingulate cortex as a critical gateway in the network supporting conscious awareness. *Brain*, 139(e24):1–2.

[Baria et al., 2017] Baria, A. T., Maniscalco, B., and He, B. J. (2017). Initial-state-dependent, robust, transient neural dynamics encode conscious visual perception. *PLoS Computational Biology*, 13(11):1–29.

[Barnes, 1992] Barnes, E. (1992). Explanatory Unification and the Problem of Asymmetry. *Philosophy of Science*, 59(4):558–571.

[Barron and Klein, 2016] Barron, A. B. and Klein, C. (2016). What insects can tell us about the origins of consciousness. *Proceedings of the National Academy of Sciences of the United States of America*, 113(18):4900–8.

[Bartsch et al., 2015] Bartsch, R. P., Liu, K. K., Bashan, A., and Ivanov, P. C. (2015). Network physiology: How organ systems dynamically interact. *PLoS ONE*, 10(11):1–36.

[Barttfeld et al., 2015] Barttfeld, P., Uhrig, L., Sitt, J. D., Sigman, M., Jarraya, B., and Dehaene, S. (2015). Signature of consciousness in the dynamics of resting-state brain activity. *Proceedings of the National Academy of Sciences*, page 201418031.

[Bashan et al., 2012] Bashan, A., Bartsch, R. P., Kantelhardt, J. W., Havlin, S., and Ivanov, P. C. (2012). Network physiology reveals relations between network topology and physiological function. *Nature Communications*, 3:702–709.

[Basti et al., 2020] Basti, A., Nili, H., Hauk, O., Marzetti, L., and Henson, R. N. (2020). Multi-dimensional connectivity: a conceptual and mathematical review. *NeuroImage*, 221(February).

[Bayne, 2018] Bayne, T. (2018). On the axiomatic foundations of the integrated information theory of consciousness. *Neuroscience of Consciousness*, 2018(1):1–8.

[Bayne and Carter, 2018] Bayne, T. and Carter, O. (2018). Dimensions of consciousness and the psychedelic state. *Neuroscience of Consciousness*, 4(1):1–8.

[Bayne et al., 2016] Bayne, T., Hohwy, J., and Owen, A. M. (2016). Are There Levels of Consciousness? *Trends in Cognitive Sciences*, 20(6):405–413.

[Beck and Eccles, 1992] Beck, F. and Eccles, J. C. (1992). Quantum aspects of brain activity and the role of consciousness. *Proceedings of the National Academy of Sciences*, 89(23):11357–11361.

[Bekinschtein et al., 2009] Bekinschtein, T., Dehaene, S., Rohaut, B., Tadel, F., Cohen, L., and Naccache, L. (2009). Neural signature of the conscious processing of auditory regularities. *PNAS*, 106(5):1672–1677.

[Bélanger et al., 2011] Bélanger, M., Allaman, I., and Magistretti, P. J. (2011). Brain energy metabolism: Focus on Astrocyte-neuron metabolic cooperation. *Cell Metabolism*, 14(6):724–738.

[Bevan et al., 1997] Bevan, J. C., Veall, G. R., Macnab, A. J., Ries, C. R., and Marsland, C. (1997). Midazolam premedication delays recovery after propofol without modifying involuntary movements. *Anesthesia and Analgesia*, 85(1):50–54.

[Biswal et al., 1995] Biswal, B., Yetkin, F. Z., Haughton, V. M., and Hyde, J. S. (1995). Functional Connectivity in the Motor Cortex of Resting Human Brain Using Echo-Planar MRI. *Magnetic Resonance in Medicine*, 34(4):537–541.

[Block, 1995] Block, N. (1995). On a confusion about a function of consciousness. *Behavioral and Brain Sciences*, 18(2):227–247.

[Block, 1996] Block, N. (1996). What is functionalism? In *The Encyclopedia of Philosophy Supplement*.

[Block, 1999] Block, N. (1999). Sexism, Racism, Ageism, and the Nature of Consciousness. *Philosophical Topics*, 26(1):39–70.

[Block, 2005] Block, N. (2005). Two neural correlates of consciousness. *Trends in Cognitive Sciences*, 9(2):46–52.

[Block, 2007] Block, N. (2007). Consciousness, accessibility, and the mesh between psychology and neuroscience. *Behavioral and Brain Sciences*, 30(5-6):481–548.

[Block, 2009] Block, N. (2009). Comparing the major theories of consciousness. In *The cognitive neurosciences, 4th ed.*, pages 1111–1122. Massachusetts Institute of Technology, Cambridge, MA, US.

[Boccaletti et al., 2014] Boccaletti, S., Bianconi, G., Criado, R., del Genio, C. I., Gómez-Gardeñes, J., Romance, M., Sendiña-Nadal, I., Wang, Z., and Zanin, M. (2014). The structure and dynamics of multilayer networks. *Physics Reports*, 544(1):1–122.

[Boly et al., 2017] Boly, M., Massimini, M., Tsuchiya, N., Postle, B. R., Koch, C., and Tononi, G. (2017). Are the neural correlates of consciousness in the front or in the back of the cerebral cortex? Clinical and neuroimaging evidence. *Journal of Neuroscience*, 37(40):9603–9613.

[Boly et al., 2012a] Boly, M., Moran, R., Murphy, M., Boveroux, P., Bruno, M. A., Noirhomme, Q., Ledoux, D., Bonhomme, V., Brichant, J. F., Tononi, G., Laureys, S., and Friston, K. (2012a). Connectivity changes underlying spectral EEG changes during propofol-induced loss of consciousness. *Journal of Neuroscience*, 32(20):7082–7090.

[Boly et al., 2012b] Boly, M., Perlbarg, V., Marrelec, G., Schabus, M., Laureys, S., Doyon, J., Pélégrini-Issac, M., Maquet, P., and Benali, H. (2012b). Hierarchical clustering of brain activity during human nonrapid eye movement sleep. *PNAS*, 109(15):5856–5861.

[Bonhomme et al., 2001] Bonhomme, V., Fiset, P., Meuret, P., Backman, S., Plourde, G., Paus, T., Bushnell, M. C., and Evans, A. C. (2001). Propofol anesthesia and cerebral blood flow changes elicited by vibrotactile stimulation: A positron emission tomography study. *Journal of Neurophysiology*, 85(3):1299–1308.

[Bonhomme et al., 2016] Bonhomme, V., Vanhaudenhuyse, A., Demertzi, A., Bruno, M. A., Jaquet, O., Bahri, M. A., Plenevaux, A., Boly, M., Boveroux, P., Soddu, A., Brichant, J. F., Maquet, P., and Laureys, S. (2016). Resting-state Network-specific Breakdown of Functional Connectivity during Ketamine Alteration of Consciousness in Volunteers. *Anesthesiology*, 125(5):873–888.

[Boveroux et al., 2010] Boveroux, P., Vanhaudenhuyse, A., Bruno, M. A., Noirhomme, Q., Lauwick, S., Luxen, A., Degueldre, C., Plenevaux, A., Schnakers, C., Phillips, C., Brichant, J. F., Bonhomme, V., Maquet, P., Greicius, M. D., Laureys, S., and Boly, M. (2010). Breakdown of within- and between-network resting state functional magnetic resonance imaging connectivity during propofol-induced loss of consciousness. *Anesthesiology*, 113(5):1038–1053.

[Box, 1976] Box, G. E. (1976). Science and statistics. *Journal of the American Statistical Association*, 71(356):791–799.

[Braun et al., 1998] Braun, A. R., Balkin, T. J., Wesensten, N. J., Gwadry, F., Carson, R. E., Varga, M., Baldwin, P., Belenky, G., and Herscovitch, P. (1998). Dissociated pattern of activity in visual cortices and their projections during human rapid eye movement sleep. *Science*, 279(5347):91–95.

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

[Brette, 2016] Brette, R. (2016). Subjective Physics. In Hady, A. E., editor, *Closed Loop Neuroscience*, chapter Chapter 11, pages 145–169. Elsevier.

[Brette, 2017] Brette, R. (2017). Is coding a relevant metaphor for the brain? *BioRxiv*, (Jul).

[Brown et al., 2010] Brown, E. N., Lydic, R., and Schiff, N. D. (2010). General Anesthesia, Sleep, and Coma. *New England Journal of Medicine*, 363(27):2638–2650.

[Brown et al., 2012] Brown, R. E., Basheer, R., McKenna, J. T., Strecker, R. E., and McCarley, R. W. (2012). Control of sleep and wakefulness. *Physiol. Rev.*, 92:1087–1187.

[Bruza et al., 2015] Bruza, P. D., Wang, Z., and Busemeyer, J. R. (2015). Quantum cognition: a new theoretical approach to psychology. *Trends in cognitive sciences*, 19(7):383–93.

[Buchsbaum et al., 2001] Buchsbaum, M. S., Hazlett, E. A., Wu, J., and Bunney, W. E. (2001). Positron emission tomography with deoxyglucose-F18 imaging of sleep. *Neuropsychopharmacology*, 25(5):S50–S56.

[Buckner et al., 2008] Buckner, R. L., Andrews-Hanna, J. R., and Schacter, D. L. (2008). The brain's default network: Anatomy, function, and relevance to disease. *Annals of the New York Academy of Sciences*, 1124:1–38.

[Busemeyer and Bruza, 2012] Busemeyer, J. R. and Bruza, P. D. (2012). *Quantum Models of Cognition and Decision*. Cambridge University Press, Cambridge, first edition.

[Cabral et al., 2017] Cabral, J., Kringelbach, M. L., and Deco, G. (2017). Functional connectivity dynamically evolves on multiple time-scales over a static structural connectome: Models and mechanisms. *NeuroImage*, 160(March):84–96.

[Canet et al., 2003] Canet, J., Raeder, J., Rasmussen, L. S., Enlund, M., Kuipers, H. M., Hanning, C. D., Jolles, J., Korttila, K., Siersma, V. D., Dodds, C., Abildstrom, H., Sneyd, J. R., Vila, P., Johnson, T., Muñoz Corsini, L., Silverstein, J. H., Nielsen, I. K., and Moller, J. T.

(2003). Cognitive dysfunction after minor surgery in the elderly. *Acta Anaesthesiologica Scandinavica*, 47(10):1204–1210.

[Cárdenas et al., 2010] Cárdenas, L. M., Letelier, J. C., Gutierrez, C., Cornish-Bowden, A., and Soto-Andrade, J. (2010). Closure to efficient causation, computability and artificial life. *Journal of Theoretical Biology*, 263(1):79–92.

[Carhart-Harris et al., 2014] 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*, 8(February):20.

[Carhart-Harris et al., 2016] Carhart-Harris, R. L., Muthukumaraswamy, S., Roseman, L., Kaelen, M., Droog, W., Murphy, K., Tagliazucchi, E., Schenberg, E. E., Nest, T., Orban, C., Leech, R., Williams, L. T., Williams, T. M., Bolstridge, M., Sessa, B., McGonigle, J., Sereno, M. I., Nichols, D., Hellyer, P. J., Hobden, P., Evans, J., Singh, K. D., Wise, R. G., Curran, H. V., Feilding, A., and Nutt, D. J. (2016). Neural correlates of the LSD experience revealed by multimodal neuroimaging. *Proceedings of the National Academy of Sciences of the United States of America*, 113(17):4853–4858.

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

[Casarotto et al., 2016] Casarotto, S., Comanducci, A., Rosanova, M., Sarasso, S., Fecchio, M., Napolitani, M., Pigorini, A., G. Casali, A., Trimarchi, P. D., Boly, M., Gosseries, O., Bodart, O., Curto, F., Landi, C., Mariotti, M., Devalle, G., Laureys, S., Tononi, G., and Massimini, M. (2016). Stratification of unresponsive patients by an independently validated index of brain complexity. *Annals of Neurology*, 80(5):718–729.

[Chakravarthi and VanRullen, 2012] Chakravarthi, R. and VanRullen, R. (2012). Conscious updating is a rhythmic process. *Proceedings of the National Academy of Sciences*, 109(26):10599–10604.

[Chalmers, 1995a] Chalmers, D. (1995a). The puzzle of conscious experience. *Scientific American*, 273(6):80–86.

[Chalmers, 2013a] Chalmers, D. (2013a). How can we construct a science of consciousness? *Annals of the New York Academy of Sciences*, 1303(1):25–35.

[Chalmers, 1995b] Chalmers, D. J. (1995b). Facing Up to the Problem of Consciousness. *Journal of Consciousness Studies*, (3):200–219.

[Chalmers, 1997] Chalmers, D. J. (1997). *The Conscious Mind*. Oxford University Press.

[Chalmers, 2013b] Chalmers, D. J. (2013b). Panpsychism and Panprotopsychism. *Amherst Lecture in Philosophy*, 8.

[Chalmers, 2018] Chalmers, D. J. (2018). The meta-problem is the problem of consciousness. *Journal of Consciousness Studies*, 25(9-10):6–61.

[Chalmers, 2019] Chalmers, D. J. (2019). Idealism and the Mind-Body Problem. In Seager, W., editor, *The Routledge Handbook of Panpsychism.*, pages 353–373. Routledge, New York.

[Chander et al., 2014] Chander, D., García, P. S., MacColl, J. N., Illing, S., and Sleigh, J. W. (2014). Electroencephalographic variation during end maintenance and emergence from surgical anesthesia. *PLoS ONE*, 9(9).

[Chaudhuri et al., 2015] Chaudhuri, R., Knoblauch, K., Gariel, M. A., Kennedy, H., and Wang, X. J. (2015). A Large-Scale Circuit Mechanism for Hierarchical Dynamical Processing in the Primate Cortex. *Neuron*, 88(2):419–431.

[Cleeremans, 2011] Cleeremans, A. (2011). The Radical Plasticity Thesis: How the Brain Learns to be Conscious. *Frontiers in psychology*, 2(May):86.

[Cocchi et al., 2017] Cocchi, L., Gollo, L. L., Zalesky, A., and Breakspear, M. (2017). Criticality in the brain: A synthesis of neurobiology, models and cognition. *Progress in Neurobiology*, 158:132–152.

[Coecke, 2011] Coecke, B., editor (2011). *New Structures for Physics*, volume 813 of *Lecture Notes in Physics*. Springer Berlin Heidelberg, Berlin, Heidelberg, lectures n edition.

[Coecke et al., 2016] Coecke, B., Duncan, R., Kissinger, A., and Wang, Q. (2016). Generalised Compositional Theories and Diagrammatic Reasoning. In Chiribella, G. and Spekkens, R., editors, *Quantum Theory: Informational Foundations and Foils. Fundamental Theories of Physics.*, volume 181, pages 309–366. Springer.

[Cole et al., 2013] Cole, M. W., Reynolds, J. R., Power, J. D., Repovs, G., Anticevic, A., and Braver, T. S. (2013). Multi-task connectivity reveals flexible hubs for adaptive task control. *Nature Neuroscience*, 16(9):1348–1355.

[Collier, 1972] Collier, B. B. (1972). Ketamine and the conscious mind. *Anaesthesia*, 27(2):120–134.

[Colombo, 2017] Colombo, M. (2017). Why build a virtual brain ? Large-scale neural simulations as jump start for cognitive computing. *Journal of Experimental & Theoretical Artificial Intelligence*, 3079(February):1–10.

[Corazzol et al., 2017] Corazzol, M., Lio, G., Lefevre, A., Deiana, G., Tell, L., André-obadia, N., Bourdillon, P., Guenot, M., Desmurget, M., Luauté, J., and Sirigu, A. (2017). Restoring Consciousness with vagus nerve stimulation. *Current Biology*, 27(18):R994–R996.

[Crick and Koch, 1998] Crick, F. and Koch, C. (1998). Consciousness and neuroscience. *Cerebral cortex*, 8:97–1007.

[Crick and Koch, 2003] Crick, F. and Koch, C. (2003). A Framework for Consciousness. *Nature neuroscience*, 6(2):119–126.

[Critchley and Harrison, 2013] Critchley, H. D. and Harrison, N. A. (2013). Visceral Influences on Brain and Behavior. *Neuron*, 77(4):624–638.

[Critchley et al., 2004] Critchley, H. D., Wiens, S., Rotshtein, P., Öhman, A., and Dolan, R. J. (2004). Neural systems supporting interoceptive awareness. *Nature Neuroscience*, 7(2):189–195.

[Daugman, 2001] Daugman, J. G. (2001). Brain metaphor and brain theory. In William P. Bechtel, Pete Mandik, J. M. &. R. S. S., editor, *Philosophy and the Neurosciences: A Reader*. Blackwell.

[de Barros et al., 2017] de Barros, J. A., Montemayor, C., and De Assis, L. P. (2017). Contextuality in the Integrated Information Theory. In de Barros, J. A., editor, *Quantum Interactions 2016*, volume 8369, pages 57–70. Springer International Publishing.

[de Pasquale et al., 2018] de Pasquale, F., Corbetta, M., Betti, V., and Della Penna, S. (2018). Cortical cores in network dynamics. *NeuroImage*, 180:370–382.

[de Rezende, 1975] de Rezende, A. M. (1975). Le point de départ dans la philosophie de Merleau-Ponty. *Revue Philosophique de Louvain*, 73(19):451–480.

[Deco et al., 2018a] 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 and Corbetta, 2011] Deco, G. and Corbetta, M. (2011). The dynamical balance of the brain at rest. *Neuroscientist*, 17(1):107–123.

[Deco et al., 2018b] 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.e6.

[Deco et al., 2019] Deco, G., Cruzat, J., and Kringelbach, M. L. (2019). Brain songs framework used for discovering the relevant timescale of the human brain. *Nature Communications*, 10(1):1–13.

[Deco et al., 2017a] Deco, G., Hartevelt, T. J. V., Fernandes, H. M., Stevner, A., and Kringelbach, M. L. (2017a). The most relevant human brain regions for functional connectivity: Evidence for a dynamical workspace of binding nodes from whole-brain computational modelling. *NeuroImage*, 146(November 2016):197–210.

[Deco et al., 2009] Deco, G., Jirsa, V., McIntosh, A., Sporns, O., and Kotter, R. (2009). Key role of coupling, delay, and noise in resting brain fluctuations. *Proceedings of the National Academy of Sciences*, 106(29):12207–12208.

[Deco and Jirsa, 2012] Deco, G. and Jirsa, V. K. (2012). Ongoing cortical activity at rest: Criticality, multistability, and ghost attractors. *Journal of Neuroscience*, 32(10):3366–3375.

[Deco et al., 2011] Deco, G., Jirsa, V. K., and McIntosh, A. R. (2011). Emerging concepts for the dynamical organization of resting-state activity in the brain. *Nature reviews. Neuroscience*, 12(1):43–56.

[Deco et al., 2013] Deco, G., Jirsa, V. K., and McIntosh, A. R. (2013). Resting brains never rest: Computational insights into potential cognitive architectures. *Trends in Neurosciences*, 36(5):268–274.

[Deco et al., 2008] 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).

[Deco and Kringelbach, 2017] Deco, G. and Kringelbach, M. L. (2017). Hierarchy of Information Processing in the Brain : A Novel 'Intrinsic Ignition' Framework. *Neuron*, 94(5):961–968.

[Deco et al., 2017b] Deco, G., Tagliazucchi, E., Laufs, H., Sanjuán, A., and Kringelbach, M. L. (2017b). Novel intrinsic ignition method measuring local-global integration characterizes wakefulness and deep sleep. *eNeuro*, 4(5):1–12.

[Deco et al., 2015] Deco, G., Tononi, G., Boly, M., and Kringelbach, M. L. (2015). Rethinking segregation and integration: contributions of whole-brain modelling. *Nature Reviews Neuroscience*, 16(7):430–439.

[DeFelipe, 2011] DeFelipe, J. (2011). The evolution of the brain, the human nature of cortical circuits, and intellectual creativity. *Frontiers in Neuroanatomy*, 5(MAY):1–17.

[Dehaene and Changeux, 2005] Dehaene, S. and Changeux, J. P. (2005). Ongoing spontaneous activity controls access to consciousness: A neuronal model for inattentional blindness. *PLoS Biology*, 3(5):0910–0927.

[Dehaene and Changeux, 2011] Dehaene, S. and Changeux, J.-P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron*, 70(2):200–27.

[Dehaene et al., 2014] Dehaene, S., Charles, L., King, J.-R., and Marti, S. (2014). Toward a computational theory of conscious processing. *Current opinion in neurobiology*, 25(1947):76–84.

[Dehaene et al., 2017] Dehaene, S., Lau, H., and Kouider, S. (2017). What is consciousness, and could machines have it? *Science*, 358(6362):484–489.

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

[Del Cul et al., 2007] Del Cul, A., Baillet, S., and Dehaene, S. (2007). Brain dynamics underlying the nonlinear threshold for access to consciousness. *PLoS Biology*, 5(10):2408–2423.

[Demertzi et al., 2014] Demertzi, A., Gómez, F., Crone, J. S., Vanhaudenhuyse, A., Tshibanda, L., Noirhomme, Q., Thonnard, M., Charland-Verville, V., Kirsch, M., Laureys, S., and Soddu, A. (2014). Multiple fMRI system-level baseline connectivity is disrupted in patients with consciousness alterations. *Cortex*, 52(1):35–46.

[Demertzi et al., 2019] Demertzi, A., Tagliazucchi, E., Dehaene, S., Deco, G., Barttfeld, P., Raimondo, F., Martial, C., Fernández-Espejo, D., Rohaut, B., Voss, H. U., Schiff, N. D., Owen, A. M., Laureys, S., Naccache, L., and Sitt, J. D. (2019). Human consciousness is supported by dynamic complex patterns of brain signal coordination. *Science Advances*, 5(2):1–11.

[Dennett, 1988] Dennett, D. C. (1988). Quining Qualia. In Marcel, A. and Bisiach, E., editors, *Consciousness in Modern Science*. Oxford University Press.

[Dennett, 1993] Dennett, D. C. (1993). *Consciousness Explained*. Penguin Books Ltd, London.

[Di Perri et al., 2016] Di Perri, C., Bahri, M. A., Amico, E., Thibaut, A., Heine, L., Antonopoulos, G., Charland-Verville, V., Wannez, S., Gomez, F., Hustinx, R., Tshibanda, L., Demertzi, A., Soddu, A., and Laureys, S. (2016). Neural correlates of consciousness in patients who have emerged from a minimally conscious state: A cross-sectional multimodal imaging study. *The Lancet Neurology*, 15(8):830–842.

[Doerig et al., 2019] 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(November 2018):49–59.

[Dosenbach et al., 2008] Dosenbach, N. U., Fair, D. A., Cohen, A. L., Schlaggar, B. L., and Petersen, S. E. (2008). A dual-networks architecture of top-down control. *Trends in Cognitive Sciences*, 12(3):99–105.

[Dresler et al., 2012] Dresler, M., Wehrle, R., Spoormaker, V. I., Koch, S. P., Holsboer, F., Steiger, A., Obrig, H., Sämann, P. G., and Czisch, M. (2012). Neural Correlates of Dream Lucidity Obtained from Contrasting Lucid versus Non-Lucid REM Sleep: A Combined EEG/fMRI Case Study. *Sleep*, 35(7):1017–1020.

[Durham et al., 2020] Durham, I., Kleiner, J., Kremnitzer, Y., Mason, J., and Prentner, R., editors (2020). *Models of Consciousness*, volume 22. Special is edition.

[Edelman, 2003] Edelman, G. M. (2003). Naturalizing consciousness: A theoretical framework. *Proceedings of the National Academy of Sciences*, 100(9):5520–5524.

[Ehresmann and Gomez-Ramirez, 2015] Ehresmann, A. C. and Gomez-Ramirez, J. (2015). Conciliating neuroscience and phenomenology via category theory. *Progress in Biophysics and Molecular Biology*, 119(3):347–359.

[Ellsberg, 1961] Ellsberg, D. (1961). Risk, ambiguity, and the Savage axioms. *The quarterly journal of economics*, 75(4):643–669.

[Engbers, 2018] Engbers, F. (2018). Is unconsciousness simply the reverse of consciousness? *Anaesthesia*, 73(1):6–9.

[Engemann et al., 2018] Engemann, D. A., Raimondo, F., King, J. R., Rohaut, B., Louppe, G., Faugeras, F., Annen, J., Cassol, H., Gosseries, O., Fernandez-Slezak, D., Laureys, S., Naccache, L., Dehaene, S., and Sitt, J. D. (2018). Robust EEG-based cross-site and cross-protocol classification of states of consciousness. *Brain*, 141(11):3179–3192.

[Escrichs et al., 2019] Escrichs, A., Sanjuán, A., Atasoy, S., López-González, A., Garrido, C., Càmara, E., and Deco, G. (2019). Characterizing the Dynamical Complexity Underlying Meditation. *Frontiers in Systems Neuroscience*, 13(July):2015–2020.

[Fahrenfort and van Gaal, 2020] Fahrenfort, J. J. and van Gaal, S. (2020). Criteria for empirical theories of consciousness should focus on the explanatory power of mechanisms, not on functional equivalence. *Cognitive Neuroscience*, 00(00):1–2.

[Faisal et al., 2008] Faisal, A. A., Selen, L. P. J., and Wolpert, D. M. (2008). Noise in the nervous system. *Nature Reviews Neuroscience*, 9(April):292–303.

[Feinberg, 2011] Feinberg, T. E. (2011). The nested neural hierarchy and the self. *Consciousness and Cognition*, 20(1):4–15.

[Fernández et al., 2014] Fernández, N., Maldonado, C., and Gershenson, C. (2014). Information Measures of Complexity, Emergence, Self-organization, Homeostasis, and Autopoiesis. In Prokopenko, M., editor, *Guided Self-Organization: Inception. Emergence, Complexity and Computation*, pages 19–51. Springer Berlin Heidelberg.

[Feshchenko et al., 2004] Feshchenko, V. A., Veselis, R. A., and Reinsel, R. A. (2004). Propofol-induced alpha rhythm. *Neuropsychobiology*, 50(3):257–266.

[Fields et al., 2018] Fields, C., Hoffman, D. D., Prakash, C., and Singh, M. (2018). Conscious agent networks: Formal analysis and application to cognition. *Cognitive Systems Research*, 47(October):186–213.

[Fosse et al., 2001] Fosse, R., Stickgold, R., and Hobson, J. A. (2001). Brain-mind states: Reciprocal variation in thoughts and hallucinations. *Psychological Science*, 12(1):30–36.

[Foster et al., 2017] Foster, J. A., Rinaman, L., and Cryan, J. F. (2017). Stress & the gut-brain axis: Regulation by the microbiome. *Neurobiology of Stress*, 7:124–136.

[Fox et al., 2016] Fox, K. C., Dixon, M. L., Nijeboer, S., Girn, M., Floman, J. L., Lifshitz, M., Ellamil, M., Sedlmeier, P., and Christoff, K. (2016). Functional neuroanatomy of meditation: A review and meta-analysis of 78 functional neuroimaging investigations. *Neuroscience and Biobehavioral Reviews*, 65:208–228.

[Fox et al., 2005] Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Van Essen, D. C., and Raichle, M. E. (2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proceedings of the National Academy of Sciences*, 102(27):9673–9678.

[Frankish, 2017] Frankish, K. (2017). *Illusionism*. Imprint Academic, Exeter.

[Franks, 2008] Franks, N. P. (2008). General anaesthesia: From molecular targets to neuronal pathways of sleep and arousal. *Nature Reviews Neuroscience*, 9(5):370–386.

[Fransson, 2006] Fransson, P. (2006). How default is the default mode of brain function? Further evidence from intrinsic BOLD signal fluctuations. *Neuropsychologia*, 44(14):2836–2845.

[Friedman et al., 2010] Friedman, E. B., Sun, Y., Moore, J. T., Hung, H. T., Meng, Q. C., Perera, P., Joiner, W. J., Thomas, S. A., Eckenhoff, R. G., Sehgal, A., and Kelz, M. B. (2010). A conserved behavioral state barrier impedes transitions between anesthetic-induced unconsciousness and wakefulness: Evidence for neural inertia. *PLoS ONE*, 5(7).

[Friston et al., 2020] Friston, K. J., Wiese, W., and Hobson, J. A. (2020). Sentience and the origins of consciousness: From cartesian duality to Markovian monism. *Entropy*, 22(5):1–31.

[Fukunaga et al., 2006] Fukunaga, M., Horovitz, S. G., van Gelderen, P., de Zwart, J. A., Jansma, J. M., Ikonomidou, V. N., Chu, R., Deckers, R. H., Leopold, D. A., and Duyn, J. H. (2006). Large-amplitude, spatially correlated fluctuations in BOLD fMRI signals during extended rest and early sleep stages. *Magnetic Resonance Imaging*, 24(8):979–992.

[Gallagher and Zahavi, 2008] Gallagher, S. and Zahavi, D. (2008). *The phenomenological mind*. Routledge, London, first edition.

[Garrison et al., 2015] Garrison, K. A., Zeffiro, T. A., Scheinost, D., Constable, R. T., and Brewer, J. A. (2015). Meditation leads to reduced default mode network activity beyond an active task. *Cognitive, Affective and Behavioral Neuroscience*, 15(3):712–720.

[Gershenson, 2013a] Gershenson, C. (2013a). Facing complexity: Prediction vs. adaptation. *Understanding Complex Systems*, pages 3–14.

[Gershenson, 2013b] Gershenson, C. (2013b). The Implications of Interactions for Science and Philosophy. *Foundations of Science*, 18(4):781–790.

[Gershenson, 2015] Gershenson, C. (2015). Requisite variety, autopoiesis, and self-organization. *Kybernetes*, 44(6-7):866–873.

[Ghosh et al., 2008] Ghosh, A., Rho, Y., McIntosh, A. R., Kötter, R., and Jirsa, V. K. (2008). Noise during rest enables the exploration of the brain's dynamic repertoire. *PLoS Computational Biology*, 4(10).

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

[Gilovich et al., 2002] Gilovich, T., Griffin, D., and Kahneman, D. (2002). *Heuristics and biases: The psychology of intuitive judgment*. Cambridge University Press, Cambridge, UK.

[Glennon et al., 1984] Glennon, R. A., Titeler, M., and McKenney, J. D. (1984). Evidence for 5-HT2 involvement in the mechanism of action of hallucinogenic agents. *Life Sciences*, 35(25):2505–2511.

[Glymour, 1980] Glymour, C. (1980). Explanations, Tests, Unity, and Necessity. *Nous*, 14:31–50.

[Gödel, 1931] Gödel, K. (1931). Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I. *Monatshefte für Mathematik und Physik*, 38(1):173–198.

[Goff, 2019] Goff, P. (2019). *Galileo's Error: Foundations for a New Science of Consciousness*. Phanteon Books, New York.

[Gómez-Ramirez, 2014] Gómez-Ramirez, J. (2014). *A New Foundation for Representation in Cognitive and Brain Science*. Springer Netherlands, Dordrecht.

[Gosseries et al., 2014] Gosseries, O., Di, H., Laureys, S., and Boly, M. (2014). Measuring consciousness in severely damaged brains. *Annual review of neuroscience*, 37:457–78.

[Greicius et al., 2008] Greicius, M. D., Kiviniemi, V., Tervonen, O., Vainionpää, V., Alahuhta, S., Reiss, A. L., and Menon, V. (2008). Persistent default-mode network connectivity during light sedation. *Human Brain Mapping*, 29(7):839–847.

[Grenham et al., 2011] Grenham, S., Clarke, G., Cryan, J. F., and Dinan, T. G. (2011). Brain-gut-microbe communication in health and disease. *Frontiers in Physiology*, 2 DEC(December):1–15.

[Guy and Staiger, 2017] Guy, J. and Staiger, J. F. (2017). The functioning of a cortex without layers. *Frontiers in Neuroanatomy*, 11(July):1–13.

[Hameroff and Penrose, 2014] Hameroff, S. and Penrose, R. (2014). Consciousness in the universe: A review of the "Orch OR" theory. *Physics of Life Reviews*, 11:39–78.

[Hannawi et al., 2015] Hannawi, Y., Lindquist, M., Caffo, B., Sair, H., and Stevens, R. (2015). Resting brain activity in disorders of consciousness. *Neurology*, 84:1272–1280.

[Hansen et al., 2015] Hansen, E. C. A., 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.

[Harrison and Connolly, 2013] Harrison, A. H. and Connolly, J. F. (2013). Finding a way in: A review and practical evaluation of fMRI and EEG for detection and assessment in disorders of consciousness. *Neuroscience and Biobehavioral Reviews*, 37(8):1403–1419.

[Haun and Tononi, 2019] Haun, A. and Tononi, G. (2019). Why does space feel theway it does? Towards a principled account of spatial experience. *Entropy*, 21(12).

[Haydon, 2017] Haydon, P. G. (2017). Astrocytes and the modulation of sleep. *Current Opinion in Neurobiology*, 44:28–33.

[Haydon and Carmignoto, 2006] Haydon, P. G. and Carmignoto, G. (2006). Astrocyte control of synaptic transmission and neurovascular coupling. *Physiological Reviews*, 86(3):1009–1031.

[Henry et al., 1999] Henry, T. R., Votaw, J. R., Pennell, P. B., Epstein, C. M., Bakay, R. A., Faber, T. L., Grafton, S. T., and Hoffman, J. M. (1999). Acute blood flow changes and efficacy of vagus nerve stimulation in partial epilepsy. *Neurology*, 52(6):1166–1173.

[Hentschke and Stüttgen, 2011] Hentschke, H. and Stüttgen, M. C. (2011). Computation of measures of effect size for neuroscience data sets. *The European journal of neuroscience*, 34(12):1887–94.

[Herbet et al., 2014] Herbet, G., Lafargue, G., de Champfleur, N. M., Moritz-Gasser, S., le Bars, E., Bonnetblanc, F., and Duffau, H. (2014). Disrupting posterior cingulate connectivity disconnects consciousness from the external environment. *Neuropsychologia*, 56(1):239–244.

[Herzog et al., 2016] Herzog, M. H., Kammer, T., and Scharnowski, F. (2016). Time Slices: What Is the Duration of a Percept? *PLoS biology*, 14(4):e1002433.

[Hilgetag and Goulas, 2020] Hilgetag, C. C. and Goulas, A. (2020). 'Hierarchy' in the organization of brain networks. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 375(1796).

[Hintze and Nelson, 1998] Hintze, J. L. and Nelson, R. D. (1998). Violin Plots: A Box Plot-Density Trace Synergism. *The American Statistician*, 52(2):181–184.

[Hobson, 2005] Hobson, J. A. (2005). Sleep is of the brain, by the brain and for the brain. *Nature*, 437(27):1254–1256.

[Hobson and Pace-Schott, 2002] Hobson, J. A. and Pace-Schott, E. F. (2002). The cognitive neuroscience of sleep: Neuronal systems, consciousness and learning. *Nature Reviews Neuroscience*, 3(9):679–693.

[Hoffman, 2008] Hoffman, D. D. (2008). Conscious Realism and the Mind-Body Problem. *Mind and Matter*, 6(1):87–121.

[Hoffman and Prakash, 2014] Hoffman, D. D. and Prakash, C. (2014). Objects of consciousness. *Frontiers in Psychology*, 5(JUN):1–22.

[Hohwy and Frith, 2004] Hohwy, J. and Frith, C. (2004). Can neuroscience explain consciousness? *Journal of Consciousness Studies*, 11(7-8):180–198.

[Honey and Sporns, 2008] Honey, C. J. and Sporns, O. (2008). Dynamical consequences of lesions in cortical networks. *Human Brain Mapping*, 29(7):802–809.

[Honey et al., 2009] 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. *PNAS*, 106(6):2035–2040.

[Honey et al., 2010] Honey, C. J., Thivierge, J. P., and Sporns, O. (2010). Can structure predict function in the human brain? *NeuroImage*, 52(3):766–776.

[Horgan et al., 2004] Horgan, T. E., Tienson, J. L., and Graham, G. (2004). Phenomenal intentionality and the brain in a vat. In Schantz, R., editor, *The Externalist Challenge*. De Gruyter.

[Horovitz et al., 2009] Horovitz, S. G., Braun, A. R., Carr, W. S., Picchioni, D., Balkin, T. J., Fukunaga, M., and Duyn, J. H. (2009). Decoupling of the brain's default mode network during deep sleep. *Proceedings of the National Academy of Sciences*, 106(27):11376–11381.

[Huang et al., 2014] Huang, Z., Wang, Z., Zhang, J., Dai, R., Wu, J., Li, Y., Liang, W., Mao, Y., Yang, Z., Holland, G., Zhang, J., and Northoff, G. (2014). Altered temporal variance and neural synchronization of spontaneous brain activity in anesthesia. *Human Brain Mapping*, 35(11):5368–5378.

[Huber et al., 2020] Huber, L., Finn, E. S., Chai, Y., Goebel, R., Stirnberg, R., Stöcker, T., Marrett, S., Uludag, K., Kim, S. G., Han, S. H., Bandettini, P. A., and Poser, B. A. (2020). Layer-dependent functional connectivity methods. *Progress in Neurobiology*, (March):101835.

[Hudetz and Mashour, 2016] Hudetz, A. G. and Mashour, G. A. (2016). Disconnecting Consciousness: Is There a Common Anesthetic End Point? *Anesthesia & Analgesia*, 123(5):1228–1240.

[Hudson et al., 2014] Hudson, A. E., Calderon, D. P., Pfaff, D. W., and Proekt, A. (2014). Recovery of consciousness is mediated by a network of discrete metastable activity states. *Proceedings of the National Academy of Sciences of the United States of America*, 111(25):9283–9288.

[Illari and Williamson, 2012] Illari, P. M. K. and Williamson, J. (2012). What is a mechanism? Thinking about mechanisms across the sciences. *European Journal for Philosophy of Science*, 2(1):119–135.

[Ipiña et al., 2020] 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(April).

[Ivanov et al., 2017] Ivanov, P., Liu, K., Lin, A., and Bartsch, R. (2017). Network Physiology: From Neural Plasticity to Organ Network Interactions. In Mantica, G., Stoop, R., and Stramaglia, S., editors, *Emergent Complexity from Nonlinearity, in Physics, Engineering and the Life Sciences*, volume 191, pages 145–165. Springer, Cham, springer p edition.

[Jha and Morrison, 2018] Jha, M. K. and Morrison, B. M. (2018). Glia-neuron energy metabolism in health and diseases: New insights into the role of nervous system metabolic transporters. *Experimental Neurology*, 309(July):23–31.

[Jobst et al., 2017] Jobst, B. M., Hindriks, R., Laufs, H., Tagliazucchi, E., Hahn, G., Ponce-al, A., Stevner, A. B. 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(February):1–16.

[Joglekar et al., 2018] Joglekar, M. R., Mejias, J. F., Yang, G. R., and Wang, X. J. (2018). Inter-areal Balanced Amplification Enhances Signal Propagation in a Large-Scale Circuit Model of the Primate Cortex. *Neuron*, 98(1):222–234.e8.

[Kahneman, 2003] Kahneman, D. (2003). A perspective on judgment and choice: mapping bounded rationality. *The American psychologist*, 58(9):697–720.

[Kaisti et al., 2003] Kaisti, K. K., Långsjö, J. W., Aalto, S., Oikonen, V., Sipilä, H., Teräs, M., Hinkka, S., Metsähonkala, L., and Scheinin, H. (2003). Effects of sevoflurane, propofol, and adjunct nitrous oxide on regional cerebral blood flow, oxygen consumption and blood volume in humans. *Anesthesiology*, 99(3):603–613.

[Kaisti et al., 2002] Kaisti, K. K., Metsähonkala, L., Teräs, M., Oikonen, V., Aalto, S., Jääskeläinen, S., Hinkka, S., and Scheinin, H. (2002). Effects of surgical levels of propofol and sevoflurane anesthesia on cerebral blood flow in healthy subjects studied with positron emission tomography. *Anesthesiology*, 96(6):1358–1370.

[Kastrup, 2017] Kastrup, B. (2017). An Ontological Solution to the Mind-Body Problem. *Philosophies*, 2(4):10.

[Kaufer and Chemero, 2015] Kaufer, S. and Chemero, A. (2015). *Phenomenology: An Introduction*. Polity, New York.

[Kennedy et al., 2013] Kennedy, H., Knoblauch, K., and Toroczkai, Z. (2013). Why data coherence and quality is critical for understanding interareal cortical networks. *NeuroImage*, 80:37–45.

[King et al., 2013] King, J. R., Faugeras, F., Gramfort, A., Schurger, A., El Karoui, I., Sitt, J. D., Rohaut, B., Wacongne, C., Labyt, E., Bekinschtein, T., Cohen, L., Naccache, L., and Dehaene, S. (2013). Single-trial decoding of auditory novelty responses facilitates the detection of residual consciousness. *NeuroImage*, 83:726–738.

[Kivela et al., 2014] Kivela, M., Arenas, A., Barthelemy, M., Gleeson, J. P., Moreno, Y., and Porter, M. A. (2014). Multilayer networks. *Journal of complex networks*, 2(July):203–271.

[Kiverstein and Miller, 2015] Kiverstein, J. and Miller, M. (2015). The embodied brain: towards a radical embodied cognitive neuroscience. *Frontiers in Human Neuroscience*, 9.

[Kleiner, 2019] Kleiner, J. (2019). On empirical well-definedness of models of consciousness. *PsyArXiv*, pages 1–6.

[Kleiner, 2020] Kleiner, J. (2020). Mathematical Models of Consciousness. *Entropy*, 22(6):609.

[Kleiner and Tull, 2020] Kleiner, J. and Tull, S. (2020). The mathematical structure of integrated information theory. *arXiv*, pages 1–22.

[Klink et al., 2015] Klink, P. C., Self, M. W., Lamme, V. A., and Roelfsema, P. R. (2015). Theories and methods in the scientific study of consciousness. In Miller, S. M., editor, *The Constitution of Phenomenal Consciousness. Toward a science and theory.*, pages 17–47. John Benjamins Publishing Company.

[Koch et al., 2016] Koch, C., Massimini, M., Boly, M., and Tononi, G. (2016). Neural correlates of consciousness: progress and problems. *Nature reviews. Neuroscience*, 17(5):307–21.

[Koler and Von Grünau, 1976] Koler, P. A. and Von Grünau, M. (1976). Shape and Color in apparent motion.

[Kötter and Wanke, 2005] Kötter, R. and Wanke, E. (2005). Mapping brains without coordinates. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1456):751–766.

[Koubeissi et al., 2014] Koubeissi, M. Z., Bartolomei, F., Beltagy, A., and Picard, F. (2014). Electrical stimulation of a small brain area reversibly disrupts consciousness. *Epilepsy & Behavior*, 37:32–35.

[Kozma and Freeman, 2017] Kozma, R. and Freeman, W. J. (2017). Cinematic operation of the cerebral cortex interpreted via critical transitions in self-organized dynamic systems. *Frontiers in Systems Neuroscience*, 11(March):1–10.

[Kringelbach et al., 2020] 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 of the United States of America*, 117(17):9566–9576.

[Laitio et al., 2007] Laitio, R. M., Kaisti, K. K., Langsjo, J. W., Aalto, S., Salmi, E., Maksimow, A., Aantaa, R., Oikonen, V., Sipila, H., Parkkola, R., and Scheinin, H. (2007). Effects of Xenon Anesthesia on Cerebral Blood Flow in Humans. *Anesthesiology*, 106:1128–1133.

[Lamme, 2003] Lamme, V. A. (2003). Why visual attention and awareness are different.

[Lamme, 2010] Lamme, V. A. (2010). How neuroscience will change our view on consciousness. *Cognitive Neuroscience*, 1(3):204–220.

[Langsjo et al., 2005] Langsjo, J., Maksimow, A., Salmi, E., Kaisti, K., Aalto, S., Oikonen, V., Hinkka, S., Aantaa, R., Sipila, H., Viljanen, T., Parkkola, R., and Scheinin, H. (2005). S-ketamine anesthesia increases cerebral blood flow in excess of the metabolic needs in humans. *Anesthesiology*, 103(2):258–268.

[Laumann et al., 2017] Laumann, T. O., Snyder, A. Z., Mitra, A., Gordon, E. M., Gratton, C., Adeyemo, B., Gilmore, A. W., Nelson, S. M., Berg, J. J., Greene, D. J., McCarthy, J. E., Tagliazucchi, E., Laufs, H., Schlaggar, B. L., Dosenbach, N. U., and Petersen, S. E. (2017). On the Stability of BOLD fMRI Correlations. *Cerebral Cortex*, 27(10):4719–4732.

[Laureys, 2005] Laureys, S. (2005). The neural correlate of (un) awareness: lessons from the vegetative state. *Trends in cognitive sciences*, 9(12):556–559.

[Laureys et al., 2006] Laureys, S., Boly, M., and Maquet, P. (2006). Tracking the recovery of consciousness from coma. *Journal of Clinical Investigation*, 116(7):1823–1825.

[Laureys et al., 1999] Laureys, S., Lemaire, C., and Maquet, P. (1999). Cerebral metabolism during vegetative state and after recovery to consciousness. *J Neurol Neurosurg Psychiatry*, 67:121.

[Lee et al., 2013] Lee, U., Ku, S., Noh, G., Baek, S., Choi, B., and Mashour, G. a. (2013). Disruption of Frontal-Parietal Communication. *Anesthesiology*, 118(6):1264–1275.

[Lee et al., 2011] Lee, U., Müller, M., Noh, G. J., Choi, B., and Mashour, G. A. (2011). Dissociable network properties of anesthetic state transitions. *Anesthesiology*, 114(4):872–881.

[Lepousé et al., 2006] Lepousé, C., Lautner, C. A., Liu, L., Gomis, P., and Leon, A. (2006). Emergence delirium in adults in the post-anaesthesia care unit. *British Journal of Anaesthesia*, 96(6):747–753.

[Letelier et al., 2011] Letelier, J. C., Cárdenas, M. L., and Cornish-Bowden, A. (2011). From L'Homme Machine to metabolic closure: Steps towards understanding life. *Journal of Theoretical Biology*, 286(1):100–113.

[Letelier et al., 2006] Letelier, J. C., Soto-Andrade, J., Guíñez Abarzúa, F., Cornish-Bowden, A., and Luz Cárdenas, M. (2006). Organizational invariance and metabolic closure: Analysis in terms of (M, R) systems. *Journal of Theoretical Biology*, 238(4):949–961.

[Lewis et al., 2012] Lewis, L. D., Weiner, V. S., Mukamel, E. A., Donoghue, J. A., Eskandar, E. N., Madsen, J. R., Anderson, W. S., Hochberg, L. R., Cash, S. S., Brown, E. N., and Purdon, P. L. (2012). Rapid fragmentation of neuronal networks at the onset of propofol-induced unconsciousness. *Proceedings of the National Academy of Sciences*, 109(49):E3377–E3386.

[Llinás, 2003] Llinás, R. (2003). Consciousness and the thalamocortical loop. *International Congress Series*, 1250(C):409–416.

[Llinas et al., 1998] Llinas, R., Ribary, U., Contreras, D., and Pedroarena, C. (1998). The neuronal basis for consciousness. *Philosophical Transactions of the Royal Society of London B*, 353:1841–1849.

[Llinás and Paré, 1991] Llinás, R. R. and Paré, D. (1991). Of dreaming and wakefulness. *Neuroscience*, 44(3):521–535.

[López-Ruiz et al., 1995] López-Ruiz, R., Mancini, H. L., and Calbet, X. (1995). A statistical measure of complexity. *Physics Letters A*, 209(5-6):321–326.

[LoTurco and Booker, 2013] LoTurco, J. and Booker, A. (2013). Neuronal Migration Disorders. *Cellular Migration and Formation of Neuronal Connections*, pages 481–494.

[Luppi et al., 2020] Luppi, A. I., Vohryzek, J., Mediano, P. A. M., Adapa, R., Pappas, I., Finoia, P., Allanson, J., Atasoy, S., and Stamatakis, E. A. (2020). Connectome Harmonic Decomposition of Human Brain Dynamics Reveals a Landscape of Consciousness. *bioRxiv*.

[Lutz et al., 2004] Lutz, A., Greischar, L., Rawlings, N., Ricard, M., and Davidson, R. (2004). Long-term meditators self-induce high-amplitude gamma synchrony during mental practice. *Proceedings of the National Academy of Sciences*, 101(46):16369–16373.

[Lutz et al., 2002] 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 of the United States of America*, 99(3):1586–1591.

[Lutz et al., 2016] Lutz, J., Brühl, A. B., Scheerer, H., Jäncke, L., and Herwig, U. (2016). Neural correlates of mindful self-awareness in mindfulness meditators and meditation-naïve subjects revisited. *Biological Psychology*, 119:21–30.

[Lynn et al., 2020] Lynn, C. W., Cornblath, E. J., Papadopoulos, L., Bertolero, M. A., and Bassett, D. S. (2020). Non-equilibrium dynamics and entropy production in the human brain. *arXiv*, pages 1–18.

[Machina, 2009] Machina, M. (2009). Risk, Ambiguity, and the Rank-Dependence axioms. *The American Economic Review*, 99(1):385–392.

[Magistretti and Allaman, 2015] Magistretti, P. J. and Allaman, I. (2015). A Cellular Perspective on Brain Energy Metabolism and Functional Imaging. *Neuron*, 86(4):883–901.

[Maillé and Lynn, 2020] Maillé, S. and Lynn, M. (2020). Reconciling current theories of consciousness. *Journal of Neuroscience*, 40(10):1994–1996.

[Maquet, 1995] Maquet, P. (1995). Sleep function(s) and cerebral metabolism. *Behavioural Brain Research*, 69(1-2):75–83.

[Maquet et al., 1996] Maquet, P., Peters, J. M., Aerts, J., Delfiore, G., Degueldre, C., Luxen, A., and Franck, G. (1996). Functional neuroanatomy of human rapid-eye-movement sleep and dreaming. *Nature*, 383(6596):163–166.

[Martuzzi et al., 2010] Martuzzi, R., Ramani, R., Qiu, M., Rajeevan, N., and Constable, R. T. (2010). Functional connectivity and alterations in baseline brain state in humans. *NeuroImage*, 49(1):823–834.

[Mashour et al., 2020] 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.

[Massimini et al., 2009] Massimini, M., Boly, M., Casali, A., Rosanova, M., and Tononi, G. (2009). *A perturbational approach for evaluating the brain's capacity for consciousness.*, volume 177. Elsevier.

[Massimini et al., 2004] Massimini, M., Huber, R., Ferrarelli, F., Hill, S., and Tononi, G. (2004). The sleep slow oscillation as a traveling wave. *Journal of Neuroscience*, 24(31):6862–6870.

[Mateos et al., 2018] Mateos, D. M., Guevara Erra, R., Wennberg, R., and Perez Velazquez, J. L. (2018). Measures of entropy and complexity in altered states of consciousness. *Cognitive Neurodynamics*, 12(1):73–84.

[Maturana and Varela, 1998] Maturana, H. and Varela, F. (1998). *De máquinas y seres vivos*. Editorial Universitaria S.A., Santiago de Chile, quinta edition.

[Maturana, 2011] Maturana, H. R. (2011). Ultrastability ... Autopoiesis? Reflective Response to Tom Froese and John Stewart. *Cybernetics and Human Knowing*, 18(4):143–152.

[Mayer, 2011] Mayer, E. A. (2011). Gut feelings: The emerging biology of gut-brain communication. *Nature Reviews Neuroscience*, 12(8):453–466.

[Mayer et al., 2015] Mayer, E. A., Tillisch, K., Gupta, A., Mayer, E. A., Tillisch, K., and Gupta, A. (2015). Gut / brain axis and the microbiota Find the latest version : Gut / brain axis and the microbiota. *Journal of Clinical Investiogation*, 125(3):926–938.

[Mazzocchi, 2008] Mazzocchi, F. (2008). Complexity in biology. *EMBO reports*, 9(1):10–14.

[McComas and Cupido, 1999] McComas, A. J. and Cupido, C. M. (1999). The RULER model. Is this how the somatosensory cortex works? *Clinical Neurophysiology*, 110(11):1987–1994.

[McCulloch and Pitts, 1943] McCulloch, W. S. and Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, 5(4):115–133.

[McGinn, 1988] McGinn, C. (1988). Consciousness and content. In *Proceedings of the British Academy*, volume 74, pages 219–239.

[McIntosh, 2000] McIntosh, A. (2000). Towards a network theory of cortical areas. *Neural Networks*, 13:861–870.

[Meisel et al., 2012] Meisel, C., Storch, A., Hallmeyer-Elgner, S., Bullmore, E., and Gross, T. (2012). Failure of adaptive self-organized criticality during epileptic seizure attacks. *PLoS Computational Biology*, 8(1).

[Merleau-Ponty, 2005] Merleau-Ponty, M. (2005). *Phenomenology of perception*. Routledge, London.

[Mesulam, 1990] Mesulam, M. (1990). Large-Scale Neurocognitive Networks and Distributed Processing for Attention, Language, and Memory. *Neurological Progress. Annals of Neurology*, 28:597'613.

[Mesulam, 1998] Mesulam, M. M. (1998). From sensation to cognition. *Brain*, 121(6):1013–1052.

[Metzinger, 2020] Metzinger, T. (2020). Minimal phenomenal experience. *Philosophy and the Mind Sciences*, 1(I):7.

[Michel et al., 2019] Michel, M., Beck, D., Block, N., Blumenfeld, H., Brown, R., Carmel, D., Carrasco, M., Chirimuuta, M., Chun, M., Cleeremans, A., Dehaene, S., Fleming, S. M., Frith, C., Haggard, P., He, B. J., Heyes, C., Goodale, M. A., Irvine, L., Kawato, M., Kentridge, R., King, J. R., Knight, R. T., Kouider, S., Lamme, V., Lamy, D., Lau, H., Laureys, S., LeDoux, J., Lin, Y. T., Liu, K., Macknik, S. L., Martinez-Conde, S., Mashour, G. A.,

Melloni, L., Miracchi, L., Mylopoulos, M., Naccache, L., Owen, A. M., Passingham, R. E., Pessoa, L., Peters, M. A., Rahnev, D., Ro, T., Rosenthal, D., Sasaki, Y., Sergent, C., Solovey, G., Schiff, N. D., Seth, A., Tallon-Baudry, C., Tamietto, M., Tong, F., van Gaal, S., Vlassova, A., Watanabe, T., Weisberg, J., Yan, K., and Yoshida, M. (2019). Opportunities and challenges for a maturing science of consciousness. *Nature Human Behaviour*, 3(2):104–107.

[Miłkowski, 2018] Miłkowski, M. (2018). From Computer Metaphor to Computational Modeling: The Evolution of Computationalism. *Minds and Machines*, 28(3):515–541.

[Millière et al., 2018] Millière, R., Carhart-Harris, R. L., Roseman, L., Trautwein, F. M., and Berkovich-Ohana, A. (2018). Psychedelics, meditation, and self-consciousness. *Frontiers in Psychology*, 9(SEP).

[Moore, 2002] Moore, D. (2002). Measuring new types of question-order effects: Additive and subtractive. *The Public Opinion Quarterly*, 66(1):80–91.

[Moore, 2005] Moore, G. W. (2005). What is a brane? *Notices of the American Mathematical Society*, 52(2):214–215.

[Muthukumaraswamy et al., 2013] 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., Feilding, A., Friston, K. J., and Nutt, D. J. (2013). Broadband cortical desynchronization underlies the human psychedelic state. *Journal of Neuroscience*, 33(38):15171–15183.

[Naccache, 2018] Naccache, L. (2018). Why and how access consciousness can account for phenomenal consciousness. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(1755).

[Nagel, 1961] Nagel, E. (1961). *The Structure of Science: Problems in the Logic of Scientific Explanation.* Harcourt, New York.

[Nagel, 1986] Nagel, T. (1986). *The View From Nowhere*. Oxford University Press.

[Nguyen et al., 2016] Nguyen, V. T., Breakspear, M., Hu, X., and Guo, C. C. (2016). The integration of the internal and external milieu in the insula during dynamic emotional experiences. *NeuroImage*, 124:455–463.

[Nichols et al., 2017] Nichols, A. L. A., Eichler, T., Latham, R., and Zimmer, M. (2017). A global brain state underlies C. elegans sleep behavior. *Science*, 356(1247).

[Nir et al., 2011] Nir, Y., Staba, R. J., Andrillon, T., Vyazovskiy, V. V., Cirelli, C., Fried, I., and Tononi, G. (2011). Regional Slow Waves and Spindles in Human Sleep. *Neuron*, 70(1):153–169.

[Noë and Thompson, 2004] Noë, A. and Thompson, E. (2004). Are There Neural correlates of consciousness? *Journal of Consciousness Studies*, 11(1):3–28.

[Noel et al., 2019] Noel, J. P., Ishizawa, Y., Patel, S. R., Eskandar, E. N., and Wallace, M. T. (2019). Leveraging Nonhuman Primate Multisensory Neurons and Circuits in Assessing Consciousness Theory. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 39(38):7485–7500.

[Northoff and Huang, 2017] Northoff, G. and Huang, Z. (2017). How do the brain's time and space mediate consciousness and its different dimensions? Temporo-spatial theory of consciousness (TTC). *Neuroscience and Biobehavioral Reviews*, 80(May):630–645.

[Northoff and Lamme, 2020] Northoff, G. and Lamme, V. (2020). Neural signs and mechanisms of consciousness: Is there a potential convergence of theories of consciousness in sight? *Neuroscience and Biobehavioral Reviews*, 118(July):568–587.

[Northoff et al., 2019] Northoff, G., Tsuchiya, N., and Saigo, H. (2019). Mathematics and the Brain: A Category Theoretical Approach to Go Beyond the Neural Correlates of Consciousness. *Entropy*, 21(12):1234.

[Northoff et al., 2011] Northoff, G., Wiebking, C., Feinberg, T., and Panksepp, J. (2011). The 'resting-state hypothesis' of major depressive disorder-A translational subcortical-cortical framework for a system disorder. *Neuroscience and Biobehavioral Reviews*, 35(9):1929–1945.

[Nowogrodzki, 2018] Nowogrodzki, A. (2018). The Strongest Scanners. *Nature*, 563(November):24–26.

[O'Connor and Adams, 2010] O'Connor, C. M. and Adams, J. U. (2010). *Essentials of Cell Biology*. NPG Education, Cambridge, MA.

[Oizumi et al., 2014] 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.

[Oizumi et al., 2016] Oizumi, M., Tsuchiya, N., and Amari, S. I. (2016). Unified framework for information integration based on information geometry. *Proceedings of the National Academy of Sciences of the United States of America*, 113(51):14817–14822.

[Olivares et al., 2015] Olivares, F. A., Vargas, E., Fuentes, C., Martínez-Pernía, D., and Canales-Johnson, A. (2015). Neurophenomenology revisited: Second-person methods for the study of human consciousness. *Frontiers in Psychology*, 6(May):1–12.

[O'Mahony et al., 2011] O'Mahony, S. M., Hyland, N. P., Dinan, T. G., and Cryan, J. F. (2011). Maternal separation as a model of brain-gut axis dysfunction. *Psychopharmacology*, 214(1):71–88.

[Pace-Schott and Hobson, 2002] Pace-Schott, E. F. and Hobson, J. A. (2002). The neurobiology of sleep: Genetics, cellular physiology and subcortical networks. *Nature Reviews Neuroscience*, 3(8):591–605.

[Park et al., 2014] Park, H. D., Correia, S., Ducorps, A., and Tallon-Baudry, C. (2014). Spontaneous fluctuations in neural responses to heartbeats predict visual detection. *Nature Neuroscience*, 17(4):612–618.

[Park and Tallon-Baudry, 2014] Park, H. D. and Tallon-Baudry, C. (2014). The neural subjective frame: From bodily signals to perceptual consciousness. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1641).

[Parr et al., 2019] Parr, T., Corcoran, A. W., Friston, K. J., and Hohwy, J. (2019). Perceptual awareness and active inference. *Neuroscience of Consciousness*, 2019(1):1–15.

[Pautz, 2014] Pautz, A. (2014). The Real Trouble with Phenomenal Externalism: New Empirical Evidence for a Brain-Based Theory of Consciousness. In Brown, R., editor, *Consciousness Inside and Out: Phenomenology, Neuroscience, and the Nature of Experience*, pages 237–317. Springer, Dordrecht.

[Peduto et al., 1991] Peduto, V. A., Concas, A., Santoro, G., Biggio, G., and Gessa, G. L. (1991). Biochemical and Electrophysiologic Evidence That Propofol Enhances GABAergic Transmission in the Rat Brain. *Anesthesiology*, 75(6):1000–1009.

[Perouansky et al., 2019] Perouansky, M., Maciver, M. B., and Pearce, R. A. (2019). Wake Up, Neurons! Astrocytes Calling. *Anesthesiology*, 130(3):361–363.

[Pessoa, 2014] Pessoa, L. (2014). Understanding brain networks and brain organization. *Physics of Life Reviews*, 11(3):400–435.

[Petit and Magistretti, 2016] Petit, J. M. and Magistretti, P. J. (2016). Regulation of neuron-astrocyte metabolic coupling across the sleep-wake cycle. *Neuroscience*, 323:135–156.

[Petitmengin et al., 2019] Petitmengin, C., Remillieux, A., and Valenzuela-Moguillansky, C. (2019). Discovering the structures of lived experience. *Phenomenology and the Cognitive Sciences*, 18(4):691–730.

[Picchioni et al., 2008] Picchioni, D., Fukunaga, M., Carr, W. S., Braun, A. R., Balkin, T. J., Duyn, J. H., and Horovitz, S. G. (2008). fMRI differences between early and late stage-1 sleep. *Neuroscience Letters*, 441(1):81–85.

[Piccinini, 2004] Piccinini, G. (2004). The first computational theory of mind and brain: A close look at McCulloch and Pitts's "logical calculus of ideas immanent in nervous activity". *Synthese*, 141(2):175–215.

[Pigorini et al., 2015] Pigorini, A., Sarasso, S., Proserpio, P., Szymanski, C., Arnulfo, G., Casarotto, S., Fecchio, M., Rosanova, M., Mariotti, M., Lo Russo, G., Palva, J. M., Nobili, L., and Massimini, M. (2015). Bistability breaks-off deterministic responses to intracortical stimulation during non-REM sleep. *NeuroImage*, 112:105–113.

[Posner et al., 2008] Posner, J. B., Saper, C. B., Schiff, N., and Plum, F. (2008). *Plum and Posner's Diagnosis of Stupor and Coma*. Oxford University Press.

[Pothos and Busemeyer, 2013] Pothos, E. M. and Busemeyer, J. R. (2013). Can quantum probability provide a new direction for cognitive modeling? *The Behavioral and brain sciences*, 36(3):255–74.

[Prakash et al., 2020] Prakash, C., Fields, C., Hoffman, D. D., Prentner, R., and Singh, M. (2020). Fact, fiction, and fitness. *Entropy*, 22(5):1–23.

[Prentner, 2017] Prentner, R. (2017). Consciousness: A Molecular Perspective. *Philosophies*, 2(4):26.

[Prentner, 2019] Prentner, R. (2019). Consciousness and topologically structured phenomenal spaces. *Consciousness and Cognition*, 70(February):25–38.

[Priesemann et al., 2013] Priesemann, V., Valderrama, M., Wibral, M., and Le Van Quyen, M. (2013). Neuronal Avalanches Differ from Wakefulness to Deep Sleep - Evidence from Intracranial Depth Recordings in Humans. *PLoS Computational Biology*, 9(3).

[Priesemann et al., 2014] Priesemann, V., Wibral, M., Valderrama, M., Pröpper, R., Le Van Quyen, M., Geisel, T., Triesch, J., Nikolić, D., and Munk, M. H. (2014). Spike avalanches in vivo suggest a driven, slightly subcritical brain state. *Frontiers in Systems Neuroscience*, 8(JUNE):1–17.

[Proekt and Hudson, 2018] Proekt, A. and Hudson, A. E. (2018). A stochastic basis for neural inertia in emergence from general anaesthesia. *British Journal of Anaesthesia*, 121(1):86–94.

[Ramadasan-Nair et al., 2019] Ramadasan-Nair, R., Hui, J., Itsara, L. S., Morgan, P. G., and Sedensky, M. M. (2019). Mitochondrial Function in Astrocytes Is Essential for Normal Emergence from Anesthesia in Mice. *Anesthesiology*, 130(3):423–434.

[Ranft et al., 2016] Ranft, A., Golkowski, D., Kiel, T., Riedl, V., Kohl, P., Rohrer, G., Pientka, J., Berger, S., Thul, A., Maurer, M., Preibisch, C., Zimmer, C., Mashour, G. A., Kochs, E. F., Jordan, D., and Ilg, R. (2016). Neural Correlates of Sevoflurane-induced Unconsciousness Identified by Simultaneous Functional Magnetic Resonance Imaging and Electroencephalography. *Anesthesiology*, 125(5):861–872.

[Reardon, 2019] Reardon, S. (2019). Rival theories face off over brain's source of consciousness. *Science*, 366(6463):293.

[Redinbaugh et al., 2020] Redinbaugh, M. J., Phillips, J. M., Kambi, N. A., Afrasiabi, M., Raz, A., Saalmann, Y. B., Redinbaugh, M. J., Phillips, J. M., Kambi, N. A., Mohanta, S., and Andryk, S. (2020). Thalamus Modulates Consciousness via Layer- Specific Control of Cortex Thalamus Modulates Consciousness via Layer-Specific Control of Cortex. *Neuron*, 0(0):1–10.

[Rescher, 2012] Rescher, N. (2012). Process Philosophy.

[Reutlinger, 2017] Reutlinger, A. (2017). Explanation beyond causation? New directions in the philosophy of scientific explanation. *Philosophy Compass*, 12(2):1–11.

[Rhee et al., 2009] Rhee, S. H., Pothoulakis, C., and Mayer, E. A. (2009). Principles and clinical implications of the brain-gut-enteric microbiota axis. *Nature Reviews Gastroenterology and Hepatology*, 6(5):306–314.

[Rodríguez, 2008] Rodríguez, E. (2008). Ideas para naturalizar el estudio de la conciencia. In Edmundo Kronmüller and Cornejo, C., editors, *Ciencias de la mente: aproximaciones desde Latinoamerica*, pages 301–324. Juan Carlos Sáez Editor.

[Rosanova et al., 2012] Rosanova, M., Gosseries, O., Casarotto, S., Boly, M., Casali, A. G., Bruno, M. A., Mariotti, M., Boveroux, P., Tononi, G., Laureys, S., and Massimini, M. (2012). Recovery of cortical effective connectivity and recovery of consciousness in vegetative patients. *Brain*, 135(4):1308–1320.

[Rosenthal, 2020] Rosenthal, D. (2020). Assessing criteria for theories. *Cognitive Neuroscience*, 00(00):1–2.

[Rosenthal, 2002a] Rosenthal, D. M. (2002a). Explaining consciousness. In *Philosophy of mind: Classical and contemporary readings.*, pages 406–421.

[Rosenthal, 2002b] Rosenthal, D. M. (2002b). How many kinds of consciousness? *Consciousness and Cognition*, 11(4):653–665.

[Rosenthal, 2008] Rosenthal, D. M. (2008). Consciousness and its function. *Neuropsychologia*, 46(3):829–840.

[Rosner, 2012] Rosner, B. (2012). *Fundamentals of biostatistics*. Cengage Learning, seventh edition.

[Ruiz-Mirazo and Moreno, 2004] Ruiz-Mirazo, K. and Moreno, A. (2004). Basic autonomy as a fundamental step in the synthesis of life. *Artificial Life*, 10(3):235–259.

[Saenger et al., 2017] Saenger, V. M., Kahan, J., Foltynie, T., Friston, K., Aziz, T. Z., Green, A. L., Van Hartevelt, T. J., Cabral, J., Stevner, A. B., Fernandes, H. M., Mancini, L., Thornton, J., Yousry, T., Limousin, P., Zrinzo, L., Hariz, M., Marques, P., Sousa, N., Kringelbach, M. L., and Deco, G. (2017). Uncovering the underlying mechanisms and whole-brain dynamics of deep brain stimulation for Parkinson's disease. *Scientific Reports*, 7(1):1–14.

[Salmon, 1990] Salmon, W. C. (1990). Scientific Explanation: Causation and Unification. *Crítica (México D. F. En línea)*, 22(66):3–23.

[Salomon et al., 2016] Salomon, R., Ronchi, R., Dönz, J., Bello-Ruiz, J., Herbelin, B., Martet, R., Faivre, N., Schaller, K., and Blanke, O. (2016). The insula mediates access to awareness of visual stimuli presented synchronously to the heartbeat. *Journal of Neuroscience*, 36(18):5115–5127.

[Sandberg et al., 2010] Sandberg, K., Timmermans, B., Overgaard, M., and Cleeremans, A. (2010). Measuring consciousness: Is one measure better than the other? *Consciousness and Cognition*, 19(4):1069–1078.

[Sanz and Tagliazucchi, 2018] Sanz, C. and Tagliazucchi, E. (2018). The experience elicited by hallucinogens presents the highest similarity to dreaming within a large database of psychoactive substance reports. *Frontiers in Neuroscience*, 12(JAN):1–19.

[Saper, 2002] Saper, C. B. (2002). The Central Autonomic Nervous System: Conscious Visceral Perception and Autonomic Pattern Generation. *Annual Review of Neuroscience*, 25(1):433–469.

[Sarasso et al., 2014] Sarasso, S., Rosanova, M., Casali, A. G., Casarotto, S., Fecchio, M., Boly, M., Gosseries, O., Tononi, G., Laureys, S., and Massimini, M. (2014). Quantifying cortical EEG responses to TMS in (Un)consciousness. *Clinical EEG and Neuroscience*, 45(1):40–49.

[Scarone et al., 2008] Scarone, S., Manzone, M. L., Gambini, O., Kantzas, I., Limosani, I., Agostino, A. D., and Hobson, J. A. (2008). The Dream as a Model for Psychosis: An Experimental Approach Using Bizarreness as a Cognitive Marker. *Schizophrenia Bulletin*, 34(3):515–522.

[Scheibner et al., 2017] Scheibner, H. J., Bogler, C., Gleich, T., Haynes, J. D., and Bermpohl, F. (2017). Internal and external attention and the default mode network. *NeuroImage*, 148:381–389.

[Schenberg et al., 2015] Schenberg, E. E., Alexandre, J. F. M., Filev, R., Cravo, A. M., Sato, J. R., Muthukumaraswamy, S. D., Yonamine, M., Waguespack, M., Lomnicka, I., Barker, S. A., and Da Silveira, D. X. (2015). Acute biphasic effects of ayahuasca. *PLoS ONE*, 10(9):1–27.

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

[Schiff, 2013] Schiff, N. D. (2013). Central thalamic deep brain stimulation for support of forebrain arousal regulation in the minimally conscious state. In Lozano, A. and Hallett, M., editors, *Handbook of Clinical Neurology*, volume 116, chapter 24, pages 295–306. Elsevier B.V.

[Schiff et al., 2007] Schiff, N. D., Giacino, J. T., Kalmar, K., Victor, J. D., Baker, K., Gerber, M., Fritz, B., Eisenberg, B., O'Connor, J., Kobylarz, E. J., Farris, S., Machado, A., McCagg, C., Plum, F., Fins, J. J., and Rezai, A. R. (2007). Behavioural improvements with thalamic stimulation after severe traumatic brain injury. *Nature*, 448(7153):600–603.

[Schroeder et al., 2016] Schroeder, K. E., Irwin, Z. T., Gaidica, M., Bentley, J. N., Patil, P. G., Mashour, G. A., and Chestek, C. A. (2016). Disruption of corticocortical information transfer during ketamine anesthesia in the primate brain. *NeuroImage*, 134:459–465.

[Schrouff et al., 2011] Schrouff, J., Perlbarg, V., Boly, M., Marrelec, G., Boveroux, P., Vanhaudenhuyse, A., Bruno, M. A., Laureys, S., Phillips, C., Pélégrini-Issac, M., Maquet, P., and Benali, H. (2011). Brain functional integration decreases during propofol-induced loss of consciousness. *NeuroImage*, 57(1):198–205.

[Scott et al., 2014] Scott, G., Fagerholm, E. D., Mutoh, H., Leech, R., Sharp, D. J., Shew, W. L., and Knöpfel, T. (2014). Voltage imaging of waking mouse cortex reveals emergence of critical neuronal dynamics. *Journal of Neuroscience*, 34(50):16611–16620.

[Searle, 1998] Searle, J. R. (1998). How to study consciousness scientifically. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 353(1377):1935–1942.

[Searle, 2000] Searle, J. R. (2000). Consciousness. *Annual Review of Neuroscience*, 23:557–578.

[Sepúlveda et al., 2018] Sepúlveda, P. O., Carrasco, E., Tapia, L. F., Ramos, M., Cruz, F., Conget, P., Olivares, Q. F., and Cortínez, I. (2018). Evidence of hysteresis in propofol pharmacodynamics. *Anaesthesia*, 73(1):40–48.

[Sergent et al., 2017] Sergent, C., Faugeras, F., Rohaut, B., Perrin, F., Valente, M., Tallon-Baudry, C., Cohen, L., and Naccache, L. (2017). Multidimensional cognitive evaluation of patients with disorders of consciousness using EEG: A proof of concept study. *NeuroImage: Clinical*, 13:455–469.

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

[Seth, 2009] Seth, A. (2009). Explanatory correlates of consciousness: Theoretical and computational challenges. *Cognitive Computation*, 1(1):50–63.

[Seth, 2013] Seth, A. K. (2013). Interoceptive inference, emotion, and the embodied self. *Trends in Cognitive Sciences*, 17(11):565–573.

[Seth, 2018] Seth, A. K. (2018). Consciousness: The last 50 years (and the next). *Brain and Neuroscience Advances*, 2:239821281881601.

[Sharon et al., 2016] Sharon, G., Sampson, T. R., Geschwind, D. H., and Mazmanian, S. K. (2016). The Central Nervous System and the Gut Microbiome. *Cell*, 167(4):915–932.

[Shea and Frith, 2016] Shea, N. and Frith, C. D. (2016). Dual-process theories and consciousness: the case for 'Type Zero' cognition. *Neuroscience of Consciousness*, 2016(1):1–10.

[Sherrington, 1906] Sherrington, C. (1906). *The Integrative action of the nervous system*. Yale University Press, sixth edition.

[Shoemaker, 2000] Shoemaker, S. (2000). Phenomenal Character Revisited. *Philosophy and Phenomenological Research*, 60(2):465.

[Shulman et al., 2009] Shulman, R. G., Hyder, F., and Rothman, D. L. (2009). Baseline brain energy supports the state of consciousness. *PNAS*, 106(27).

[Siclari et al., 2017] Siclari, F., Baird, B., Perogamvros, L., Bernardi, G., LaRocque, J. J., Riedner, B., Boly, M., Postle, B. R., and Tononi, G. (2017). The neural correlates of dreaming. *Nature Neuroscience*, 20(6):872–878.

[Signorelli, 2018a] Signorelli, C. M. (2018a). Can Computers become Conscious and overcome Humans? *Frontiers Robotics and Artificial Intelligence*, 5(121).

[Signorelli, 2018b] Signorelli, C. M. (2018b). Can Computers Overcome Humans? Consciousness Interaction and its Implications. In *Proceedings of 2018 IEEE 17th International Conference on Cognitive Informatics and Cognitive Computing, ICCI*CC 2018*, pages 61–69. IEEE.

[Signorelli et al., 2021a] Signorelli, C. M., Diaz Boils, J., Tagliazucchi, E., Jarraya, B., and Deco, G. (2021a). From brain-body function to consciousness interaction. *(under review)*.

[Signorelli et al., 2020a] Signorelli, C. M., Dundar-Coecke, S., Wang, V., and Coecke, B. (2020a). Cognitive Structures of Space-Time. *Frontiers in Psychology*.

[Signorelli and Joaquin Diaz Boils, 2020] Signorelli, C. M. and Joaquin Diaz Boils (2020). Multilayer networks as embodied consciousness interactions. A formal model approach. *To be submited*.

[Signorelli and Meling, 2021] Signorelli, C. M. and Meling, D. (2021). Towards new concepts for a biological neuroscience of consciousness. *Cognitive Neurodynamics*.

[Signorelli et al., 2021b] Signorelli, C. M., Szczotka, J., and Prentner, R. (2021b). Explanatory profiles of models of consciousness - towards a systematic classification.

[Signorelli et al., 2021c] Signorelli, C. M., Uhrig, L., Kringelbach, M., Jarraya, B., and Deco, G. (2021c). Hierarchical disruption in the cortex of anesthetized monkeys as a new signature of consciousness loss. *NeuroImage*, 227(June 2020):117618.

[Signorelli et al., 2020b] Signorelli, C. M., Wang, Q., and Coecke, B. (2020b). Reasoning about conscious experience with axiomatic and graphical mathematics. *Submitted*.

[Signorelli et al., 2021d] Signorelli, C. M., Wang, Q., and Khan, I. (2021d). A Compositional model of Consciousness based on Consciousness-only. *Entropy*.

[Simon and Emmons, 1956] Simon, C. W. and Emmons, W. H. (1956). EEG, consciousness, and sleep. *Science*, 124(3231):1066–1069.

[Sitt et al., 2014] 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.

[Solms, 1997] Solms, M. (1997). *The Neuropsychology of Dreams: A Clinico-anatomical Study*. Lawrence Erlbaum Associates, New Jersey.

[Stamatakis et al., 2010] Stamatakis, E. A., Adapa, R. M., Absalom, A. R., and Menon, D. K. (2010). Changes in resting neural connectivity during propofol sedation. *PLoS ONE*, 5(12).

[Stankovski et al., 2017] Stankovski, T., Pereira, T., McClintock, P. V., and Stefanovska, A. (2017). Coupling functions: Universal insights into dynamical interaction mechanisms. *Reviews of Modern Physics*, 89(4):1–50.

[Stankovski et al., 2016] Stankovski, T., Petkoski, S., Raeder, J., Smith, A. F., McClintock, P. V., and Stefanovska, A. (2016). Alterations in the coupling functions between cortical and cardio-respiratory oscillations due to anaesthesia with propofol and sevoflurane. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2067).

[Stender et al., 2016] Stender, J., Mortensen, K. N., Thibaut, A., Darkner, S., Laureys, S., Gjedde, A., and Kupers, R. (2016). The Minimal Energetic Requirement of Sustained Awareness after Brain Injury. *Current Biology*, 26(11):1494–1499.

[Steriade, 2000] Steriade, M. (2000). Corticothalamic resonance, states of vigilance and mentation. *Neuroscience*, 101(2):243–276.

[Steriade et al., 1993] Steriade, M., Nunez, A., and Amzica, F. (1993). A novel slow (< 1 Hz) oscillation of neocortical neurons in vivo: Depolarizing and hyperpolarizing components. *Journal of Neuroscience*, 13(8):3252–3265.

[Steriade and Timofeev, 2003] Steriade, M. and Timofeev, I. (2003). Neuronal plasticity in thalamocortical networks during sleep and waking oscillations. *Neuron*, 37(4):563–576.

[Steriade et al., 2001] Steriade, M., Timofeev, I., and Grenier, F. (2001). Natural Waking and Sleep States: A View From Inside Neocortical Neurons. *Journal of Neurophysiology*, 85(5):1969–1985.

[Stern, 2017] Stern, P. (2017). Neuroscience: In search of new concepts. *Science*, 358(6362):464–465.

[Steyn-Ross et al., 2004] Steyn-Ross, M. L., Steyn-Ross, D. A., and Sleigh, J. W. (2004). Modelling general anaesthesia as a first-order phase transition in the cortex. *Progress in Biophysics and Molecular Biology*, 85(2-3):369–385.

[Steyn-Ross et al., 2001] Steyn-Ross, M. L., Steyn-Ross, D. A., Sleigh, J. W., and Wilcocks, L. C. (2001). Toward a theory of the general-anesthetic-induced phase transition of the cerebral cortex. I. A thermodynamics analogy. *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, 64(1):16.

[Stickgold et al., 2001] Stickgold, R., Malia, A., Fosse, R., and Hobson, J. A. (2001). Brain-mind states: I. Longitudinal field study of sleep/wake factors influencing mentation report length. *Sleep*, 24(2):171–179.

[Storm et al., 2017] Storm, J. F., Boly, M., Casali, A. G., Massimini, M., Olcese, U., Pennartz, C. M., and Wilke, M. (2017). Consciousness Regained: Disentangling Mechanisms, Brain Systems, and Behavioral Responses. *The Journal of Neuroscience*, 37(45):10882–10893.

[Strawson, 2006] Strawson, G. (2006). Realistic monism: Why physicalism entails panpsychism. *Journal of Consciousness Studies*, 13(10).

[Strevens, 2004] Strevens, M. (2004). The causal and unification approaches to explanation unified - Causally. *Nous*, 38(1):154–176.

[Suzuki and Larkum, 2020] Suzuki, M. and Larkum, M. E. (2020). General Anesthesia Decouples Cortical Pyramidal Neurons. *Cell*, 180(4):666–676.e13.

[Szymusiak, 1995] Szymusiak, R. (1995). Magnocellular nuclei of the basal forebrain: Substrates of sleep and arousal regulation. *Sleep*, 18(6):478–500.

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

[Tagliazucchi et al., 2012] Tagliazucchi, E., Balenzuela, P., Fraiman, D., and Chialvo, D. R. (2012). Criticality in large-scale brain fmri dynamics unveiled by a novel point process analysis. *Frontiers in Physiology*, 3 FEB(February):1–12.

[Tagliazucchi et al., 2014] Tagliazucchi, E., Carhart-Harris, R., Leech, R., Nutt, D., and Chialvo, D. R. (2014). Enhanced repertoire of brain dynamical states during the psychedelic experience. *Human Brain Mapping*, 35(11):5442–5456.

[Tagliazucchi et al., 2016a] 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).

[Tagliazucchi et al., 2016b] Tagliazucchi, E., Crossley, N., Bullmore, E. T., and Laufs, H. (2016b). Deep sleep divides the cortex into opposite modes of anatomical-functional coupling. *Brain Structure and Function*, 221(8):4221–4234.

[Tagliazucchi et al., 2016c] Tagliazucchi, E., Roseman, L., Kaelen, M., Orban, C., Muthukumaraswamy, S. D., Murphy, K., Laufs, H., Leech, R., McGonigle, J., Crossley, N., Bullmore, E., Williams, T., Bolstridge, M., Feilding, A., Nutt, D. J., and Carhart-Harris, R. (2016c). Increased Global Functional Connectivity Correlates with LSD-Induced Ego Dissolution. *Current Biology*, 26(8):1043–1050.

[Tagliazucchi et al., 2013] Tagliazucchi, E., von Wegner, F., Morzelewski, A., Brodbeck, V., Jahnke, K., and Laufs, H. (2013). Breakdown of long-range temporal dependence in default mode and attention networks during deep sleep. *Proceedings of the National Academy of Sciences*, 110(38):15419–15424.

[Tallon-Baudry et al., 2018] Tallon-Baudry, C., Campana, F., Park, H. D., and Babo-Rebelo, M. (2018). The neural monitoring of visceral inputs, rather than attention, accounts for first-person perspective in conscious vision. *Cortex*, 102:139–149.

[Tanabe et al., 2020] Tanabe, S., Huang, Z., Zhang, J., Chen, Y., Fogel, S., Doyon, J., Wu, J., Xu, J., Zhang, J., Qin, P., Wu, X., Mao, Y., Mashour, G. A., Hudetz, A. G., and Northoff, G. (2020). Altered Global Brain Signal during Physiologic, Pharmacologic, and Pathologic States of Unconsciousness in Humans and Rats. *Anesthesiology*, 132(6):1392–1406.

[Tarnal et al., 2016] Tarnal, V., Vlisides, P. E., and Mashour, G. A. (2016). The Neurobiology of Anesthetic Emergence. *J Neurosurg Anesthesiol*, 118(24):6072–6078.

[Tasserie, 2020] Tasserie, J. (2020). *Functional neuro-imaging study of Deep Brain Stimulation mechanisms for the restoration of consciousness using a Non-Human Primate model*. PhD thesis.

[Tasserie et al., 2019] Tasserie, J., Grigis, A., Uhrig, L., Dupont, M., Amadon, A., and Jarraya, B. (2019). Pypreclin: An automatic pipeline for macaque functional MRI preprocessing. *NeuroImage*, 207(September 2019).

[Taylor et al., 2013] Taylor, V. A., Daneault, V., Grant, J., Scavone, G., Breton, E., Roffe-vidal, S., Courtemanche, J., Lavarenne, A. S., Marrelec, G., Benali, H., and Beauregard, M. (2013). Impact of meditation training on the default mode network during a restful state. *Social Cognitive and Affective Neuroscience*, 8(1):4–14.

[Thomas Nagel, 1974] Thomas Nagel (1974). What is it like to be a bat? *The Philosophical Review*, 83(4):435–450.

[Thompson, 2004] Thompson, E. (2004). Life and mind: From autopoiesis to neurophenomenology. A tribute to Francisco Varela. *Phenomenology and the Cognitive Sciences*, 3(4):381–398.

[Thompson, 2007] Thompson, E. (2007). *Mind in Life*. Harvard University Press.

[Thompson, 2014] Thompson, E. (2014). *Waking, Dreaming, Being*. Columbia University Press.

[Thompson and Varela, 2001] Thompson, E. and Varela, F. (2001). Radical embodiment: neural dynamics and consciousness. *Trends in cognitive sciences*, 5(October):418–425.

[Tononi et al., 2016] Tononi, G., Boly, M., Massimini, M., and Koch, C. (2016). Integrated information theory: from consciousness to its physical substrate. *Nature reviews. Neuroscience*, 17(7):450–61.

[Tononi and Edelman, 1998] Tononi, G. and Edelman, G. M. (1998). Consciousness and complexity. *Science*, 282(5395):1846–1851.

[Tononi and Koch, 2015] Tononi, G. and Koch, C. (2015). Consciousness: here, there and everywhere? *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 370(1668).

[Tsuchiya et al., 2016] Tsuchiya, N., Taguchi, S., and Saigo, H. (2016). Using category theory to assess the relationship between consciousness and integrated information theory. *Neuroscience Research*, 107:1–7.

[Tsuchiya et al., 2015] 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.

[Tull and Kleiner, 2020] Tull, S. and Kleiner, J. (2020). Integrated Information in Process Theories.

[Uhrig et al., 2014] Uhrig, L., Dehaene, S., and Jarraya, B. (2014). A Hierarchy of Responses to Auditory Regularities in the Macaque Brain. *Journal of Neuroscience*, 34(4):1127–1132.

[Uhrig et al., 2016] Uhrig, L., Janssen, D., Dehaene, S., and Jarraya, B. (2016). Cerebral responses to local and global auditory novelty under general anesthesia. *NeuroImage*, 141:326–340.

[Uhrig et al., 2018] Uhrig, L., Sitt, J. D., Jacob, A., Tasserie, J., Barttfeld, P., Dupont, M., Dehaene, S., and Jarraya, B. (2018). Resting-state Dynamics as a Cortical Signature of Anesthesia in Monkeys. *Anesthesiology*, 129(5):942–958.

[van den Heuvel and Sporns, 2011] van den Heuvel, M. P. and Sporns, O. (2011). Rich-club organization of the human connectome. *Journal of Neuroscience*, 31(44):15775–15786.

[van den Heuvel and Sporns, 2013] van den Heuvel, M. P. and Sporns, O. (2013). An anatomical substrate for integration among functional networks in human cortex. *Journal of Neuroscience*, 33(36):14489–14500.

[Van Gulick, 2018] Van Gulick, R. (2018). Consciousness.

[Van Vugt et al., 2018] Van Vugt, B., Dagnino, B., Vartak, D., Safaai, H., Panzeri, S., Dehaene, S., and Roelfsema, P. R. (2018). The threshold for conscious report: Signal loss and response bias in visual and frontal cortex. *Science*, 360(6388):537–542.

[Vanduffel et al., 2001] Vanduffel, W., Fize, D., Mandeville, J. B., Nelissen, K., Van Hecke, P., Rosen, B. R., Tootell, R. B., and Orban, G. A. (2001). Visual motion processing investigated using contrast agent-enhanced fMRI in awake behaving monkeys. *Neuron*, 32(4):565–577.

[Vanhaudenhuyse et al., 2010] Vanhaudenhuyse, A., Noirhomme, Q., Tshibanda, L. J., Bruno, M. A., Boveroux, P., Schnakers, C., Soddu, A., Perlbarg, V., Ledoux, D., Brichant, J. F., Moonen, G., Maquet, P., Greicius, M. D., Laureys, S., and Boly, M. (2010). Default network connectivity reflects the level of consciousness in non-communicative brain-damaged patients. *Brain*, 133(1):161–171.

[VanRullen, 2016] VanRullen, R. (2016). Perceptual Cycles. *Trends in Cognitive Sciences*, 20(10):723–735.

[VanRullen and Koch, 2003] VanRullen, R. and Koch, C. (2003). Is perception discrete or continuous? *Trends in Cognitive Sciences*, 7(5):207–213.

[Varela, 1979] Varela, F. (1979). *Principles of Biological Autonomy*. Elsevier North Holland, New York.

[Varela et al., 2001] Varela, F., Lachaux, J., Rodriguez, E., and Martinerie, J. (2001). The brainweb: phase synchronization and large-scale integration. *Nature reviews Neuroscience*, 2(April):229–239.

[Varela et al., 1974] Varela, F., Maturana, H., and Uribe, R. (1974). Autopoiesis: The organization of living systems, its characterization and a model. *BioSystems*, 5:187–196.

[Varela, 1975] Varela, F. J. (1975). A Calculus for Self-Reference.

[Varela, 1991] Varela, F. J. (1991). Organism: A Meshwork of Selfless Selves. In Tauber A.I., editor, *Organism and the Origins of Self*, pages 79–107. Springer, Dordrecht.

[Varela, 1996] Varela, F. J. (1996). Neurophenomenology: A Methodological Remedy for the Hard Problem. *Journal of Consciousness Studies*, 3(4):330–349.

[Varela, 1997] Varela, F. J. (1997). Patterns of life intertwining identity and cognition. *Brain and Cognition*, 87(34):72–87.

[Varela et al., 2016] Varela, F. J., Thompson, E., and Rosch, E. (2016). *The Embodied Mind*. MIT Press, revised ed edition.

[Velazquez, 2020] Velazquez, J. L. P. (2020). On the emergence of cognition: from catalytic closure to neuroglial closure. *Journal of Biological Physics*.

[Vidaurre et al., 2016] Vidaurre, D., Quinn, A. J., Baker, A. P., Dupret, D., Tejero-Cantero, A., and Woolrich, M. W. (2016). Spectrally resolved fast transient brain states in electrophysiological data. *NeuroImage*, 126:81–95.

[Vincent et al., 2007] Vincent, J. L., Patel, G. H., Fox, M. D., Snyder, A. Z., Baker, J. T., Van Essen, D. C., Zempel, J. M., Snyder, L. H., Corbetta, M., and Raichle, M. E. (2007). Intrinsic functional architecture in the anaesthetized monkey brain. *Nature*, 447(7140):83–86.

[Vogt and Derbyshire, 2009] Vogt, B. A. and Derbyshire, S. W. G. (2009). Visceral circuits and cingulate-mediated autonomic functions. In *Cingulate Neurobiology and Disease*, chapter 10, pages 219–236. Oxford University Press, Oxford.

[von Neumann, 1958] von Neumann, J. (1958). *The Computer and the Brain*. Yale University Press.

[Voss et al., 2009] Voss, U., Holzmann, R., Tuin, I., and Hobson, J. A. (2009). Lucid dreaming: a state of consciousness with features of both waking and non-lucid dreaming. *Sleep*, 32(9):1191–1200.

[Wang et al., 2014] Wang, Z., Solloway, T., Shiffrin, R. M., and Busemeyer, J. R. (2014). Context effects produced by question orders reveal quantum nature of human judgments. *Proceedings of the National Academy of Sciences of the United States of America*, 111(26):9431–6.

[Ward and Wegner, 2013] Ward, A. F. and Wegner, D. M. (2013). Mind-blanking: When the mind goes away. *Frontiers in Psychology*, 4(SEP):1–15.

[Warnaby et al., 2017] Warnaby, C. E., Sleigh, J. W., Hight, D., Jbabdi, S., and Tracey, I. (2017). Investigation of Slow-wave Activity Saturation during Surgical Anesthesia Reveals a Signature of Neural Inertia in Humans. *Anesthesiology*, 127(4):645–657.

[Werner, 2012] Werner, G. (2012). From brain states to mental phenomena via phase space transitions and renormalization group transformation: Proposal of a theory. *Cognitive Neurodynamics*, 6(2):199–202.

[Werner, 2013] Werner, G. (2013). Consciousness viewed in the framework of brain phase space dynamics, criticality, and the Renormalization Group. *Chaos, Solitons and Fractals*, 55:3–12.

[White, 2018] White, P. A. (2018). Is conscious perception a series of discrete temporal frames? *Consciousness and Cognition*, 60(February):98–126.

[Wiener, 1985] Wiener, N. (1985). *Cybernetics: Or Control and Communication in the Animal and the Machine*. The MIT Press, second edition.

[Wiese and Friston, 2020] Wiese, W. and Friston, K. J. (2020). The neural correlates of consciousness under the free energy principle : From computational correlates to computational explanation. *PsyArXiv*, pages 1–27.

[Wittmann, 2011] Wittmann, M. (2011). Moments in time. *Frontiers in Integrative Neuroscience*, 5(October):1–9.

[Woodward, 2004] Woodward, J. (2004). *Making Things Happen*. Oxford University Press.

[Woodward, 2013] Woodward, J. (2013). II-James Woodward: Mechanistic Explanation: Its Scope and Limits. *Aristotelian Society Supplementary Volume*, 87(1):39–65.

[Woodward, 2019] Woodward, J. (2019). Scientific Explanation.

[Wu et al., 1996] Wu, J., Harata, N., and Akaike, N. (1996). Potentiation by sevoflurane of the γ-aminobutyric acid-induced chloride current in acutely dissociated CA1 pyramidal neurones from rat hippocampus. *British Journal of Pharmacology*, 119(5):1013–1021.

[Yoshimi, 2007] Yoshimi, J. (2007). Mathematizing phenomenology. *Phenomenology and the Cognitive Sciences*, 6(3):271–291.

[Zeeman, 1976] Zeeman, E. C. (1976). Catastrophe theory. *Scientific American*, 234(4):65–83.

[Zong et al., 2017] Zong, W., Wu, R., Li, M., Hu, Y., Li, Y., Li, J., Rong, H., Wu, H., Xu, Y., Lu, Y., Jia, H., Fan, M., Zhou, Z., Zhang, Y., Wang, A., Chen, L., and Cheng, H. (2017). Fast high-resolution miniature two-photon microscopy for brain imaging in freely behaving mice. *Nature Methods*, 14(7):713–719.

[Zonta et al., 2003] Zonta, M., Angulo, M. C., Gobbo, S., Rosengarten, B., Hossmann, K. A., Pozzan, T., and Carmignoto, G. (2003). Neuron-to-astrocyte signaling is central to the dynamic control of brain microcirculation. *Nature Neuroscience*, 6(1):43–50.