UPPSALA
UNIVERSITET

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine 1597

Brain, consciousness and disorders of consciousness at the intersection of neuroscience and philosophy

MICHELE FARISCO

ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2019

ISSN 1651-6206
ISBN 978-91-513-0749-7
urn:nbn:se:uu:diva-392187

Dissertation presented at Uppsala University to be publicly examined in Sal IX, Universitetshuset, Biskopsgatan 3, Uppsala, Wednesday, 30 October 2019 at 09:00 for the degree of Doctor of Philosophy (Faculty of Medicine). The examination will be conducted in English. Faculty examiner: Professor Wolf Singer (Max Planck Institute for Brain Research, Frankfurt am Main, Germany).

Abstract
Farisco, M. 2019. Brain, consciousness and disorders of consciousness at the intersection of neuroscience and philosophy. *Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine* 1597. 63 pp. Uppsala: Acta Universitatis Upsaliensis. ISBN 978-91-513-0749-7.

The present dissertation starts from the general claim that neuroscience is not neutral, with regard to theoretical questions like the nature of consciousness, but it needs to be complemented with dedicated conceptual analysis. Specifically, the argument for this thesis is that the combination of empirical and conceptual work is a necessary step for assessing the significant questions raised by the most recent study of the brain. Results emerging from neuroscience are conceptually very relevant in themselves but, notwithstanding its theoretical sophistication, neuroscience is not sufficient to provide a complete interpretation or an appropriate understanding of their impact. Consequently, the present thesis starts from the need for an interdisciplinary and hybrid field of research, i.e. fundamental neuroethics.

Within this framework, the thesis takes consciousness and related disorders (i.e. Vegetative State/Unresponsive Wakefulness Syndrome, Minimally Conscious State and Coma) and the addicted brain as illustrative cases of the potential fruitful collaboration between empirical and conceptual investigations.

The general goal of the thesis is to contribute to the overall development of bridging the gap between empirical and conceptual understandings of consciousness. The first paper sets the theoretical framework, providing an empirically-based description of the brain with significant philosophical implications for an understanding of consciousness. The last three papers of the thesis try to apply the theoretical framework to illustrative cases. Papers II and III analyse the possible application of science and technology for an easier detection and clinical care of patients with disorders of consciousness, with particular attention to communication mediated by neurotechnology and the simulation of the conscious brain, respectively; paper IV provides a potentially new ethical analysis of addiction within the elaborated general conceptual framework.

The conclusion of the thesis is that the impact of neuroscientific results needs that a dedicated conceptual approach reveals and investigates their conceptual meaning. This conceptual analysis is not exclusive but integrative and complementary to the empirical science. The case of consciousness, analysed from both an ethical and conceptual point of view, is highly illustrative in this respect. In the end, a conceptual/linguistic work of clarification is urgently needed.

Keywords: Brain; consciousness; disorders of consciousness; neuroethics; neurophilosophy

Michele Farisco, Centre for Research Ethics and Bioethics, Box 564, Uppsala University, SE-751 22 Uppsala, Sweden.

ISSN 1651-6206
ISBN 978-91-513-0749-7
urn:nbn:se:uu:diva-392187 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-392187)

To my beloved family: my wife Olimpia
and my children Ilde and Andrea.
You shared me with this inspiring
but often tough journey.
I will never thank you enough.

Consciousness, however small, is an illegitimate birth in any philosophy that starts without it, and yet professes to explain all facts by continuous evolution.

William James (1890), *The Principles of Psychology*, Vol. I, Ch. VI, p. 149

List of Papers

This thesis is based on the following papers[1], which are referred to in the text by their Roman numerals.

I Farisco, M., Laureys, S., Evers, K. (2017) The Intrinsic Activity of the Brain and Its Relation to Levels and Disorders of Consciousness. *Mind&Matter*, 15(2):197-219

II Farisco, M., Laureys, S., Evers, K. (2015) Externalization of consciousness. Clinical and ethical issues. *Current Topics in Behavioral Neuroscience*, 19:205-222

III Farisco, M., Hellgren Kotaleski, J., Evers, K. (2018) Large-scale brain simulation and disorders of consciousness. Mapping technical and conceptual issues. *Frontiers in Psychology*, 9:585

IV Farisco, M., Evers, K., Changeux, J.P. (2018) Drug Addiction: from Neuroscience to Ethics. *Frontiers in Psychiatry*, 9:595

Reprints were made with permission from the respective publishers.

[1] Other relevant publications by Michele Farisco, which are part of his research but not included in the thesis are: Farisco, M., & Evers, K. (2017). The ethical relevance of the unconscious. *Philos Ethics Humanit Med, 12*(1), 11. doi:10.1186/s13010-017-0053-9; Farisco, M., Salles, A., & Evers, K. (2018). Neuroethics: A Conceptual Approach. *Camb Q Healthc Ethics, 27*(4), 717-727. doi:10.1017/S0963180118000208.

Contents

Abbreviations

AI	Artificial Intelligence
DMN	Default Mode Network
DOCs	Disorders of Consciousness
fMRI	Functional Magnetic Resonance Imaging
GNWT	Global Neuronal Workspace Theory
ICT	Intrinsic Consciousness Theory
IIT	Integrated Information Theory
MCS	Minimally Conscious State
NCC	Neural Correlates of Consciousness
UWS	Unresponsive Wakefulness Syndrome
VS	Vegetative State

Introduction

The instrumental investigation and assessment of consciousness and its disorders (DOCs, i.e. coma, vegetative state/unresponsive wakefulness syndrome, VS/UWS and minimally conscious state, MCS) have witnessed remarkable progress over the last few years. Among other things, this progress has resulted in the passage from a monolithic way of looking at severe brain damage to a more graded nosology, based on a quantitative assessment of consciousness and on functional neuroimaging technologies.

The so-called 'neuro-technologies', especially the application of technology to the assessment and investigation of consciousness, has led to impressive and unpredictable results with important theoretical and practical implications.

While the technical advances in the study of consciousness, particularly of its correlates, have been remarkable (even if several important issues remain unanswered), the conceptual investigation of consciousness still seems slowed down by controversies about how much explanatory power should be attributed to science, notably empirical psychology and cognitive neuroscience, and how to handle the empirical evidence emerging therefrom.

We ultimately still lack a comprehensive conceptual assessment of consciousness. This gap in the clarification of consciousness risks to affect the scientific investigation itself: the meaningfulness of science is grounded on background concepts that need to be made explicit, elaborated and analysed from a scientific as well as from an extra-scientific perspective. Specifically, neuroscience knowledge is epistemically normative, in the sense that it is based on presupposed models that make neuroscientific results scientifically sound even when still limited. This is particularly relevant to the study of subjective experience, which as such requires both a third- and first-person perspective.

A conceptual investigation of consciousness trying to integrate both empirical and theoretical perspectives is urgently needed. This thesis aims to propose a conceptual assessment of consciousness through a multidisciplinary approach.

By conceptual clarification, I mean two things in particular: 1) a philosophical analysis of theoretical and epistemological premises and categories of science and 2) the application of a naturalistically oriented philosophical reasoning, assessing the impact of scientific inductive and deductive explanations and justifications and their logical consistency. As will be further explained below, I assume conceptual investigation to be part of a naturalistic philosophy, i.e. in strong connection with empirical science.

In this sense, I take a specific position in the *querelle* about the theoretical relevance of neuroscience for conceptualising consciousness. As will be argued in detail below, I defend at the same time the relevance of neuroscience for conceptualising consciousness and the need for complementing neuroscientific investigation with explicitly philosophical reasoning. This might seem like a kind of compromise between scientific and philosophical reasons, but I prefer to describe it as the parallel recognition of the intrinsic insufficiency and of the necessary complementarity of both science and philosophy in order to handle the fundamental issue of the nature of consciousness. It is true that neuroscience is primarily interested in and focused on the *conscious brain*, avoiding reference to abstract and potentially misleading notions like consciousness as such, but I think that neuroscientific knowledge is highly relevant for assessing the philosophical issue of the nature of subjective experience (e.g. Where does it come from? Why is the brain conscious? Can other objects be conscious? among others). These are philosophical, specifically metaphysical issues, and neuroscience is not neutral to them.

There is a further point on neuroscience methods that is important to outline. Notwithstanding important developments in theoretical neuroscience, to date the neuroscience of consciousness is mainly grounded on an empirical, inductive methodology. As a result, the theoretical component of the neuroscientific analysis of consciousness is still neither very mature nor adequately developed. Maybe the Integrated Information Theory (IIT) represents one exception. Nevertheless, I think that its postulates are ultimately not sufficiently justified, i.e. not sufficiently grounded on available empirical evidence and exposed to criticism about their logical consistency (Bayne, 2018; Tononi, 2008). Compared to other disciplines like, for instance, physics, neuroscience still seems to lack a solid theoretical framework necessary for giving an inferential twist to its methodology.

For these reasons, this thesis starts from the most recent advancements in the scientific description of the brain to suggest a new conceptual model of consciousness, named the Intrinsic Consciousness Theory (ICT). ICT is an attempt to overcome the intrinsic limitation of our everyday language, which affects our explanatory power regarding consciousness. Instead, ICT suggests a semantic stretching of consciousness, on the basis of an inferential reasoning applied to empirical knowledge. Such 'conceptual experiment' has significant implications for thinking its ethical relevance in the context of DOCs and addiction.

Furthermore, this new philosophical theory of consciousness has potential implications beyond the theoretical context, at the ethical, clinical and social levels. In this thesis, I provide a description of particular cases from these contexts, specifically the care of people with DOCs, the neurotechnological assessment of consciousness through computer models and simulations, and the ethics of addiction, to conclude by outlining further possible directions that can be explored in future work.

Rationale

The impact of the growing scientific knowledge on clinical practice, particularly regarding DOCs, has been remarkable, and it will realistically continue to grow. The vast scientific knowledge emerging from the investigation of the brain is driving the development of new technologies that are changing the way of looking at and treating DOCs.

From an ethical point of view, the impact of these new knowledge and technology on clinical practice is significant; clinicians and families have to assess emerging challenges in order to decide the right treatment for their loved ones affected by DOCs.

A sound ethical assessment of these issues is impossible without a conceptual clarification of consciousness. Such clarification is necessary to epistemically assess neuroscience as well as its extra-scientific impact; it is important to analyse the categories used by science, its methodology, the potential biases affecting its methods and consequently its results as well as the categories' possible theoretical impact, and the meaning of what science is studying (e.g. consciousness). In brief, while neuroscience is moving towards an increasing sophistication in the understanding of consciousness and related disorders, it still needs a conceptual assessment in order to avoid epistemic traps and to go from the bench to the clinics.

Moreover, given the enduring lack of an overarching theory of consciousness, I think it is timely to attempt to proceed in this direction, on the basis of a dialogue between neuroscience and philosophy.

Aim

The goal of the thesis is to contribute to the attempt to develop a general conceptual model of consciousness.

The specific aims of the thesis are:

1. to develop a conceptual analysis of the most recent achievements of the neuroscience of consciousness and related technologies
2. to use this conceptual analysis as the ground for developing an ethical analysis of consciousness
3. to apply the resulting ethical framework to different contexts, i.e. the clinical treatment of DOCs and the social issue of addiction.

Specific aims

Paper I

To develop a new, broad definition of consciousness, starting from the empirical understanding of the brain as intrinsically active and plastic (Farisco, Laureys, & Evers, 2017). Consciousness is equated with the intrinsic projective intentionality of the brain, i.e. with the basic, intrinsic predisposition of the brain to develop models of the world.

The paper then explores the possible impacts of this notion of consciousness on our understanding of its disorders, and its potential role in the diagnosis and care of patients with DOCs.

Paper II

The paper starts by analysing recent advances in neurotechnological assessment of residual consciousness in patients with DOCs and in neurotechnology-mediated communication with them (Farisco, Laureys, & Evers, 2015).

Specifically, the paper discusses some technical aspects of functional magnetic resonance imaging (fMRI) and brain-computer interfaces (BCI) and their prospective use for communicating with patients with DOCs. The paper aims at setting the scientific stage, i.e. the potential and actual clinical application of neuro-imaging for diagnosing and assessing DOCs, through two specific tasks:

- describing the theoretical and technical premises of "mind-reading" and "externalisation of mind"
- analysing the ethical issues emerging from the clinical application of neuroimaging technologies.

Paper III

This paper aims to assess the plausibility of simulation technologies for emulation of consciousness and the potential clinical impact of large-scale brain simulation on the assessment and care of DOCs (Farisco, Kotaleski, & Evers, 2018). Notwithstanding their technical limitations, the paper suggests that simulation technologies may offer new solutions to old practical problems, particularly in clinical contexts. Specifically, the paper argues that the simulation of neural correlates of consciousness (NCC) is potentially useful for improving treatments of patients with DOCs.

Paper IV

This paper aims to elaborate a new ethical analysis of addiction, focusing on the relationship between aware and unaware processing in the brain (Farisco, Evers, & Changeux, 2018). It takes the case of the opioids epidemics to argue that a consideration of both aware and unaware processing provides a more comprehensive ethical framework to discuss the ethical issues raised by addiction. Finally, the hypothesis is that in addition to identified Central Nervous System's neuronal/neurochemical factors contributing to addictive dynamics, socio-economic status plays a causal role through epigenetic processes, originating the need for additional reward in the brain. This provides a strong base for a socio-political form of responsibility for preventing and managing addiction crisis.

Methods

Within the framework of a broadly defined natural philosophy (i.e. an analytical conceptual approach informed by the empirical scientific data), the thesis develops a conceptual analysis of the most recent empirical findings emerging from the neuroscientific investigation of consciousness. Particularly, the first paper attempts to elaborate a conceptual model of consciousness on the basis of recent empirical evidence and related interpretation; the last three papers join conceptual and ethical analysis, with the aim of applying the models elaborated in the first study to illustrative contexts (i.e. DOCs and drug addiction).

In general, the present work is developed within the methodological framework of fundamental neuroethics, originally introduced by Kathinka Evers (Evers, 2007, 2009) and recently developed by the Uppsala University Neuroethics group (Evers, Salles, & Farisco, 2017; Farisco, Salles, & Evers, 2018).

Briefly, three main approaches in neuroethics are distinguished: neurobioethics, empirical neuroethics and conceptual neuroethics (Evers et al., 2017). Neurobioethics is conceived as mainly normative, i.e. the application of ethical theory to practical issues arising from neuroscientific research and its clinical applications, as well as to issues arising from the public communication of neuroscience research. Empirical neuroethics is assumed as being mainly descriptive and occasionally explanatory; it uses data to assess theoretical (e.g. the definition of moral reasoning) and practical issues (e.g. the definition of a moral agent). Conceptual neuroethics, including fundamental neuroethics as a particular form, is primarily theoretical in the sense that it uses conceptual analysis of key notions to assess, among other things, why and how empirical knowledge of the brain can be relevant to philosophical, social and ethical concerns.

This thesis is an attempt to apply the conceptual neuroethics model and related methodology to the specific issue of consciousness. It starts from the need for a philosophical assessment of consciousness because of the limitation of the neuroscientific approach. This is not to deny that neuroscience is conceptually relevant, or that it has a conceptual component itself. However, I think that neuroscience is conceptually intrinsically limited for the following reasons (Farisco, Salles, et al., 2018): 1. Neuroscience is a relatively new field of research, especially if compared to other fields like physics, so that its conceptual component is much less developed; 2. The possibility and the need to approach consciousness from both third- and first-person perspectives implies the epistemic insufficiency of neuroscience, which is confined to the third-person perspective; 3. Neuroscience is necessarily linked to models, which epistemically mediate between the world and us, impacting and eventually determining what we can know about it. The definition of these models depends on both scientific and extra-scientific factors, and a dedicated philosophical analysis might be a necessary complementation for neuroscience to build the most reliable conceptual models and 4. Because of the multiscale and multilevel structure of the brain, organised in different spatiotemporal scales, from molecules to cells to multicellular assemblies to long-distance networks to behaviour (Changeux, 2017), a conceptual work of refinement, interpretation and synthesis is necessary, and neuroscience does not suffice to clear all the needed concepts (e.g. space, time, level).

Fundamental neuroethics is a particular form of conceptual neuroethics (Evers, 2007). It aims to be not simply an analysis of the potential impact of neuroscience on fundamental notions like self, responsibility and freedom, like theorised in neuroethics from the beginning (Roskies, 2002), but rather an analysis of fundamental concepts and methods used in the neuroscientific investigation of notions like identity, morality and consciousness, among others. Methodologically, fundamental neuroethics is multidisciplinary and interdisciplinary. It is multidisciplinary because it uses elements from different disciplines, including philosophy of science, philosophy of language, philosophy of mind and moral philosophy. It is interdisciplinary because it combines both empirical and conceptual disciplines, so that fundamental neuroethics cannot be subsumed under any particular classical discipline. In the end, fundamental neuroethics recognises the mutual relevance of philosophy and neuroscience; the first needs to take relevant empirical data and their interpretation into account when addressing issues such as consciousness, while the second needs the conceptual complementation of philosophy in explaining its results.

There is a specific aspect of the conceptual insufficiency of science that I think is important to stress in the context of the following discussion. When the question of the nature of consciousness is the issue at stake, neuroscience, with its empirical methodology, is of course relevant but insufficient. As Uriah Kriegel writes, addressing 'over-and-above' claims (i.e. claims that conscious experience is nothing-over-and-above its neuronal underpinnings) is not a scientific but rather a philosophical task (Kriegel, Forthcoming). When the question of 'What is consciousness' arises, neuroscientific findings (e.g. NCC) seem open to a number of different interpretations. A focused conceptual analysis might provide support for a particular view. For this reason, philosophical reflection on consciousness is not an optional complement to the neuroscience of consciousness, but rather necessarily integral to it.

Background

The problem of consciousness

What is consciousness? This is a puzzling question, similar to what Saint Augustine first and David Hume later wrote about time: I know what it is, but if you ask me what it is, I don´t know it. The elusiveness of the concept of consciousness is not a limitation, but a challenge that fuels both scientific and philosophical investigation.

The empirical investigation of consciousness still lacks a comprehensive conceptual theory of its object. The clinical assessment of conscious brain activity and its detection, as well as the diagnosis and the prognosis of DOCs, have been greatly improved during the last few decades, especially thanks to extraordinary scientific and technological advances in fields like brain imaging and computational modelling, among others. This means that science, as well as clinical practice, cannot wait for a comprehensive definition of consciousness; to know particular and limited aspects of what can be generally defined as conscious experience is sufficient for developing technology and making medical decisions that can be qualified as appropriate, at least provisionally. Yet, big uncertainties remain, a lot still has to be cleared, and this clarification needs an improvement of the conceptual assessment of consciousness.

Historical and conceptual controversies about consciousness

The reflection on consciousness can be traced back to the origins of humanity. Both pre-historical and pre-literate societies left signs of attention and reflection about the nature of consciousness (Lewis-Williams, 2002). Nevertheless, it is reasonable to think that people of that period retained a different conceptualisation (if any) and a different experience of consciousness. In addition to the disagreement regarding how we understand consciousness, the question of the historical origin in Europe of the concept of consciousness we refer to today is also still open and debated (Lewis-Williams, 2002). Despite the lack of agreement regarding the origin of the concept, consciousness becomes central in philosophical definition of self from the 17^{th} century, notably with the work of René Descartes, who defines the mental as essentially related to consciousness. Descartes was followed by John Locke and Gottfried Wilhelm Leibniz. While Leibniz recognised the possibility of a dissociation between mind and consciousness, admitting the existence of unconscious thoughts, the associationist psychology developed in the Anglo-American world from the 18th century identified mind and consciousness. Associationist psychology basically affirms that consciousness is a succession of associated ideas. This position was first criticised by Kant, who argued for the need to invoke a more structured self, with particular reference to space, time and causality, and then by phenomenology, which included body and society in the study of consciousness.

Modern scientific psychology in the mid 19th century still equated mind and consciousness, while behaviourism, from the beginning of the 20th century, excluded consciousness from scientific psychology. In the second half of 20th century, cognitive psychology reintroduced mental processes, basically defined as information processing, in the debate, even though consciousness remained largely neglected for several further years.

The situation deeply changed in the 1980s and 1990s, when considerable philosophical and neuroscientific interest in consciousness study, still ongoing, started.

Although the definitions of consciousness are many and even greatly different, it is possible to make a distinction between different types according to particular criteria (Van Gulick, 2014).

Among the most influential theories, we can identify the following:

a. Higher-order theories

Reflexive meta-mental self-awareness is critical for the definition of a conscious mental state, which is mental as long as it is related to a simultaneous and non-inferential higher-order state whose content is the one actually in the mental state (Carruthers, 2000; Rosenthal, 1997). To be conscious requires to be conscious of being conscious.

b. Reflexive theories

Like higher-order theories, reflexive theories stress a strong link between consciousness and self-consciousness, but unlike the abovementioned theories, reflexive theories locate self-awareness within the conscious state itself rather than in a distinct meta-state. In short, every conscious perception is at the same time directed towards an external object and towards itself (Gennaro, 2012; Kriegel & Williford, 2006).

c. Representationalist theories

According to these theories, the representationalist features of consciousness exhaust all its mental features; conscious mental states are one and the same with representational states (Tye, 1995).

d. Narrative interpretative theories

Consciousness is the result not of determinate facts, but of a larger context of interpretative judgments (Dennett, 1991; Gazzaniga, 2011), finally emerging as a narrative process devoid of intrinsic reality.

e. Cognitive theories

Consciousness is associated with a distinct cognitive architecture or with a special pattern of cognitive activities. A prominent example of such theories is the Global Neuronal Workspace Theory (GNWT), which describes consciousness as a competition among processors and outputs for a limited capacity resource that broadcasts information, which is conscious as long as it is available to the global workspace (Baars, 1988; Dehaene & Naccache, 2001).

f. Information Integration Theory

Consciousness is essentially defined by the integration of information; this integration is necessary and sufficient for consciousness, regardless of the substrate in which it is realised (e.g. artificial or biological) (Koch, 2012). The most famous example of this theory is Tononi's IIT, according to which consciousness is an information-theoretic property of a cognitive system (Tononi, 2008). According to this account, consciousness is a graded feature.

g. Neural theories

Consciousness has neural correlates, and at least some of them are essential substrates of consciousness. Different specific explanations of consciousness are included under the umbrella term of neural theories, with the difference arising from the neural processes or properties assumed as essential to consciousness (system-level or more local and specific mechanisms) and from the particular aspect of consciousness assumed as explanandum (Metzinger, 2000a).

h. Quantum theories

The natural locus of consciousness is placed beyond the neural, at the microphysical level of quantum phenomena (Hameroff & Penrose, 2014). Different specific versions of this kind of theories have been elaborated, with the shared character being the stressed necessity to go beyond classical physics to explain consciousness.

i. Non-physical theories

Consciousness is described as a non-physicalist aspect of reality, i.e. something that cannot be reduced to the natural/physical world (Chalmers, 1996). A possible version of non-physical theories consists of asserting the fundamental character of consciousness, i.e. consciousness is a fundamental, non-reducible entity, as it is stated in different forms of panpsychism.

Controversy exists not only about the concept of consciousness, but also about the relationship between neuroscience and philosophy, with respect to issues such as the explanatory power of neuroscience (or lack thereof), the epistemic primacy of *a posteriori* over *a priori* knowledge (or vice versa), and the categorical differentiation between the respective objects of investigation, etc. (Bennett, Dennett, Hacker, & Searle, 2007).

Among the possible options regarding the relationship between neuroscience and philosophy, the so called ‘neutrality thesis’ has historically had many supporters (Whiteley, 2019). According to the neutrality thesis, science in general and neuroscience in particular are neutral in the matter of theoretical/explanatory issues, i.e. in the assessment of the nature of consciousness. A related position, the so called ‘compatibility thesis’, argues that philosophy should be compatible with empirical science, but can proceed completely *a priori* because science is not crucial in assessing theoretical issues.

The methodological premise of the present thesis (already described) goes in a different direction: neuroscience is not neutral to theoretical/explanatory issues, and yet intrinsically insufficient for analysing them. With specific reference to consciousness, I propose complementarity rather than neutrality as necessary for advancing the understanding of consciousness; neuroscience has (or should have) a significant impact on the elaboration of philosophical theories of consciousness, which, on the other hand, should not simply be compatible with empirical findings but also try to overcome the intrinsic conceptual limitation of empirical, *a posteriori,* knowledge.

Uriah Kriegel has recently summarised the different conceptual options regarding matter-consciousness relationship as follows (Kriegel, Forthcoming).

A first main distinction can be drawn between monism (i.e. matter and consciousness are unified at the fundamental level) and dualism (i.e. matter and consciousness are fundamentally different).

Monism can be physicalist (i.e. reality is at bottom physical), neutral (i.e. reality is at bottom neither physical nor mental) or idealist (i.e. reality is at bottom mental).

Physicalist monism can be eliminative (i.e. there are no experiential properties or types), reductive (i.e. experiential properties or types are identical with physical ones) and nonreductive (i.e. experiential properties and types are not identical with physical ones but constitutively dependent upon them).

Nonreductive physicalist monism is divided into a priori (i.e. experiential supervenes on the physical with conceptual necessity) and a posteriori (i.e. experiential supervenes on the physical with merely metaphysical necessity) nonreductive physicalist monism.

Dualism is divided into substance (i.e. matter and consciousness are two separate substances) and property (i.e. matter and consciousness are two separate properties of an underlying reality) dualism. Property dualism is then divided into naturalistic (i.e. consciousness is causally or nomically dependent upon the physical) and non-naturalistic (i.e. there is no causal or nomic dependence on consciousness upon the physical) property dualism. Naturalistic property dualism is divided into interactionist (i.e. consciousness has a causal efficacy on the physical) and epiphenomenalist (i.e. consciousness has no causal efficacy on the physical) naturalistic property dualism.

Taking inspiration from John Searle´s biological naturalism (Searle, 2007), I think that the relationship between consciousness and matter should be framed within the biological context, because organic life is (at least so far)[2] the only (ontological) level where we experience the existence of consciousness.

The alleged explanatory and metaphysical neutrality of the neuroscience of consciousness defended, for instance, by David Chalmers (Chalmers, 2000) is challenged by recent developments in the field. In fact, consciousness research is not limited to the correlative description of the NCC programme, but includes sophisticated theories with strong empirical grounds and undeniable explanatory ambitions (Whiteley, 2019). It is as though neuroscience itself feels the need for elaborating its explanatory models of empirical findings.

As illustrative cases, I briefly describe the two most influential scientific theories of consciousness, the Global Neuronal Worskspace Theory and the Integrated Information Theory, to show that neuroscientific accounts of consciousness aim at explaining it but necessarily start from specific preliminary theoretical assumptions, requiring further conceptual (i.e. philosophical) analysis.

The Global Neuronal Workspace Theory of Consciousness

Starting from the idea of a cognitive global workspace (GNW) originally suggested by Baars (Baars, 1988), Dehaene, Kerszberg and Changeux specifically proposed the Global Neuronal Workspace Theory of Consciousness (GNWT) (Dehaene, Kerszberg, & Changeux, 1998). GNWT identifies consciousness with conscious access of information, which corresponds to global information availability: "What we subjectively experience as conscious access is the selection, amplification and global broadcasting, to many dis-

[2] This is not do deny in principle the possibility that consciousness might exist also in other non-biological contexts, like Artificial Intelligence. The point being to strategically focus on consciousness as actually manifested to then eventually infer properties that can possibly be relevant for exploring the possibility of consciousness existence also in other contexts.

tant areas, of a single piece of information selected for its salience or relevance to current goals" (Dehaene, Changeux, & Naccache, 2011).

As explicitly written by Changeux, the aim of GNWT is not to solve the general problem of consciousness but to model the independent processing of several and different signals passing through distinct parallel pathways and their integration in a unified field or a common workspace (Changeux, 2004a). For this reason, GNWT has been deliberately focused on specific aspect of consciousness in order to offer the possibility to test the models experimentally on defined tasks.

In short, GNWT suggests that a subset of cortical pyramidal cells with long-range excitatory axons, particularly dense in prefrontal, cingulate and parietal regions, together with the relevant thalamocortical loops, form a horizontal 'neuronal workspace', interconnecting the multiple specialised, automatic and non-conscious processors (Dehaene & Changeux, 2011). The difference between conscious and non-conscious information is that while non conscious information is encapsulated within discrete processors, conscious information is globally broadcasted within the GNW. In this way, information is better processed and can be verbally reported. In the end, what we experience as a conscious state is global availability of information (Dehaene & Naccache, 2001).

Notwithstanding its elegance and parsimony in the explanation of consciousness, GNWT seems to focus only on one aspect of the complex phenomenon of consciousness. Namely, as explicitly affirmed by its proponents, GNWT focuses on conscious access or conscious processing of a stimulus (Dehaene & Changeux, 2011). This is indeed an important dimension of the phenomenon of consciousness but does not seem to cover all its conceptual and empirical aspects. Moreover, another form of consciousness that seems not to be explainable within GNWT is a kind of conscious state devoid of any content (contentless consciousness)[3] (Thompson, 2015).

The Integrated Information Theory of consciousness

IIT is a scientific theory of consciousness developed by Giulio Tononi and Gerald Edelman with a strong inferential theoretical element. In fact, its

[3] Actually, as mentioned above, this limited focus of GNWT on specific aspects of consciousness is a deliberate choice of its proponents in order to make possible experimental test of the model on specific tasks. Thus, the limitation of GNWTs focus is not in principle, but for practical reasons. In fact, the theory has been subsequently developed further to also cover other dimensions of consciousness, like self-consciousness (Lou, Changeux, & Rosenstand, 2016) and social interaction (Changeux, 2017). Regardless, in my analysis, I refer to the original formulation of the theory.

premise is that understanding consciousness requires not only empirical studies of its neural correlates but also a principled theoretical approach that can provide explanatory, inferential and predictive power (Tononi, 2008). In other words, IIT starts from the assumption that it is necessary to complement scientific explanations of consciousness with theoretical, conceptual investigations.

Particularly, according to IIT, it is not possible to infer the existence of consciousness from physical systems, which implies that the opposite approach is necessary; starting from experience, identifying its essential properties (axioms) and then inferring what kind of properties (postulates) physical systems must have to account for the essential properties of consciousness (Tononi, Boly, Massimini, & Koch, 2016).

Ultimately, consciousness is equated with maximal integrated information[4], which is defined as the amount of information generated by a complex of elements, over and above the information generated by its parts. Information is assumed as uncertainty reduction; the more perceptual possibilities are ruled out, the more information is available and the higher is the level of consciousness. Intrinsic information is defined as differences that make a difference within a system. Intrinsic information is raised by mechanisms that exist intrinsically, without the need for an external observer-interpreter. Consciousness is ultimately identical with intrinsic information; a system is conscious if it generates information over and above its constituting parts and independently from external observers-interpreters. This is the reason why, according to IIT, a digital simulation of the brain cannot be conscious, either in principle or in practice, while a neuromorphic silicon made computer could be conscious, because it could be composed in order to realise neuron-like macro-elements intrinsically existing and characterised by conceptual structures (i.e. cause-effect repertoires) similar to ours (Tononi, 2015).

Hence, for IIT, there is consciousness when there is a difference within a physical system. Specifically, the subset of elements causally connected in a re-entrant architecture with maximal causal power is conscious. The brain architecture is an excellent example of such an organisation, but IIT does not limit consciousness to human brains (Fallon, 2016).

Conceptually, proponents of IIT state that it offers a parsimonious explanation for empirical evidence, makes testable predictions, and permits inferences and extrapolations (Tononi, 2015). Moreover, IIT is suggested to pro-

[4] Historically, the ideas proposed by Tononi and Edelman have been expressed earlier by Henri Atlan, who applied integrated information to living organisms in general, and consequently to consciousness in particular (Atlan, 1979; Atlan & Fessard, 1972). The connections between Atlan´s perspective and IIT would deserve to be explored further.

vide an assessment of the multiple dimensions of consciousness, namely of quantity and quality, i.e. experience, which is defined as "feel like something".

IIT is open to a number of objections. The most serious criticism concerns its philosophical part.[5] First, concerning the alleged "axioms", i.e. how they are chosen and justified (Bayne, 2018).

Other issues, both empirical and conceptual, arise from the core element of IIT, i.e. integration. From the empirical point of view, recent findings suggest the possibility of information integration without awareness, particularly the possibility of integrative mechanisms established consciously but later instantiated without consciousness (Mudrik, Faivre, & Koch, 2014). Specifically, consciousness could be necessary for new, multisensory, long-range, high-level semantic integration, but not for already learned, short-range and low-level semantic integration. It follows that integration is necessary but not sufficient for consciousness. This requires a redefinition of the relationship between consciousness and unconsciousness for which IIT seems insufficient.

Furthermore, at the conceptual level, integration is a concept that needs to be defined. As outlined by Mudrik et al., it seems likely that IIT refers to the phenomenological concept of integration, which explicitly refers to the so-called "binding problem", i.e. combining different features into a unified percept (Mudrik, Faivre, & Koch, 2016). Integrated information defined as the information possessed by a system as a whole, above and beyond its parts, is also very relevant for IIT. Integrated information so defined seems to be a systemic emergent property. If so, this is in contrast with the alleged intrinsicality of consciousness (Mørch, 2019). Finally, another open conceptual issue is the notion of information that IIT refers to, which according to some critics is a purely structural-dynamical notion that makes IIT unable to avoid an explanatory gap with regard to the nature of consciousness (Mindt, 2017). Moreover, it is a contextual and relational concept rather than an intrinsic feature of physical systems (Searle, 2013).

Looking for a definition

Both GNWT and IIT illustrate that neuroscientific accounts of consciousness are not immune from theoretical and explanatory aims; however, at the same time, they need explicit philosophical complementation in the attempt to approximate a definition of consciousness.

According to widely shared neuroscientific and philosophical views, consciousness is a system-level feature of the brain shaped by its structural and

[5] It is important to stress that the philosophical criticism does not necessarily affect or deny IITs clinical relevance and usefulness.

functional organisation (Dehaene & Changeux, 2011; Northoff, 2014). This means that conscious activity is a property of the brain as a whole. This holistic definition of consciousness can be a limitation of any scientific attempt to detect its so-called neural correlates, i.e. the cerebral regions underpinning conscious activity. There is not a specific *topos* dedicated to consciousness, but the whole brain is involved in conscious activity. However, even though consciousness is a global cerebral phenomenon, it is possible to identify its specific sub-components and to detect the cerebral regions critical for those sub-components.

As briefly summarised above, there are different specific philosophical approaches and related definitions of consciousness, but the main conceptual distinction between access and phenomenal consciousness seems to be conceptually well grounded as well as scientifically and clinically useful. The former refers to the interaction between different mental states, particularly the availability of one state´s content for use in reasoning and rationally guiding speech and action; the latter is the subjective feeling of a particular experience, "what it is like to be" in a particular state (Block, 1995).

Also, the clinical/operational distinction between two components of consciousness, i.e. level (wakefulness) and content (awareness) (Laureys, 2005), seems to be highly relevant, especially to explore the clinical impact of the empirical and conceptual investigations of consciousness.

If we look at the scientific investigations of the last few years, it appears that the most important efforts have been dedicated to investigating access consciousness, while investigation of phenomenal consciousness seems to be more problematic. Moreover, the distinction between access and phenomenal consciousness is not universally accepted among neuroscientists (Schier, 2009) (Kouider, de Gardelle, Sackur, & Dupoux, 2010) (Baars & Laureys, 2005). For instance, Kouider and Dehaene suggest replacing the distinction between different forms of consciousness with that between different levels of conscious access. According to them, the subject is able to access the phenomenal contents, while he cannot verbally report them (Dehaene, Changeux, Naccache, Sackur, & Sergent, 2006). The distinction is accordingly drawn between access consciousness and reportability, not between phenomenal consciousness and access consciousness, since the former is ultimately reduced to the latter.

Moreover, Kouider et al. object to those who identify specific neural mechanisms involved in phenomenal consciousness (Lamme, 2006) that all we can infer from these mechanisms is that the brain processes information without access consciousness, not that it is phenomenally conscious; unconscious phenomenal experiences cannot be demonstrated (Kouider & Dehaene, 2007).

I contested this conclusion (Farisco et al., 2017) because I think that the existence of unconscious phenomenal experiences could theoretically be deduced by an inference to the best explanation, independently of our direct experience of it. This is standard practice in physical theory, for instance. Moreover, from an epistemological point of view, even if it were the case that we could not study phenomenal consciousness without accessing it, this does not imply that the first is identical with the second; the phenomenon to study is different from the means of studying it (Schier, 2009).

There are also empirical arguments supporting the thesis that our phenomenal experience is widespread, so to speak, rather than limited to our conscious access. Namely, on the basis of psychological and neuronal data on iconic and working visual memory, showing that the neural correlates of iconic memory representation share all its essential qualities with the working memory representation, except those that enable access and report. Lamme argues that the existence of phenomenality without report seems to be the more parsimonious conclusion (Lamme, 2004, 2010). With an interesting reference to the need for a more comprehensive account of consciousness, Lamme concludes: 'There are (…) no reasons whatsoever to assume that taking away the modules that enable access and report (…) also takes away the visual phenomenality. In fact, linking visual phenomenality to access and report gives the whole notion of consciousness a poor ontological status' (Lamme, 2010).

Thus, it seems useful to draw a distinction between phenomenal and access consciousness, both from a conceptual and from an empirical point of view, at least by an inference to the best explanation; we do not experience the content of the information without a subjective quality associated to it. As we will see in more details below, this defence of the phenomenal character of consciousness also emerges from recent research on brain development, which stresses the intrinsic activity of the brain (i.e. the brain activity independent from external inputs) and the impossibility to reduce the brain to a simple input/output machine.

Summary of findings

Study I: A new model of consciousness

The starting point of Study I is the description of the brain provided by recent neuroscientific accounts. Neuroscience has moved away from depicting the brain as a simple mechanistic input-output device, and towards a view that describes it as a complex, dynamic and plastic organ that is spontaneously active and projective (Changeux, 1986; Edelman, 1987; Evers, 2009; Laureys, 2015; LeDoux, 2002). From the embryonic to the adult stage, an ongoing spontaneous activity is present throughout the nervous system, particularly but not exclusively at the level of cortical "workspace" neurons, which send and receive projections to many distant areas (Dehaene & Changeux, 2005). Moreover, a wide distributed network of areas has been detected to be more active at rest, i.e. in absence of actual stimulation, than during active task, constituting the so-called resting state (RS) brain activity. The resting state network includes dorsal and ventral medial prefrontal, lateral parietotemporal and posterior cingulate cortices (Gusnard, Raichle, & Raichle, 2001; Vanhaudenhuyse et al., 2011). The spontaneous activity of the brain and its RS activity are the foundation of the brain's relative autonomy from external stimuli. The brain develops spontaneous representations in what has been described as its "projective style", even in absence of actual external signals (Changeux, 2004b; Sanders, Tononi, Laureys, & Sleigh, 2012). Projective here means that the brain is predisposed to build a model of the world that is useful for the satisfaction of its needs and survival. Model here does not directly mean a mental representation, but a particular neuronal configuration corresponding to a specific prediction of the world. As a consequence, even if feedback and feedforward activity in the brain, particularly in cortical layers, is continuous and on-going, our perception results from comparing an internal representation of the world, resulting from both previous feedback loops stored in memory and the spontaneous projective style of the brain, with what is actually perceived (Friston, 2010; Frith, 2007).

Drawing from these scientific views, I argue that consciousness is an intrinsic characteristic of the brain, that the brain is intrinsically conscious, as long as it retains the ability to evaluate and model the world, i.e. it retains an appropriate intrinsic and RS activities. This model can be qualified as Intrinsic Consciousness Theory (ICT). In this perspective, consciousness in its broadest sense corresponds to the phenomenal, evaluative and modelling abilities of the brain. Thus, consciousness is an overarching brain characteristic. More specifically, consciousness thus conceived can express itself in two distinct modalities: explicit (i.e. aware or reflective) consciousness and implicit (i.e. unaware or unreflective) consciousness. The main point is that consciousness exists on a continuum and is not reducible to our higher cognitive abilities (See Fig. 1).

The two modalities of consciousness can have different levels of elaboration and content, and they are dynamical because they exist along a continuum merging with each other. Moreover, they are asymmetrically interconnected, for while it is impossible to have aware consciousness without an underlying unaware consciousness, unaware consciousness can exist without awareness (e.g. in non responsive DOCs) (Dehaene & Changeux, 2005).

If true, it follows that consciousness (both aware and unaware) is a multilevel non-linear dynamic configuration of the brain. According to my framework, the intrinsic and RS activities of the brain are both necessary and sufficient for one modality of consciousness (i.e. unaware consciousness), while they are necessary but not sufficient for the other modality (i.e. aware consciousness).

More specifically, the basic level of consciousness characterising the intrinsically conscious brain can be qualified as non-cognitive rather than cognitive, i.e. not related to the high cognitive functions consciousness is usually identified with (Cerullo, 2015) but rather with a basic capacity of meaningfully interacting with the environment, i.e. of evaluating it by an interaction in which a central role is played by emotions. This basic level of unaware consciousness is phenomenal in itself, a non-cognitive and non-reflective modality of phenomenal consciousness.

Finally, not only does ICT relate consciousness to the intrinsic activity of the brain, it defines such activity as itself conscious, i.e. non-cognitively unaware conscious.[6] My hypothesis is that consciousness is inherent to the architecture of the brain, and that the relationship between brain and consciousness may be defined in terms of conditional necessity. So long as the brain is alive and satisfies some minimal conditions, there will be some level of consciousness. This is why I call the model I propose Intrinsic Consciousness Theory.

[6] I would like to stress that the idea that the living intrinsically active brain is inherently conscious does not entail panpsychism (i.e. consciousness is a metaphysical fundamental constituent of reality). A possible implication of ICT is rather a form of biopsychism, according to which consciousness (at different levels) is intrinsic to biological life. This point is out of the scope of the present thesis, while I plan to analyse it in future works.

FIG.1

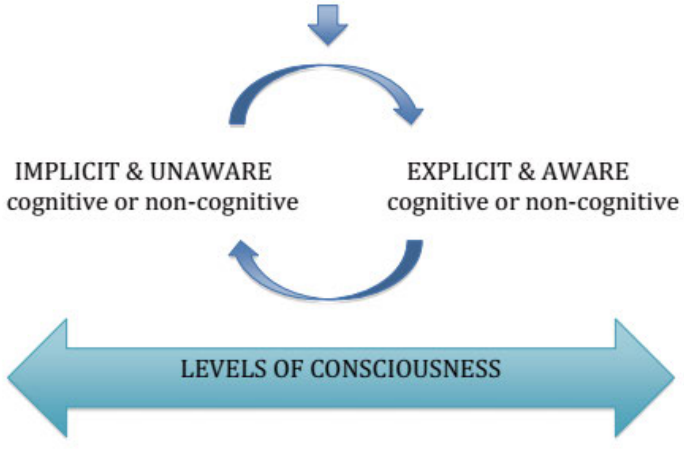
CONSCIOUSNESS

<=>

THE BRAIN´S INTRINSIC & RESTING STATE ACTIVITY

=

THE BRAIN'S PREDISPOSITION TO EVALUATE AND TO MODEL THE WORLD



Consciousness has a deep relation with the brain's intrinsic and resting state activities, specifically with the brain's predisposition to model and evaluate the world. This predisposition may express itself in two modalities, implicit/unaware or explicit/aware, which are not binary but existing on a continuum. Both of them can be cognitive or non-cognitive. Consciousness is an over-arching brain characteristic with different levels.

Study II: Communication with speechless patients

Functional neuroimaging technologies have allowed neuroscientists to monitor the activity of specific brain areas in real time during the execution of particular tasks (Laureys, Boly, & Tononi, 2009). Notwithstanding some technical and procedural limitations (e.g. the risk of motion artefacts and the length of the procedure), functional magnetic resonance imaging (fMRI) is the most commonly used technology in the study of DOCs, especially for its non-invasive nature, ever-increasing availability, relatively high spatiotemporal resolution, capacity to show the entire network of brain areas activated in particular tasks, and capacity to provide both anatomical and functional information in the scanned subject (Laureys et al., 2009).

Among other things, the identification of relevant areas and their monitoring make it possible to implement new forms of interaction with other people and communication based on brain´s reaction to external stimuli rather than on verbal responses.

More specifically, advances in neuroimaging research allow the development of novel investigational paradigms that provide an imaging indication of volition and awareness, even though the reliability of such indication is still discussed (Laureys & Schiff, 2012). One of the earliest studies, conducted by Owen, Laureys and colleagues in 2006 (Owen et al., 2006), is particularly relevant in showing the possible dissociation between the clinical examination based on the behavioural appearance and the results of a neuroimaging assessment (in this case, an fMRI examination). A young woman who survived a car accident was behaviourally diagnosed as being in VS/UWS according to the international guidelines. The research team used some sentences (e.g. ‘There was milk and sugar in his coffee’) and measured her neural responses using fMRI comparing them with responses to acoustically matched noise sequences. Interestingly, the woman’s neural reaction to the sentences was equivalent to the control subjects’ reactions. Nonetheless, this result alone is not sufficient to conclude that the woman is aware because of the possibility of implicit processing; some aspects of human cognition, as language perception and understanding, can go on without awareness (Fine & Florian Jaeger, 2013). For this reason, the research team developed a complementary fMRI study asking the woman to mentally perform two tasks: imagining playing tennis and imagining visiting her house. The relevant result was that the brain activation in the woman was not distinguishable from that of the control subjects, a group of conscious volunteers.

Similar results were obtained in the follow-up study jointly conducted in Liege and Cambridge; fifty-four patients with severe acquired brain injuries were scanned using fMRI. In response to the request to perform imagery tasks, 5 of them were able to modulate their brain activity by generating blood-oxygenation-level-dependent (BOLD) responses that were judged by the researchers as voluntary, reliable and repeatable (Monti et al., 2010). Additional tests in one of the 5 responsive subjects revealed his ability to correctly answer yes–no questions through imagery tasks, showing the feasibility of communication.

The determination of whether an alternative form of communication is feasible is critical for assessing particular ethical issues, such as whether it is conceivable to involve patients with DOCs in their clinical treatment decisions and how?

Research on implementing an fMRI-based communication with patients with DOCs is currently in progress (Sorger, Reithler, Dahmen, & Goebel, 2012). Yet, to date, all these attempts are still at the stage of proofs of concept rather than being practical means to really ensure long-term communication.

For the above-mentioned difficulties, EEG-based communication devices, the so-called brain–computer interfaces (BCI), are being developed as a potentially more practical, transportable and cheaper alternative to fMRI for communicating with patients with DOCs (Bruno, Gosseries, Ledoux, Hustinx, & Laureys, 2011; Lule et al., 2013; Naci et al., 2012; Sellers, 2013; Sorger et al., 2009).

BCI is a direct connection between living neuronal tissue and artificial devices that establishes a non-muscular communication pathway between a computer and a brain (Wolpaw, Birbaumer, McFarland, Pfurtscheller, & Vaughan, 2002). Through BCI, it is possible to detect changes in neuroelectrical activity or brain activity in response to sensory stimulation. The user is then trained to use these changes to select items, words or letters in communication software or to make choices for neuroprosthesis control (Kubler et al., 2009).

While these technologies (fMRI and BCI) are promising for giving back to patients with DOC the ability to communicate their thoughts, their actual use is very difficult and challenging (Lule et al., 2013). For instance, it is possible that patients retain the ability to partially understand commands, to understand but not to follow commands, or to understand and to follow commands but not well enough to make BCI useful. Other variables to take into account in the evaluation of the results emerging from experiments with BCI involving patients with DOCs are the possibility of questions too difficult to answer or asked when the patients were sleeping, and the fact that movement, ocular, and respiration artefacts are involuntary and can interfere with the instrumental assessment with false-positive results (Boly et al., 2005).

Despite these difficulties, the use of neuroimaging-based technologies like BCI is also very promising for better evaluation of patients with DOCs, whose misdiagnosis is a major clinical and ethical problem (Farisco & Petrini, 2014).

A different question that still remains open is whether and how these responding patients may be able to use their brain responses for controlling a BCI and how much integrity and connectivity of the brain is necessary for a minimal communication through BCI (Kubler et al., 2009), and for possibly calling for revisiting the notion of informed consent ascribing to speechless patients like those with DOCs the right of self-determination (Jox, 2016).

Actual requirements for informed consent in clinical context are quite demanding; thus, they seem hardly applicable to patients with DOCs (Farisco et al., 2015). The following table summarises the main challenging cognitive capacities to assess in patients with DOCs for a valid informed consent (APA, 1998; Petrini, 2010):

Understanding the provided information
Appreciating the provided information, i.e. to understand that it is applicable to her/him at a specific time
Executive function, i.e. to organise, plan, and categorise information
Communicating a personal choice, e.g. absence of volitional impairment

These abilities (i.e. understanding, appreciation, reasoning, and choice) are gradable abilities; it is possible that a patient with DOCs retains them only partially. This raises the question as to whether the notion of informed consent requires too much and should then be re-defined. Furthermore, evidence of residual awareness could be flickering and fluctuating. In addition, the patient could decide not to execute the command or be asleep during the execution of the task.

In addition to these cognitive conditions, the emotional dimension of informed consent should be carefully assessed as well (Northoff, 2006). Both cognitive and emotional conditions are particularly challenging to assess in patients with DOCs. Since emotions have strong roots in our unaware consciousness, ICT particularly suggests that focusing only on aware retained abilities might be insufficient for assessing the emotional conditions required for informed consent. This is not to say that the concept of informed consent, as it is now or partly changed, should also be applied to patients with DOCs, but that ICT suggests how to find relevant information about residual conscious activity in patients with DOCs for making more appropriate decisions about their clinical care.

Study III: Brain simulation

Computer modelling and simulations are increasingly used in contemporary attempts to describe, explain and quantitatively predict the human brain's operations.

Study III analyses the specific questions of whether it is plausible to use simulation technologies for emulation of consciousness and what is the potential clinical impact of large-scale brain simulation on the diagnosis and care of DOCs.

I addressed this question in the context of the discussion about NCC, which since its formal introduction at the beginning of the 90s have been widely scrutinised, from both an empirical and conceptual point of view (Chalmers, 2000; Crick & Koch, 1990; Fink, 2016; Koch, Massimini, Boly, & Tononi, 2016; Metzinger, 2000b; Overgaard, 2017).

NCC can be basically defined as minimal neuronal activations sufficient for consciousness (Chalmers, 2000). There are two distinctions that are relevant for my analysis. One is the distinction between NCC of state of consciousness, i.e. marking the difference between being and not being conscious, and NCC of specific consciousness' contents. The second, suggested by Chalmers, is between total NCC (comprising the totality of physical processes absolutely required for a conscious state) and core NCC (comprising only the core processes correlated with the target conscious state) (Chalmers, 2000).

Recent empirical research gives important indication about the identification of both content- and state-consciousness NCC. The most accepted hypothesis is that NCC of specific consciousness' contents correspond to systems in occipital/parietal cortices (early activations) (Aru, Bachmann, Singer, & Melloni, 2012; Koch et al., 2016), while the best candidates for NCC of state-consciousness are localised in a temporo-parietal-occipital zone of the posterior cerebral cortex (Koch et al., 2016). Nevertheless, the questions are open as to whether NCC are localised in the front or the back of the cerebral cortex (Boly et al., 2017) and how to clear the connection of NCC with background conditions, neural prerequisites and neural consequences of conscious experiences (de Graaf, Hsieh, & Sack, 2012).

In the clinical context, specifically in healthcare of DOCs, the identification of the neuronal areas of wakefulness (or level of consciousness) and awareness (or consciousness' content) is particularly relevant (Laureys, 2005; Laureys & Schiff, 2012). To illustrate, the functional and structural integrity of ascending ponto-mesodiencephalic reticular pathways and widespread thalamocortical projections has been shown to be essential for igniting and maintaining the level of consciousness (i.e. wakefulness) (Laureys, Owen, & Schiff, 2004; Steriade, 1996).

Besides the activation of low-level specialised cortices (Boly et al., 2012), awareness requires the activation of a wide frontoparietal network, including lateral and medial frontal regions bilaterally, parieto-temporal and posterior parietal areas bilaterally, posterior cingulate and precuneal cortices (Laureys et al., 1999). Equally correlated with awareness are the connections within the frontoparietal network and between the frontoparietal network and the thalamus (cortico-cortical and cortico-thalamo-cortical connectivity) (Laureys et al., 2000), as well as the general level of functional integrity within the nested hierarchy of neuronal assemblies and ever increasing complex spatial-temporal structures of synchronised neuronal assemblies (Fingelkurts, Fingelkurts, Bagnato, Boccagni, & Galardi, 2012). Significantly, different networks for internal or self-awareness (i.e. relative to the self) and for external or sensory awareness (i.e. relative to the external world) have been identified (midline fronto-parietal and lateral fronto-parietal networks, respectively) (Fingelkurts, Bagnato, Boccagni, & Galardi, 2012; Vanhaudenhuyse et al., 2011).

In theory, two kinds of simulation are possible: a global or large-scale simulation and a discrete or subsystem simulation. While the latter has been used in the neuroscience of consciousness for many years (Dehaene & Changeux, 2005, 2011; Dehaene et al., 2006; Dehaene, Sergent, & Changeux, 2003), the former is much more controversial. More specifically, a large-scale simulation of the brain can in principle be implemented in two ways: as a simulation of the whole brain at different scales simultaneously in runtime or as a simulation of the whole brain at specific scales and levels.

A number of research projects with different goals and methodologies have tried or are still trying to implement a large-scale brain simulation (de Garis, Shuo, Goertzel, & Ruiting, 2010; Markram, 2011; Serban, 2017). They raise several conceptual and methodological concerns (Colombo, 2017; Eliasmith & Trujillo, 2014; Milkowski, 2016; Serban, 2017).

A large-scale simulation of the *conscious* brain, in particular, raises additional technical and conceptual issues. For instance, we have a limited understanding of how the brain is organised at different levels and how these levels interact with each other; thus, important details for a reliable brain simulation might be missed (Dudai & Evers, 2014). Particularly, the simulation of the brain's ability to model the world through cognitive or emotional experiences is presently very challenging because many available models of consciousness do not fill the gap between neurons and the representational capacity of the brain (Pennartz, 2015).

While the simulation of conscious experience as such is hardly conceivable and achievable at least at present, consciousness might be operationalised in terms of NCC and then simulated.

The situation can be more challenging if no specific NCC is identified, or if consciousness correlates with more than one specific brain subcomponent, or if consciousness is considered as emerging from the brain as a whole system.

The following table summarises the identified main challenges for a large-scale simulation of the conscious brain:

The brain is far more than an input-output machine, and its activity between the input and output layer often cannot be precisely reconstructed mathematically.
At the local level, the properties of the brain components are relatively changeable, depending on their reciprocal interaction.
At the global level, the brain exhibits properties and functions that supervene its different, particular components.
In its basic form, both neuroscience and philosophy have described consciousness as a simulation-based interaction with the external environment. Thus, to simulate the conscious brain means to simulate a simulating system, resulting in a kind of second order simulation (or metasimulation).

Study IV: The ethics of addiction

There is a growing discussion about whether addiction should be understood as a brain disease/disorder or as resulting from a non pathological brain dynamics/development (George & Koob, 2017; Hall, Carter, & Forlini, 2015; Leshner, 1997; M. Lewis, 2015; M. Lewis, 2017; Satel & Lilienfeld, 2013; Thibaut & Hoehe, 2017; Volkow, Koob, & McLellan, 2016). These two alternative views result in different interventions: if addiction is a neurobiological pathology, medication is the only way to treat it; if addiction is a non-pathological brain development, then changing the factors causing it might be sufficient to restore a non-addicted brain state.

Against the background of ICT, in study IV, I argue that new scientific perspectives on brain development and the dynamics of consciousness offer the possibility of conceptualizing addiction beyond the abovementioned dualistic interpretation.

Moreover, the ethical discussion about addiction seems to be somewhat limited, and it is mainly focused on normative and practical issues (Carter & Hall, 2015; Carter, Hall, & Illes, 2012), i.e. on the regulatory and practical questions related to the off-label abuse of opioid medication. However, in Study IV, I propose an analysis of factors that are ethically relevant, leading to addictive behaviours and, specifically, of the responsibility for such behaviours.

More specifically, I argue that in addition to the central nervous system's neuronal/neurochemical bases of addiction, socio-economic status, i.e. individual background, modulates through aware and unaware processing what can be described as the person's subjective "global well-being", raising the need for additional rewards in the brain. This need is the basis of addictive behaviour.

Behaviourally, addiction may be described as the result of the loss of or the serious impairment of self-control, decision-making and emotion processing by the subject, where an initially voluntary substance use or behaviour gradually becomes compulsive (Changeux & Lou, 2011; Verdejo-Garcia, Bechara, Recknor, & Perez-Garcia, 2007).

Addiction causes a chemical impact on consciousness. There is evidence that self-awareness, an important component of conscious experience, is determined by a paralimbic circuitry of ɣ synchrony regulated by GABAergic interneurons under the control of acetylcholine and dopamine (Changeux

& Lou, 2011). Accordingly, specific chemical agents and their respective balance modulate awareness.

At the neurophysiological and neurobiological levels, addiction causes the impairment of the paralimbic circuitry, which we have seen to be critical for self-awareness and self-control. Consequently, addiction results in a pharmacological disorder or chemical impairment of conscious self-control and self-regulation through the impairment of paralimbic medial circuitry normal function (Changeux & Lou, 2011; Romer Thomsen et al., 2013). In the end, addiction causes the disruption of the chemical balance critical for self-awareness and self-control, causing a vicious circle for which the dependence from the substance constantly increases (Changeux & Lou, 2011).

I take opioids addiction as an illustrative case. Considering only the US, the numbers concerning such addiction are impressive (Cicero & Ellis, 2017; Volkow & McLellan, 2016). The rate of opioid addiction affected about 2.5 million adults in 2014. However, in 2016, 91.8 million US civilians used prescription opioids, 11.5 million of which misused them (SAMHSA, 2017).

Increases in opioids abuse are related to increases in therapeutic opioids prescription (Cicero & Ellis, 2017). The main claimed reason for those prescriptions is chronic pain, for which the prevalence among adult Americans is between 30 and 40% (Johannes, Le, Zhou, Johnston, & Dworkin, 2010). Opioid medication is now the most prescribed medication in the US (CDC, 2014).

What are the real causes leading to an improper use of opioids?

One of the main factors is the perception of pain as a negative experience to be cured and eventually eliminated. Furthermore, the Joint Commission on the Accreditation of Healthcare Organizations (JCAHO), incorrectly assuming that clinical use of opioids rarely generates addiction, reported that effective narcotic analgesics were wrongly *not* used in US because of an irrational fear of addiction (Phillips, 2000). This report gave an important impulse to the medical prescription of opioid drugs.

An important reason for the continued use of opioids is that they are prescribed by physicians, hence perceived as either less or not dangerous at all (Daniulaityte, Falck, & Carlson, 2012). This fact points not just to the physicians' responsibility (both as causal role and accountable agents) but also shows the influence of implicit biases on the resulting addiction behaviours. In general, the doctor is implicitly seen as an ethically normative actor, someone who clearly makes the difference between licit and illicit behaviours (Rigg & Murphy, 2013). Moreover, because opioids are legally prescribed as painkillers, there is a tendency to regard them as safer than other drugs (Inciardi, Surratt, Cicero, & Beard, 2009).

Another reason for addictive opioids consumption is likely the lack of well-being understood in its widest sense. The epigenetic theory of neuronal development, revealing a deep relationship between the brain and its environment, including social and cultural contexts (Changeux, Courrège, &

Danchin, 1973), suggests that “being well” is a multilevel and multidimensional condition. Well-being can generally be perceived at both the unaware and aware levels, and it results from different factors, both internal and external to the subject (e.g. bodily components and environmental influences). Among the factors impacting on brain development are the influences on subjective well-being coming from the socio-cultural environment, including political, cultural and educational contexts. The information coming from these sources are internalised by the subject and contribute to shaping his or her personal aware and unaware well-being. The relevance of external factors in shaping individual actions raises the issue of social responsibility, if not in ethical terms, at least in terms of public policy.

In particular, an ethical warning informed by scientific data about aware and unaware brain processes should be part of both drug companies’ policies and medical doctors’ professional skills, and relevant tools should be implemented to increase understanding of these topics.

Hence, the argument developed in Study IV rests on the following:

1. scientific: unaware brain processes are massively influenced by socio-economic and ecological factors
2. psychological: addiction is mainly dependent on unaware brain processes, i.e. from loss of aware control
3. ethical: given the scientific and psychological factors mentioned, socio-economic and ecological contexts are highly relevant to addictive dynamics, especially through the influence they have on unaware brain processes.

Discussion

ICT raises a number of criticisms, both conceptual and empirical. For instance, why would we use 'consciousness' as an umbrella term to cover both aware and unaware brain processes? Isn't this counterintuitive and even useless, if not detrimental for the advancement of the scientific investigation of consciousness?

I think that, even if initially counterintuitive, to say that consciousness exists on a continuum (i.e. there is no sharp distinction between conscious and unconscious), and to understand it as an intrinsic brain characteristic (i.e. as a feature fundamental to the brain rather than emerging from it) is clearer and more parsimonious than a binary and dialectic view of the conscious/unconscious and an emergentist model of consciousness.

The counterintuitive and controversial character of ICT is somewhat analogous to the initial attempt to think humankind as part of the animal kingdom; even in that case, a conceptual and metaphysical category (i.e. animality) was stretched beyond its original limits to include something originally kept outside of it (i.e. humankind). Even if problematic at the beginning, this re-conceptualisation of animal/human relationship gradually became widely accepted.

Regarding the relationship between aware and unaware brain operations, to see them as two different modalities of consciousness recognises two important features of what traditionally qualified as unconscious. It is an active and subjectively characterised (i.e. phenomenally shaped) dimension of cerebral life, and it makes a significant contribution to our conscious life (actually being part of it). This is supported also by empirical evidence showing that the alleged monopoly of awareness also on sophisticated cognitive activities is actually misleading.

ICT is monistic; the object to investigate is one (the brain as a modelling and evaluative organ), and we do not need to bridge gaps between different dimensions. There is only a relative and blurring differentiation between modalities of the same (empirical and metaphysical) reality. However, ICT is also monistic in a broader sense, opposing any dualistic view of the brain-consciousness relationship. Consciousness is not assumed as an abstract reality emerging or supervening on the brain, but consciousness is understood as a brain characteristic, i.e. the brain's projective intentionality. To use a word that has been intentionally kept outside the specific focus of this thesis, at the metaphysical level consciousness is conceived of as a biological phenomenon that corresponds to the brain's operation.

I do not think that an explanatory gap between consciousness and brain necessarily entails an ontological gap between them, as argued by others (Chalmers, 1996; Chalmers & Jackson, 2001). I do think there is an explanatory gap between consciousness and brain, but not because they are ontologically different (in fact, I think that consciousness is not different from the brain being conscious), but rather because of our epistemic limitation in understanding biological life in general and the brain in particular. In other words, consciousness and the brain appear different to us because we are unable to know exactly what their mutual relation is. According to ICT, it is reasonable to conceive this relation in terms of identity. Does this entail a form of reductionism? It depends on what we mean by reductionism. If we conceive it as the view that a theory or phenomenon can be reduced to some other theory or phenomenon, particularly that entities of a given kind are collections or combination of entities of a simpler or more basic kind, then ICT is not reductionist. It does not claim that consciousness can be reduced to the sum of more basic brain components, e.g. neurons or neurons' ensembles. According to ICT, even if some brain components do play a critical role, consciousness is a global brain characteristic. This is not to embrace emergentism, i.e. the view that global system properties emerge beyond their particular components; for ICT, consciousness does not emerge from the brain, but it is intrinsic to it.

The different theoretical options regarding the matter/consciousness relationship settled in the Background include nonreductive physicalist monism; experiential properties are not identical to but nonetheless constitutively dependent upon biological properties of the brain. Although sympathetic to this view, I think it is not radical enough and ultimately it results in a kind of compromise. For me, talking about dependence is still a form of dualism, while the very being of consciousness corresponds to the being of the brain. In this sense, ICT is reductionist, because consciousness is nothing more than a brain characteristic. More specifically, to use the above-mentioned conceptual framework, my model supports an a priori physicalist monism. The constitutive relationship between brain and consciousness is a conceptual necessity implied (provided some conditions) by the definition of the brain itself.

ICT is also naturalistic. As stressed by different authors, although naturalism is an attribute fashionable in contemporary philosophy of mind and philosophy of science, its meaning risks to remain unclear (Horst, 2015). A possible approximation to the meaning of naturalism is that it is the view that all the features of the investigated domain should be accommodated within the framework of nature as it is understood by the natural sciences (Horst, 2015). Beyond all the ambiguities of this definition, ICT qualifies as naturalistic. ICT assumes that consciousness is a natural object that can and should be investigated with the tools and methods of natural science, even if at the same time ICT considers natural science as epistemically insufficient and argues for its necessary integration with philosophy. Yet even if interdisciplinary, the theoretical framework of ICT is inspired by the need for logical reasoning on the empirical evidence.

Another point of discussion is whether ICT is explanatory, or rather descriptive. The goal of ICT is to settle the fundamental issue of the relationship between brain and consciousness. Specifically, ICT explores consciousness within the context of the brain, ultimately arguing for their identity. Conceptually, ICT proposes that consciousness is intrinsic to the brain, while it does not aim to describe how specifically this is the case. This is the reason why ICT is theoretically compatible with other more descriptive theories of consciousness, or some parts thereof, like IIT or GNWT. The compatibility lies in the fact that other theories prevalently focus only on one modality of what ICT qualifies as consciousness, i.e. on aware consciousness.

The holistic view proposed by ICT (i.e. consciousness composed by two modalities, aware and unaware) is not less complex than the traditional way of defining consciousness as clearly distinguished from and even opposed to the unconscious, but it offers a different theoretical framework that is simpler and clearer because it conceives the brain as a unified reality with different levels of the same conscious activity.

I think that this framework is also potentially useful for the scientific advancement of the study of consciousness (i.e. for a better description of it), because it suggests a strategy for overcoming traditional gaps and shortcomings of our everyday language as a language of scientific enquiry.

ICT is relevant not only to conceptualise consciousness but also for the related ethical analysis. In general, ICT settles the stage for overcoming some implicit and/or explicit limitations and shortcomings of the usual way of analysing ethical issues emerging from consciousness.

Specifically in clinical contexts, ICT is potentially useful for providing more refined diagnoses and better management of patients with DOCs. In particular, ICT stresses the importance of analysing residual brain intrinsic activity, which potentially reveals residual conscious activity, even if at very basic, possibly unaware levels. The point is that even such levels deserve to be accounted for in the ethical assessment of patients with DOCs.

In fact, while ethical analyses of DOCs traditionally focus on residual awareness, assumed as the only ethically relevant dimension of consciousness, according to ICT this approach is limited.

Contemporary cognitive science increasingly reveals that the traditional way of depicting what is usually called the unconscious as a dimension completely disconnected from and even opposed to consciousness is misleading and overly simplistic (Farisco et al., 2017). Unawareness[7] is far more than a passive repository of information: a lot of monitory work (i.e. active exploration of the environment) takes place at the unaware level, which shows remarkable similarities with awareness as well as a deep connection with it (Evers, 2009) .

There is an extensive amount of empirical data showing both that the unaware brain is able to do many things we usually attribute to awareness only and that the unaware brain operations affect the resulting aware ones. For instance, without awareness, the brain is able to correlate information, associate meanings, reason quickly, develop complex computations, perform sophisticated mathematical operations, selectively focus on information, develop complex inferences (Dehaene, 2014), and even perceive the affective value of stimuli and to influence motivation, value judgment and goal-directed behaviour (Berlin, 2011).

Summarising the empirical evidence, Hassin is quite radical in concluding that the unconscious is able to perform every fundamental high-level cognitive function performed by consciousness (e.g. cognitive control, pursuit of goals, information broadcasting, reasoning) (Hassin, 2013). Hassin´s conclusions are in line with other studies and related interpretations (Dijksterhuis, 2006; Kastrup, 2017) within a new scientific approach that has been developed, called "the new unconscious" (Hassin, Uleman, & Bargh, 2005).

Furthermore, on the basis of extensive empirical evidence, Hassin outlines that the large amount of unaware operations that the brain is able to perform is not disconnected from but has an important impact on the aware operations. For instance, several findings show that subliminal information can drive executive functions (Lau & Passingham, 2007) and that subliminal priming of stimuli changes how we feel about them when we are actually aware of them (Monahan, Murphy, & Zajonc, 2000).

In the ethical discussions about DOCs, two abilities are usually considered central: experiencing well-being and having interests. In the dominant literature on the ethical issue arising from patients with DOCs, these abilities are typically analysed in relation to the residual awareness retained (Fins; Giacino, Fins, Laureys, & Schiff, 2014; Graham et al., 2015; Hawkins, 2016; Kahane & Savulescu, 2009; Levy, 2014; Sheperd, 2016; Sinnott-Armstrong; Wilkinson, Kahane, & Savulescu, 2008). Yet, if we analyse the meaning of both concepts, in light of recent empirical evidence about the unaware brain, we can reasonably conclude that both well-being and having interests might be experienced also at the unaware level.

Well-being can broadly be understood as the positive effect related to what makes life good (according to specific standards) for the individual in question. This broad, abstract understanding of well-being need not be as-

[7] In what follows, the unconscious and unawareness as well as consciousness and awareness are used as synonyms.

signed only to specific levels or modes of consciousness; for instance, we can meaningfully attribute a so defined well-being also to non-human animals. The only condition for well-being broadly considered is the ability to experience its 'positiveness'. Experiencing positiveness is basically an emotional process, and there is increasing evidence that the negative or positive reactions to stimuli (i.e. affective reactions) can occur at the unaware level.

Having an interest in a specific domain can be understood as having a stake in something that can potentially affect what makes our life good within that domain. An interest is what directly and immediately improves life from a certain point of view or within a particular domain, or greatly increases the likelihood of life improvement enabling the subject to realise some good (Hawkins, 2016). Even if there is no general consensus on the definition of good, for the sake of the present discussion we can understand it as what is appropriate for fulfilling a particular need, which of course can be of different kinds. Thus, ‘good’ is something we can benefit from.

What are the minimal capacities that an individual should have for life to be a good for her/him? Specifically analysing DOCs, Hawkins identifies two possible answers (Hawkins, 2016): 1. A life is good if the subject is able to value, or more basically if the subject is able to care. Importantly, Hawkins stresses that caring has no need for cognitive commitment, i.e. for high-level cognitive activities. It requires being able to distinguish something, track it for a while, recognise it over time, and have certain emotional dispositions vis-à-vis something and 2. A life is good if the subject has the capacity to have a relationship with others, i.e. for meaningfully interacting with other people.

As said above, the brain can be described as evaluative also in its unaware operations, in the sense that it is able to distinguish relevant inputs, to track and recognise them, and to emotionally react to them in a relevant way. Sensitivity to reward signals is a fundamental element in the learning process, both in aware and unaware modes (Changeux, 2004b). Moreover, the unaware brain is able to interact with its surroundings in a meaningful way and to produce meaningful information processing of stimuli coming from the external environment, including other people (Farisco & Evers, 2016).

This suggests that unawareness may (at least partly) fulfil both conditions identified by Hawkins for life to be good for a subject, thus making unawareness ethically relevant. It is, of course, a different kind of good than what a paradigmatic cognitive subject with a healthy brain can experience. While I affirm the ethical relevance of unawareness, I leave open the question of its potential ethical value (for example, is the good that an unaware patient with DOC can get from life sufficient for ethically requiring that the patient be kept alive?). I consider this kind of practical and clinical question to be important, but I do not aim to answer it here; rather, I suggest a possible framework allowing a more comprehensive reflection about them.

In the end, I use 'ethical relevance' in the same sense that Goodpaster uses the term 'moral considerability' (i.e. deserving moral consideration) as distinguished from moral significance (i.e. valuable, characterised by a specific moral value) (Goodpaster, 1978). Keeping this definition as reference, two arguments can be provided in support of the ethical relevance of the unconscious / unawareness (Farisco & Evers, 2017):
If awareness is ethically relevant because of what it can enable, and unawareness may enable comparable things, then unawareness is also ethically relevant.

A weaker (less controversial) argument is the following:
If awareness is ethically relevant, and unawareness is (at least in part) the result of awareness, then unawareness is also ethically relevant. Or conversely: if awareness is ethically relevant, and unawareness plays an important role in shaping awareness, then unawareness is also ethically relevant.

I stress that ethical relevance is different than ethical value; to acknowledge that unawareness should be considered in the ethical evaluation does not imply that unawareness retains an intrinsic ethical value. To illustrate, the existence and ethical relevance of interests at the level of unawareness, as I argued for, does not imply that all life is worth living. The kind of good a wholly unaware patient is able to enjoy is not necessarily sufficient for her/his life to be worth living.

The ethical implications of ICT are not only theoretical (i.e. for a wider ethical understanding of both aware and unaware modalities of consciousness) but also practical. In particular, ICT is relevant for the definition of DOCs and for their clinical care decisions. In fact, the semantic enlargement of consciousness and the identification of (very basic levels of) consciousness with the brain's intrinsic and RS activities implies that patients with DOCs are still conscious if their brains are still alive and retain appropriate residual intrinsic and RS activities, specifically a residual ability to model and evaluate the world. This implies that the question will not be 'whether' the patient has any level of consciousness (since this would necessarily be the case), but on what level is it present? Assessing a DOC means assessing the residual intrinsic and RS activities of the brain (Giacino et al., 2014; Northoff, 2014). Promising studies using fMRI, PET or EEG have recently been developed to assess neural signatures of conscious Default Mode Network (DMN) (Barttfeld et al., 2015; Chennu et al., 2017; Demertzi et al., 2015; Sitt et al., 2014; Stender et al., 2014); nonetheless, even in this case, consciousness is understood as awareness, and the unaware levels are not taken into account.

Moreover, on the bases of empirical evidence (Barttfeld et al., 2015; Huang et al., 2015; Kouider, de Gardelle, Dehaene, Dupoux, & Pallier, 2010; Mudrik et al., 2014; Raichle, 2015), the ICT argues that consciousness is heterogeneous but continuous. Empirical studies also suggest that DOCs are heterogeneous but continuous (Bruno et al., 2012; Noirhomme et al., 2010). The differentiation between different forms of DOCs might not be as sharp as traditionally thought (Fisher & Truog, 2017). Residual consciousness in DOCs can be assessed in terms of both aware and unaware consciousness as defined above. It is important to take into account this bimodal and multilevel characterisation of consciousness in assessing patients with DOCs; otherwise, we risk looking only for the highest levels of consciousness, underestimating the possibility that the patient retains other lower levels of consciousness, particularly the unaware and possibly non-cognitive consciousness related to the intrinsic and RS brain activities.

From this fact, within the ICT framework, a possible ethical consequence arises, i.e. that we should treat the damaged brain so as to respect its ability to retain a certain degree of phenomenality.

From the long-term perspective, holding that as long as the brain is intrinsically active it is *ipso facto* intrinsically phenomenal (even if at different levels) might support the view that we should care about living brains, and we should make efforts in approaching patients with DOCs, not only on the basis of their retained aware abilities but also on the basis of their retained unaware abilities, which are currently not well investigated and may be underestimated.

More specifically, acknowledging the ethical relevance of unawareness as it is reasonable to do starting from ICT, notably calls for further development and refinement of:

4. diagnostics (including also the need for nosological revision)
5. assessments and interpretations of subjective states (trying to identify unaware, e.g. emotional, states)
6. adaptations of living conditions (taking into account the possibility of unaware positive and negative conditions, e.g. emotions)
7. therapeutic interventions.

In particular, the interpretation of subjective states, as well as the adaptation of living conditions, should acknowledge that negative and positive emotions are not necessarily aware. To illustrate, new stimulations of patients' unaware perception might be implemented in order to increase their unaware well-being (e.g. through tactile, olfactory or acoustic inputs), possibly involving those who are close to the patient.

End-of-life clinical decisions could be affected by including unawareness in the ethical assessment. The end-of-life case clearly illustrates that the ethical relevance of unawareness *per se* is neutral with regard to specific clinical decisions. One person could consider it right to withdraw life-sustaining care because of the high risk of aware and/or unaware negative emotions, while another would consider such withdrawal unjust because of the possibility of residual positive aware and/or unaware emotions.

As pointed out in Study II, recent investigations of residual consciousness in patients with DOCs led to new possibilities for communicating with them, i.e. in a 'cerebral' communication without external behaviour. This new form of communication with patients with DOCs raises two fundamental issues, with related ethical implications: 1. The necessity to develop proper clinical and/or neuroimaging protocols in order to assess this risk of false positive, in other terms, the necessity to clear when the activation of a cerebral region is equivalent to or evidence of the presence of awareness and 2. The necessity to refine the tools we have for exploring and possibly interacting with the retained unaware levels of brain activity in patients with DOCs. At the ethical level, 1. arises a number of issues, e.g. the possibility to involve patients with DOCs in informed consent process. As seen above, even if a form of cerebral communication with patients with DOCs is possible, assessing their capacity to understand the provided information in order to express a valid informed consent is a challenge, and it seems as yet premature to assume that a 'cerebral communication' is enough to assess important ethical issues like informed consent. 2. outlines the necessity to enlarge the scope of our approach to retained consciousness in patients with DOCs, in order to also include possible unaware levels, which are highly relevant, for instance, to a reliable informed consent.

Another technology that can potentially be used in the assessment of DOCs is brain simulation. As stressed above, there are both conceptual and technical limitations especially to a large-scale brain computer simulation. Notwithstanding such limitations, there are important advantages coming from the application of simulation to DOCs (Markram, 2013):

- no limit on what we can record, i.e. we can obtain a potentially unlimited amount of data from a simulation (as everything in the model is measurable)
- no limit on the number of manipulations we can perform (i.e. all model parameters can be manipulated)
- enhanced replicability and interpretation of experiments
- the possibility of building bridges between different levels of brain organisation (i.e. the possibility of understanding the relative correlation between different space and time scales within the brain)
- the possibility to simulate brain diseases with major clinical diagnostic, prognostic and possibly therapeutic implications.

I suggest that computer simulation, ultimately combined with classical brain measurements and neuroimaging, particularly for verification and/or validation, might help overcome the limitation of traditional measurements of brain activity. Specifically, in the case of patients with DOCs whose involvement in neuroimaging measurements may be both technically and ethically challenging, a computer simulation could fill the gap of missing data, or give clinicians the tool for predicting the future development of a disorder or the outcome of a particular treatment. Furthermore, it is theoretically possible to simulate different scenarios (e.g. different medications) through a computer simulation, particularly the effects of treatment at different brain levels (e.g. molecular, neuronal and synaptic) and at the intersection of different levels. It is also possible to manipulate and to replicate experiments in order to get the most informative data on the patients' present and future conditions. In this way, simulation may allow for more informed decisions about the patients' treatment.

In the light of ICT, simulating consciousness means simulating the intrinsic, modelling activity of the brain, whether that is aware or unaware.

ICT is relevant for an ethical analysis, not only of DOCs but also of other social contexts where recognising both modalities of consciousness is key to better assess emerging issues. The case of addiction is highly emblematic; the ICT framework suggests a strategy to overcome the classical dichotomy between disease/not disease interpretations of addicted states, focusing instead on the underlying brain-environment interactions and related individual/social responsibilities.

Specifically, it is necessary to acknowledge the continued oscillation of the addicted subject between aware and unaware drives, which denotes different psychological, neurological and pharmacological processes in the brain. Since neuroscience is providing increasing knowledge of these processes, management strategies should consider both the aware and unaware brain. Of course, such strategies can be implemented in different ways, e.g. through a direct pharmacological approach or through an indirect approach aiming at influencing the brain by altering external environmental conditions, including cultural and social institutions. The epigenetic development of the brain opens the door to the massive influence of external environment on the subjective lifestyle.

Conclusions

A conceptual analysis of recent neuroscientific findings shows the plausibility of a monistic view of consciousness conceived as an overarching concept covering both aware and unaware brain processes. Accordingly, ICT describes consciousness as an intrinsic characteristic of the brain: if the brain retains appropriate intrinsic and resting state activities, then it is necessarily conscious, even if at very basic levels. Even if controversial, this approach promises to give a new impulse to the scientific research on consciousness and to provide new insights for the ethical assessment of related issues, namely of DOCs care and addiction.

According to ICT, the ethical analysis of DOCs should also include unaware levels of brain activity, recognising the possibility that the subjects might retain the ability to process information behind awareness (e.g. experiencing positive or negative emotions).

Accordingly, the attempt to implement a form of speechless, cerebral communication with patients with DOCs, as well as the attempt to simulate their retained brain activity, should try to also include unaware processes, e.g. intrinsic brain activity.

The ethical analysis of addiction should acknowledge the involvement of both aware and unaware levels of brain processing in causing the addictive state, including important influences from external environment.

Future Research

Several further issues arise from the research behind the present thesis. Among them, two are particularly sensitive and deserve a careful analysis: the potential panpsychist implications of ICT and the implication of ICT for artificial intelligence (AI), namely for the possibility of artificial consciousness.

In fact, the description of consciousness provided by ICT might suggest that life comprises (or is intrinsically predisposed to) a fundamental, very basic, level of conscious activity. This could be understood as organic consciousness because it is an intrinsic characteristic of a living organism. If so, ICT would entail a form of panpsychism (i.e. consciousness is fundamental to physical reality), or, as I am tempted to think, a form of biopsychism (i.e. consciousness is fundamental to biological life).

Yet, this is not a necessary conclusion. ICT is intended to describe consciousness within the brain context, and it identifies consciousness with the intrinsic brain activity, provided some necessary conditions are satisfied. This means that ICT is not biopsychist *strictu sensu*; rather, it supports what we can call a conditional form of brainpsychism. If the brain retains appropriate intrinsic and resting state activities, then it retains a (possibly very basic) level of consciousness. Still, ICT as such leaves the door open to a possible interpretation in biopsychistic terms.

Since ICT is very sympathetic to biological naturalism (Searle, 2007), it raises the issue of the plausibility of AI, and more specifically of artificial consciousness. I am inclined to think that consciousness is a biological feature. A very naïve reason in support of this is that the only form of consciousness I know is the one instantiated in biological organisms. The issue is whether biology is a necessary condition for consciousness. I am not closed to the possibility of artificial forms of consciousness, but I also think that if this were to be the case, such artificial consciousness would be qualitatively different from biological consciousness.

In conclusion, the most fascinating aspect of consciousness is that it finally emerges not as a single problem, but rather as a bundle of many fascinating questions, requiring a multidisciplinary and collaborative never ending effort.

Acknowledgments

Sometimes life is surprising. Before starting my experience at CRB under the supervision of Kathinka Evers, I was inclined to take for granted that life's surprises are most of the time negative: my philosophical inclination (not to mention my professional ambitions) was on the brink of depression. But then, six years ago, at the end of the spring, my phone rang showing an unknown number from abroad. And that was the beginning of a kind of new life for me.

Philosophy is not simply an activity I do for living, but rather my way of living. This is why Kathinka gave me the chance to be really myself again. As a philosopher, but also as a man. After these years of shared reflection, discussions, agreements, disagreements, I feel very changed and closer to my true identity. I have always felt supported and welcomed, free to go on my way, never asked to make compromises. This is really rare. Thank you Kathinka for giving me back the joy to be a philosopher and for having opened my mind to the multifaceted beauty of life.

My work has greatly benefited also from my friendship with Arleen Salles: our conversations (where you actually speaks slightly more than me…) are always provoking and extremely productive for me. Your passion for thought is really contagious: I'm very happy to have you on my side not only as a colleague, but first as a friend. Together with Kathinka you are my philosophical family, where I like to imagine myself under your protecting wing.

My greatest thanks to all the CRB colleagues who welcomed me as a member from the beginning, particularly to Mats Hansson, Amal Matar, Anna Höglund, Daniel Pallarés Domínguez, Jennifer Viberg Johansson, Josepine Fernow, Karl Sallin, Manuel Guerrero, Stefan Eriksson, Ulrik Kihlbom, and the "Little Italy" members Deborah Mascalzoni and Mirko Ancillotti. A special thanks to Pär Segerdahl, a true philosopher, who inspired me with his words and his silence.

A special thanks also to my co-supervisor, Steven Laureys, who welcomed me in his lab nine years ago introducing me to the fascinating world of consciousness.

My work would not have been possible without the important support of the European Union's Horizon 2020 Framework Programme for Research and Innovation under the Specific Grant Agreement No. 785907, Human Brain Project, which funded my research activity. I really feel privileged to have had the opportunity to be paid for doing what I really love. And my thanks to all the colleagues of the Sub-Project 12, Ethics and Society: the multidisciplinarity of the group has been an invaluable help to free me from many biases.

I am very grateful to the scientists that collaborated with me during my doctoral research in addition to my co-supervisor: Jeannette Hellgren Kotaleski and Jean-Pierre Changeux. I would have never thought to have the opportunity to think and write together with so important scientists. This is actually an achievement in itself.

My thanks to my brother, who often visited me during my periods in Uppsala showing me again his love and support.

Finally, my thanks to my wife, Olimpia, and to my children, Ilde and Andrea. I remember the burden of every second I spent away from you: sometimes it was almost devastating. Yet in the end it was a krisis which broke the banality of a relationship taken for granted, giving me the opportunity to re-discover our bond every day. Your support (particularly from you, Olimpia) has been invaluable. In this way we have been really a family.

References

APA. (1998). Guidelines for assessing the decision-making capacities of potential research subjects with cognitive impairment *Am J Psychiatry, 155*(11), 1649-1650.

Aru, J., Bachmann, T., Singer, W., & Melloni, L. (2012). Distilling the neural correlates of consciousness. *Neurosci Biobehav Rev, 36*(2), 737-746. doi:10.1016/j.neubiorev.2011.12.003

Atlan, H. (1979). *Entre le cristal et la fumée essai sur l'organisation du vivant*. Paris: Éditions du Seuil.

Atlan, H., & Fessard, A. (1972). *L'Organisation biologique et la théorie de l'information préface de A. Fessard*. Paris: Hermann.

Baars, B. J. (1988). *A cognitive theory of consciousness*. Cambridge Cambridgeshire ; New York: Cambridge University Press.

Baars, B. J., & Laureys, S. (2005). One, not two, neural correlates of consciousness. *Trends Cogn Sci, 9*(6), 269; author reply 270. doi:10.1016/j.tics.2005.04.008

Barttfeld, P., Uhrig, L., Sitt, J. D., Sigman, M., Jarraya, B., & Dehaene, S. (2015). Signature of consciousness in the dynamics of resting-state brain activity. *Proc Natl Acad Sci U S A, 112*(3), 887-892. doi:10.1073/pnas.1418031112

Bayne, T. (2018). On the axiomatic foundations of the integrated information theory of consciousness. *Neurosci Conscious, 2018*(1), niy007. doi:10.1093/nc/niy007

Bennett, M. R., Dennett, D. C., Hacker, P. M. S., & Searle, J. (2007). *Neuroscience and philosophy : brain, mind, and language*. New York: Columbia University Press.

Berlin, H. A. (2011). The neural basis of the dynamic unconscious. *Neuropsychoanalysis, 13*(1), 5-31.

Block, N. (1995). On a confusion about a function of consciousness. *Behavioral and Brain Sciences, 18*(2), 227-287.

Boly, M., Faymonville, M. E., Peigneux, P., Lambermont, B., Damas, F., Luxen, A., . . . Laureys, S. (2005). Cerebral processing of auditory and noxious stimuli in severely brain injured patients: differences between VS and MCS. *Neuropsychol Rehabil, 15*(3-4), 283-289. doi:10.1080/09602010443000371

Boly, M., Massimini, M., Garrido, M. I., Gosseries, O., Noirhomme, Q., Laureys, S., & Soddu, A. (2012). Brain connectivity in disorders of consciousness. *Brain Connect, 2*(1), 1-10. doi:10.1089/brain.2011.0049

Boly, M., Massimini, M., Tsuchiya, N., Postle, B. R., Koch, C., & Tononi, G. (2017). Are the Neural Correlates of Consciousness in the Front or in the Back of the Cerebral Cortex? Clinical and Neuroimaging Evidence. *J Neurosci, 37*(40), 9603-9613. doi:10.1523/JNEUROSCI.3218-16.2017

Bruno, M. A., Gosseries, O., Ledoux, D., Hustinx, R., & Laureys, S. (2011). Assessment of consciousness with electrophysiological and neurological imaging techniques. *Curr Opin Crit Care, 17*(2), 146-151. doi:10.1097/MCC.0b013e328343476d

Bruno, M. A., Majerus, S., Boly, M., Vanhaudenhuyse, A., Schnakers, C., Gosseries, O., . . . Laureys, S. (2012). Functional neuroanatomy underlying the clinical subcategorization of minimally conscious state patients. *J Neurol, 259*(6), 1087-1098. doi:10.1007/s00415-011-6303-7
Carruthers, P. (2000). *Phenomenal consciousness : a naturalistic theory*. Cambridge, UK ; New York: Cambridge University Press.
Carter, A., & Hall, W. (2015). What is addiction neuroethics? In J. Clausen & N. Levy (Eds.), *Handbook of Neuroethics* (pp. 995-998). Dordrecht Springer Science+Business Media
Carter, A., Hall, W., & Illes, J. (2012). *Addiction neuroethics : the ethics of addiction neuroscience research and treatment* (1st ed.). London: Academic Press.
CDC. (2014). FastStats. Therapeutic drug use. Retrieved from http://www.cdc.gov/nchs/fastats/drug-use-therapeutic.htm
Cerullo, M. A. (2015). The Problem with Phi: A Critique of Integrated Information Theory. *PLoS Comput Biol, 11*(9), e1004286. doi:10.1371/journal.pcbi.1004286
Chalmers, D. (1996). *The conscious mind : in search of a fundamental theory*. New York: Oxford University Press.
Chalmers, D. (2000). What is a neural correlate of consciousness? In T. Metzinger (Ed.), *Neural Correlates of Consciousness: Empirical and Conceptual Questions* (pp. 17-39). Cambridge, MA: MIT Press.
Chalmers, D., & Jackson, F. (2001). Conceptual analysis and reductive explanation. *Philosophical Review 110*(3), 315-361.
Changeux, J. P. (1986). *Neuronal man : the biology of mind*. New York: Oxford University Press.
Changeux, J. P. (2004a). *The physiology of truth : neuroscience and human knowledge*. Cambridge, Mass.: Belknap Press of Harvard University Press.
Changeux, J. P. (2004b). *The physiology of truth : neuroscience and human knowledge*. Cambridge, Mass.: Belknap Press of Harvard University Press.
Changeux, J. P. (2017). Climbing Brain Levels of Organisation from Genes to Consciousness. *Trends Cogn Sci, 21*(3), 168-181. doi:10.1016/j.tics.2017.01.004
Changeux, J. P., Courrège, P., & Danchin, A. (1973). A theory of the epigenesis of neuronal networks by selective stabilization of synapses. *Proc Natl Acad Sci U S A, 70*(10), 2974-2978.
Changeux, J. P., & Lou, H. C. (2011). Emergent pharmacology of conscious experience: new perspectives in substance addiction. *FASEB J, 25*(7), 2098-2108. doi:10.1096/fj.11-0702ufm
Chennu, S., Annen, J., Wannez, S., Thibaut, A., Chatelle, C., Cassol, H., . . . Laureys, S. (2017). Brain networks predict metabolism, diagnosis and prognosis at the bedside in disorders of consciousness. *Brain, 140*(8), 2120-2132. doi:10.1093/brain/awx163
Cicero, T. J., & Ellis, M. S. (2017). The prescription opioid epidemic: a review of qualitative studies on the progression from initial use to abuse. *Dialogues Clin Neurosci, 19*(3), 259-269.
Colombo, M. (2017). Why build a virtual brain? Large-scale neural simulations as jumps start for cognitive computing. *Journal of Experimental & Theoretical Artificial Intelligence, 29*(2), 361-370.
Crick, F., & Koch, K. (1990). Towards a neurobiological theory of consciousness. *Semin. Neurosci., 2*, 263-275.
Daniulaityte, R., Falck, R., & Carlson, R. G. (2012). "I'm not afraid of those ones just 'cause they've been prescribed": perceptions of risk among illicit users of

pharmaceutical opioids. *Int J Drug Policy, 23*(5), 374-384. doi:10.1016/j.drugpo.2012.01.012
de Garis, H., Shuo, C., Goertzel, B., & Ruiting, L. (2010). A world survey of artificial brain projects, Part I: Large-scale brain simulations. *Neurocomputing, 74*, 3-29.
de Graaf, T. A., Hsieh, P. J., & Sack, A. T. (2012). The 'correlates' in neural correlates of consciousness. *Neurosci Biobehav Rev, 36*(1), 191-197. doi:10.1016/j.neubiorev.2011.05.012
Dehaene, S. (2014). *Consciousness and the brain : deciphering how the brain codes our thoughts*. New York, New York: Viking Adult.
Dehaene, S., & Changeux, J. P. (2005). Ongoing spontaneous activity controls access to consciousness: a neuronal model for inattentional blindness. *PLoS Biol, 3*(5), e141. doi:10.1371/journal.pbio.0030141
Dehaene, S., & Changeux, J. P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron, 70*(2), 200-227. doi:10.1016/j.neuron.2011.03.018
Dehaene, S., Changeux, J. P., & Naccache, L. (2011). The Global Neuronal Workspace Model of Conscious Access: From Neuronal Architectures to Clinical Applications. In S. Dehaene, Christen, Y. (Ed.), *Characterizing Consciousness: From Cognition to the Clinic?* (pp. 55-84). Berlin-Heidelberg: Springer-Verlag.
Dehaene, S., Changeux, J. P., Naccache, L., Sackur, J., & Sergent, C. (2006). Conscious, preconscious, and subliminal processing: a testable taxonomy. *Trends Cogn Sci, 10*(5), 204-211. doi:10.1016/j.tics.2006.03.007
Dehaene, S., Kerszberg, M., & Changeux, J. P. (1998). A neuronal model of a global workspace in effortful cognitive tasks. *Proc Natl Acad Sci U S A, 95*(24), 14529-14534.
Dehaene, S., & Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework. *Cognition, 79*(1-2), 1-37.
Dehaene, S., Sergent, C., & Changeux, J. P. (2003). A neuronal network model linking subjective reports and objective physiological data during conscious perception. *Proc Natl Acad Sci U S A, 100*(14), 8520-8525. doi:10.1073/pnas.1332574100
Demertzi, A., Antonopoulos, G., Heine, L., Voss, H. U., Crone, J. S., de Los Angeles, C., . . . Laureys, S. (2015). Intrinsic functional connectivity differentiates minimally conscious from unresponsive patients. *Brain, 138*(Pt 9), 2619-2631. doi:10.1093/brain/awv169
Dennett, D. C. (1991). *Consciousness explained* (1st ed.). Boston: Little, Brown and Co.
Dijksterhuis, A., Nordgren, L.F. (2006). A theory of unconscious thought. *Perspectives on Psychological Science, 1*(2), 95-109.
Dudai, Y., & Evers, K. (2014). To simulate or not to simulate: what are the questions? *Neuron, 84*(2), 254-261. doi:10.1016/j.neuron.2014.09.031
Edelman, G. M. (1987). *Neural Darwinism : the theory of neuronal group selection*. New York: Basic Books.
Eliasmith, C., & Trujillo, O. (2014). The use and abuse of large-scale brain models. *Curr Opin Neurobiol, 25*, 1-6. doi:10.1016/j.conb.2013.09.009
Evers, K. (2007). Towards a philosophy for neuroethics. An informed materialist view of the brain might help to develop theoretical frameworks for applied neuroethics. *EMBO Rep, 8 Spec No*, S48-51. doi:10.1038/sj.embor.7401014
Evers, K. (2009). *Neuroetique. Quand la matière s'éveille*. Paris: Odile Jacob.

Evers, K., Salles, A., & Farisco, M. (2017). Theoretical framing of neuroethics: the need for a conceptual approach. In E. Racine, Aspler, J. (Ed.), *Debates about Neuroethics: perspectives on its development, focus and future* (pp. 89-107). Dordrecht: Springer International Publishing.
Fallon, F. (2016). Integrated Information Theory of Consciousness *Internet Encyclopedia of Philosophy*.
Farisco, M., & Evers, K. (2016). *Neurotechnology and direct brain communication. New insights and responsibilities concerning speechless but communicative subjects*. New York: Routledge.
Farisco, M., & Evers, K. (2017). The ethical relevance of the unconscious. *Philos Ethics Humanit Med, 12*(1), 11. doi:10.1186/s13010-017-0053-9
Farisco, M., Evers, K., & Changeux, J. P. (2018). Drug Addiction: From Neuroscience to Ethics. *Front Psychiatry, 9*, 595. doi:10.3389/fpsyt.2018.00595
Farisco, M., Kotaleski, J. H., & Evers, K. (2018). Large-Scale Brain Simulation and Disorders of Consciousness. Mapping Technical and Conceptual Issues. *Front Psychol, 9*, 585. doi:10.3389/fpsyg.2018.00585
Farisco, M., Laureys, S., & Evers, K. (2015). Externalization of consciousness. Scientific possibilities and clinical implications. *Curr Top Behav Neurosci, 19*, 205-222. doi:10.1007/7854_2014_338
Farisco, M., Laureys, S., & Evers, K. (2017). The intrinsic activity of the brain and its relation to levels and disorders of consciousness. *Mind&Matter, 15*(2), 197-219.
Farisco, M., & Petrini, C. (2014). Misdiagnosis as an ethical and scientific challenge. *Ann Ist Super Sanita, 50*(3), 229-233. doi:DOI: 10.4415/ANN_14_03_05
Farisco, M., Salles, A., & Evers, K. (2018). Neuroethics: A Conceptual Approach. *Camb Q Healthc Ethics, 27*(4), 717-727. doi:10.1017/S0963180118000208
Fine, A. B., & Florian Jaeger, T. (2013). Evidence for implicit learning in syntactic comprehension. *Cognitive Science, 37*(3), 578-591.
Fingelkurts, A., Bagnato, S., Boccagni, C., & Galardi, G. (2012). DMN Operational Synchrony Relates to Self-Consciousness: Evidence from Patients in Vegetative and Minimally Conscious States. *Open Neuroimag J, 6*, 55-68. doi:10.2174/1874440001206010055
Fingelkurts, A., Fingelkurts, A., Bagnato, S., Boccagni, C., & Galardi, G. (2012). Toward operational architectonics of consciousness: basic evidence from patients with severe cerebral injuries. *Cogn Process, 13*, 111-131.
Fink, S. B. (2016). A Deeper Look at the "Neural Correlate of Consciousness". *Front Psychol, 7*, 1044. doi:10.3389/fpsyg.2016.01044
Fins, J. J. *Rights come to mind : brain injury, ethics, and the struggle for consciousness*.
Fisher, D. B., & Truog, R. D. (2017). The Problems with Fixating on Consciousness in Disorders of Consciousness. *AJOB Neuroscience, 8*(3), 135-140.
Friston, K. (2010). The free-energy principle: a unified brain theory? *Nat Rev Neurosci, 11*(2), 127-138. doi:10.1038/nrn2787
Frith, C. (2007). *Making up the Mind: How the Brain Creates Our Mental World*. London: Wiley-Blackwell.
Gazzaniga, M. (2011). *Who's In Charge? Free Will and the Science of the Brain*. New York: Harper Collins.
Gennaro, R. J. (2012). *The consciousness paradox : consciousness, concepts, and higher-order thoughts*. Cambridge, Mass.: MIT Press.
George, O., & Koob, G. F. (2017). Individual differences in the neuropsychopathology of addiction. *Dialogues Clin Neurosci, 19*(3), 217-229.

Giacino, J. T., Fins, J. J., Laureys, S., & Schiff, N. D. (2014). Disorders of consciousness after acquired brain injury: the state of the science. *Nat Rev Neurol, 10*(2), 99-114. doi:10.1038/nrneurol.2013.279
Goodpaster, K. (1978). On being morally considerable. *Journal of Philosophy, 75*, 308-325.
Graham, M., Weijer, C., Cruse, D., Fernandez-Espejo, D., Gofton, T., Gonzalez-Lara, L. E., . . . Owen, A. M. (2015). An Ethics of Welfare for Patients Diagnosed as Vegetative With Covert Awareness. *AJOB Neuroscience, 6*(2), 31-41. doi:10.1080/21507740.2015.1014072
Gusnard, D. A., Raichle, M. E., & Raichle, M. E. (2001). Searching for a baseline: functional imaging and the resting human brain. *Nat Rev Neurosci, 2*(10), 685-694. doi:10.1038/35094500
Hall, W., Carter, A., & Forlini, C. (2015). The brain disease model of addiction: is it supported by the evidence and has it delivered on its promises? *Lancet Psychiatry, 2*(1), 105-110. doi:10.1016/S2215-0366(14)00126-6
Hameroff, S., & Penrose, R. (2014). Consciousness in the universe: a review of the 'Orch OR' theory. *Phys Life Rev, 11*(1), 39-78. doi:10.1016/j.plrev.2013.08.002
Hassin, R. R. (2013). Yes It Can: On the Functional Abilities of the Human Unconscious. *Perspectives on Psychological Science, 8*(2), 195-207.
Hassin, R. R., Uleman, J. S., & Bargh, J. A. (2005). *The new unconscious*. Oxford: Oxford University Press.
Hawkins, J. (2016). What is good for them? Best interests and severe disorders of consciousness. In W. Sinnott-Armstrong (Ed.), *Finding Consciousness: The Neuroscience, Ethics, and Law of Severe Brain Damage* (pp. 181-206). Oxford: Oxford University Press.
Horst, P. (2015). Beyond Reduction: From Naturalism to Cognitive Pluralism. *Mind&Matter, 12*(2), 197-244.
Huang, Z., Zhang, J., Longtin, A., Dumont, G., Duncan, N. W., Pokorny, J., . . . Northoff, G. (2015). Is There a Nonadditive Interaction Between Spontaneous and Evoked Activity? Phase-Dependence and Its Relation to the Temporal Structure of Scale-Free Brain Activity. *Cereb Cortex*. doi:10.1093/cercor/bhv288
Inciardi, J. A., Surratt, H. L., Cicero, T. J., & Beard, R. A. (2009). Prescription opioid abuse and diversion in an urban community: the results of an ultrarapid assessment. *Pain Med, 10*(3), 537-548. doi:10.1111/j.1526-4637.2009.00603.x
Johannes, C. B., Le, T. K., Zhou, X., Johnston, J. A., & Dworkin, R. H. (2010). The prevalence of chronic pain in United States adults: results of an Internet-based survey. *J Pain, 11*(11), 1230-1239. doi:10.1016/j.jpain.2010.07.002
Jox, R. (2016). Disorders of consciousness and informed consent. In M. Farisco, Evers, K. (Ed.), *Neurotechnology and Direct Brain Communication. New insights and responsibilities concerning speechless but communicative subjects* (pp. 133-142). London - New York: Routledge.
Kahane, G., & Savulescu, J. (2009). Brain damage and the moral significance of consciousness. *J Med Philos, 34*(1), 6-26. doi:10.1093/jmp/jhn038
Kastrup, B. (2017). There Is an 'Unconscious,' but It May Well Be Conscious. *Eur J Psychol, 13*(3), 559-572. doi:10.5964/ejop.v13i3.1388
Koch, C. (2012). *Consciousness: Confessions of a Romantic Reductionist*. Cambridge, MA: MIT Press.
Koch, C., Massimini, M., Boly, M., & Tononi, G. (2016). Neural correlates of consciousness: progress and problems. *Nat Rev Neurosci, 17*(5), 307-321. doi:10.1038/nrn.2016.22

Kouider, S., de Gardelle, V., Dehaene, S., Dupoux, E., & Pallier, C. (2010). Cerebral bases of subliminal speech priming. *Neuroimage, 49*(1), 922-929. doi:10.1016/j.neuroimage.2009.08.043
Kouider, S., de Gardelle, V., Sackur, J., & Dupoux, E. (2010). How rich is consciousness? The partial awareness hypothesis. *Trends Cogn Sci, 14*(7), 301-307. doi:10.1016/j.tics.2010.04.006
Kouider, S., & Dehaene, S. (2007). Levels of processing during non-conscious perception: a critical review of visual masking. *Philos Trans R Soc Lond B Biol Sci, 362*(1481), 857-875. doi:10.1098/rstb.2007.2093
Kriegel, U. (Forthcoming). What is the philosophy of consciousness? In U. Kriegel (Ed.), *Oxford Handbook of the Philosophy of Consciousness*. Oxford: Oxford University Press.
Kriegel, U., & Williford, K. (2006). *Self-representational approaches to consciousness*. Cambridge, Mass. ; London: MIT.
Kubler, A., Furdea, A., Halder, S., Hammer, E. M., Nijboer, F., & Kotchoubey, B. (2009). A brain-computer interface controlled auditory event-related potential (p300) spelling system for locked-in patients. *Ann N Y Acad Sci, 1157*, 90-100. doi:10.1111/j.1749-6632.2008.04122.x
Lamme, V. A. (2004). Separate neural definitions of visual consciousness and visual attention; a case for phenomenal awareness. *Neural Netw, 17*(5-6), 861-872. doi:10.1016/j.neunet.2004.02.005
Lamme, V. A. (2006). Towards a true neural stance on consciousness. *Trends Cogn Sci, 10*(11), 494-501. doi:10.1016/j.tics.2006.09.001
Lamme, V. A. (2010). How neuroscience will change our view on consciousness. *Cogn Neurosci, 1*(3), 204-220. doi:10.1080/17588921003731586
Lau, H. C., & Passingham, R. E. (2007). Unconscious activation of the cognitive control system in the human prefrontal cortex. *J Neurosci, 27*(21), 5805-5811. doi:10.1523/JNEUROSCI.4335-06.2007
Laureys, S. (2005). The neural correlate of (un)awareness: lessons from the vegetative state. *Trends Cogn Sci, 9*(12), 556-559. doi:10.1016/j.tics.2005.10.010
Laureys, S. (2015). *Un si brillant cerveau. Les états limites de consciences*: Odile Jacob.
Laureys, S., Boly, M., & Tononi, G. (2009). Functional neuroimaging. In S. Laureys & G. Tononi (Eds.), *The neurology of consciousness. Cognitive Neuroscience and Neuropathology* (pp. 31-42). London: Elsevier.
Laureys, S., Faymonville, M. E., Luxen, A., Lamy, M., Franck, G., & Maquet, P. (2000). Restoration of thalamocortical connectivity after recovery from persistent vegetative state. *Lancet, 355*(9217), 1790-1791.
Laureys, S., Goldman, S., Phillips, C., Van Bogaert, P., Aerts, J., Luxen, A., . . . Maquet, P. (1999). Impaired effective cortical connectivity in vegetative state: preliminary investigation using PET. *Neuroimage, 9*(4), 377-382. doi:10.1006/nimg.1998.0414
Laureys, S., Owen, A. M., & Schiff, N. D. (2004). Brain function in coma, vegetative state, and related disorders. *Lancet Neurol, 3*(9), 537-546. doi:10.1016/S1474-4422(04)00852-X
Laureys, S., & Schiff, N. D. (2012). Coma and consciousness: paradigms (re)framed by neuroimaging. *Neuroimage, 61*(2), 478-491. doi:10.1016/j.neuroimage.2011.12.041
LeDoux, J. E. (2002). *Synaptic self : how our brains become who we are*. New York: Viking.

Leshner, A. I. (1997). Addiction is a brain disease, and it matters. *Science, 278*(5335), 45-47.
Levy, N. (2014). The Value of Consciousness. *J Conscious Stud, 21*(1-2), 127-138.
Lewis-Williams, J. D. (2002). *The mind in the cave : consciousness and the origins of art*. London: Thames & Hudson.
Lewis, M. (2015). *The biology of desire : why addiction is not a disease* (First edition. ed.). New York, NY: PublicAffairs.
Lewis, M. (2017). Addiction and the Brain: Development, Not Disease. *Neuroethics, 10*(1), 7-18. doi:10.1007/s12152-016-9293-4
Lou, H. C., Changeux, J. P., & Rosenstand, A. (2016). Towards a cognitive neuroscience of self-awareness. *Neurosci Biobehav Rev*. doi:10.1016/j.neubiorev.2016.04.004
Lule, D., Noirhomme, Q., Kleih, S. C., Chatelle, C., Halder, S., Demertzi, A., . . . Laureys, S. (2013). Probing command following in patients with disorders of consciousness using a brain-computer interface. *Clin Neurophysiol, 124*(1), 101-106. doi:10.1016/j.clinph.2012.04.030
Markram, H. (2011). Introducing the Human Brain Projeect. *Procedia Comput. Sci., 7*, 39-42.
Markram, H. (2013). Seven challenges for neuroscience. *Funct Neurol, 28*(3), 145-151. doi:10.11138/FNeur/2013.28.3.144
Metzinger, T. (2000a). *Neural Correlates of Consciousness: Empirical and Conceptual Questions*. Cambridge, MA: MIT Press.
Metzinger, T. (Ed.) (2000b). *Neural Correlates of Consciousness: Empirical and Conceptual Issues*. Cambridge, MA: MIT Press.
Milkowski, M. (2016). Explanatory completeness and idealization in large brain siulations: a mechanistic perspective. *Synthese, 193*, 1457-1478.
Mindt, G. (2017). The Problem with the 'Information' in Integrated Information Theory. *Journal of Consciousness Studies, 24*(7-8), 130-154.
Monahan, J. L., Murphy, S. T., & Zajonc, R. B. (2000). Subliminal mere exposure: specific, general, and diffuse effects. *Psychol Sci, 11*(6), 462-466. doi:10.1111/1467-9280.00289
Monti, M. M., Vanhaudenhuyse, A., Coleman, M. R., Boly, M., Pickard, J. D., Tshibanda, L., . . . Laureys, S. (2010). Willful modulation of brain activity in disorders of consciousness. *N Engl J Med, 362*(7), 579-589. doi:10.1056/NEJMoa0905370
Mørch, H. H. (2019). Is Consciousness Intrinsic? A Problem for the Integrated Information Theory. *Journal of Consciousness Studies, 26*(1-2), 133-162.
Mudrik, L., Faivre, N., & Koch, C. (2014). Information integration without awareness. *Trends Cogn Sci, 18*(9), 488-496. doi:10.1016/j.tics.2014.04.009
Mudrik, L., Faivre, N., & Koch, C. (2016). Information integration without awareness: (Trends in Cognitive Sciences 18, 488-496; September, 2014). *Trends Cogn Sci, 20*(7), 559. doi:10.1016/j.tics.2016.05.005
Naci, L., Monti, M. M., Cruse, D., Kubler, A., Sorger, B., Goebel, R., . . . Owen, A. M. (2012). Brain-computer interfaces for communication with nonresponsive patients. *Ann Neurol, 72*(3), 312-323. doi:10.1002/ana.23656
Noirhomme, Q., Soddu, A., Lehembre, R., Vanhaudenhuyse, A., Boveroux, P., Boly, M., & Laureys, S. (2010). Brain connectivity in pathological and pharmacological coma. *Front Syst Neurosci, 4*, 160. doi:10.3389/fnsys.2010.00160
Northoff, G. (2006). Neuroscience of decision making and informed consent: an investigation in neuroethics. *J Med Ethics, 32*(2), 70-73. doi:10.1136/jme.2005.011858

Northoff, G. (2014). *Unlocking the brain. Volume 2, Consciousness*. Oxford: Oxford University Press.
Overgaard, M. (2017). The Status and Future of Consciousness Research. *Front Psychol, 8*.
Owen, A. M., Coleman, M. R., Boly, M., Davis, M. H., Laureys, S., & Pickard, J. D. (2006). Detecting awareness in the vegetative state. *Science, 313*(5792), 1402. doi:10.1126/science.1130197
Pennartz, C. M. (2015). *The Brain´s Representational Power. On Consciousness and the Integration of Modalities*. Cambridge, MA: MIT Press.
Petrini, C. (2010). Informed consent in experimentation involving mentally impaired persons: ethical issues. *Ann Ist Super Sanita, 46*(4), 411-421. doi:10.4415/ANN_10_04_09
Phillips, D. M. (2000). JCAHO pain management standards are unveiled. Joint Commission on Accreditation of Healthcare Organizations. *JAMA, 284*(4), 428-429.
Raichle, M. E. (2015). The restless brain: how intrinsic activity organizes brain function. *Philos Trans R Soc Lond B Biol Sci, 370*(1668). doi:10.1098/rstb.2014.0172
Rigg, K. K., & Murphy, J. W. (2013). Understanding the etiology of prescription opioid abuse: implications for prevention and treatment. *Qual Health Res, 23*(7), 963-975. doi:10.1177/1049732313488837
Romer Thomsen, K., Joensson, M., Lou, H. C., Moller, A., Gross, J., Kringelbach, M. L., & Changeux, J. P. (2013). Altered paralimbic interaction in behavioral addiction. *Proc Natl Acad Sci U S A, 110*(12), 4744-4749. doi:10.1073/pnas.1302374110
Rosenthal, D. M. (1997). A theory of consciousness. In N. Block, Flanagan, O., Guzeldere, G. (Ed.), *The Nature of Consciousness* (pp. 729-754). Cambridge, MA: MIT Press.
Roskies, A. (2002). Neuroethics for the new millenium. *Neuron, 35*(1), 21-23.
SAMHSA. (2017). Results from the 2016 National Survey on Drug Use and Health: Detailed Tables. Retrieved from https://www.samhsa.gov/data/sites/default/files/NSDUH-DetTabs-2016/NSDUH-DetTabs-2016.pdf
Sanders, R. D., Tononi, G., Laureys, S., & Sleigh, J. W. (2012). Unresponsiveness not equal unconsciousness. *Anesthesiology, 116*(4), 946-959. doi:10.1097/ALN.0b013e318249d0a7
Satel, S., & Lilienfeld, S. O. (2013). Addiction and the brain-disease fallacy. *Front Psychiatry, 4*, 141. doi:10.3389/fpsyt.2013.00141
Schier, E. (2009). Identifying phenomenal consciousness. *Conscious Cogn, 18*(1), 216-222. doi:10.1016/j.concog.2008.04.001
Searle, J. R. (2007). Biological Naturalism. In M. Velmans & S. Schneider (Eds.), *The Blackwell Companion to Consciousnss* (pp. 325-334). Malden MA, Oxford, Victoria: Blackwell Publishing Ltd.
Searle, J. R. (2013). Can Information Theory Explain Consciousness? *The New York Review of Books*.
Sellers, E. W. (2013). New horizons in brain-computer interface research. *Clin Neurophysiol, 124*(1), 2-4. doi:10.1016/j.clinph.2012.07.012
Serban, M. (2017). Learning from large-scale neural simulations. *Prog Brain Res, 233*, 129-148. doi:10.1016/bs.pbr.2017.05.004
Sheperd, J. (2016). Moral conflict in the minimally conscious state. In W. Sinnott-Armstrong (Ed.), *Finding Consciousness: The Neuroscience, Ethics, and Law of Severe Brain Damage* (pp. 160-179). Oxford: Oxford University Press.

Sinnott-Armstrong, W. *Finding consciousness : the neuroscience, ethics, and law of severe brain damage* (First edition. ed.).
Sitt, J. D., King, J. R., El Karoui, I., Rohaut, B., Faugeras, F., Gramfort, A., . . . Naccache, L. (2014). Large scale screening of neural signatures of consciousness in patients in a vegetative or minimally conscious state. *Brain, 137*(Pt 8), 2258-2270. doi:10.1093/brain/awu141
Sorger, B., Dahmen, B., Reithler, J., Gosseries, O., Maudoux, A., Laureys, S., & Goebel, R. (2009). Another kind of 'BOLD Response': answering multiple-choice questions via online decoded single-trial brain signals. *Prog Brain Res, 177*, 275-292. doi:10.1016/S0079-6123(09)17719-1
Sorger, B., Reithler, J., Dahmen, B., & Goebel, R. (2012). A real-time fMRI-based spelling device immediately enabling robust motor-independent communication. *Curr Biol, 22*(14), 1333-1338. doi:10.1016/j.cub.2012.05.022
Stender, J., Gosseries, O., Bruno, M. A., Charland-Verville, V., Vanhaudenhuyse, A., Demertzi, A., . . . Laureys, S. (2014). Diagnostic precision of PET imaging and functional MRI in disorders of consciousness: a clinical validation study. *Lancet, 384*(9942), 514-522. doi:10.1016/S0140-6736(14)60042-8
Steriade, M. (1996). Awakening the brain. *Nature, 383*(6595), 24-25. doi:10.1038/383024a0
Thibaut, F., & Hoehe, M. (2017). Addictive behaviors: where do we stand, and where are we going? *Dialogues Clin Neurosci, 19*(3), 215.
Thompson, E. (2015). *Waking, dreaming, being : self and consciousness in neuroscience, meditation, and philosophy*. New York: Columbia University Press.
Tononi, G. (2008). Consciousness as integrated information: a provisional manifesto. *Biological Bulletin, 2015*, 216-242.
Tononi, G. (2015). Integrated information theory *Scholapedia* (Vol. 10 (1)).
Tononi, G., Boly, M., Massimini, M., & Koch, C. (2016). Integrated information theory: from consciousness to its physical substrate. *Nat Rev Neurosci, 17*(7), 450-461. doi:10.1038/nrn.2016.44
Tye, M. (1995). *Ten problems of consciousness : a representational theory of the phenomenal mind*. Cambridge, Mass.: MIT Press.
Van Gulick, R. (2014). Consciousness. *The Stanford Encyclopedia of Philosophy.* Retrieved from http://plato.stanford.edu/archives/spr2014/entries/consciousness/>
Vanhaudenhuyse, A., Demertzi, A., Schabus, M., Noirhomme, Q., Bredart, S., Boly, M., . . . Laureys, S. (2011). Two distinct neuronal networks mediate the awareness of environment and of self. *J Cogn Neurosci, 23*(3), 570-578. doi:10.1162/jocn.2010.21488
Verdejo-Garcia, A., Bechara, A., Recknor, E. C., & Perez-Garcia, M. (2007). Negative emotion-driven impulsivity predicts substance dependence problems. *Drug Alcohol Depend, 91*(2-3), 213-219. doi:10.1016/j.drugalcdep.2007.05.025
Volkow, N. D., Koob, G. F., & McLellan, A. T. (2016). Neurobiologic Advances from the Brain Disease Model of Addiction. *N Engl J Med, 374*(4), 363-371. doi:10.1056/NEJMra1511480
Volkow, N. D., & McLellan, A. T. (2016). Opioid Abuse in Chronic Pain--Misconceptions and Mitigation Strategies. *N Engl J Med, 374*(13), 1253-1263. doi:10.1056/NEJMra1507771
Whiteley, C. (2019). The Neutrality Thesis, Neural Correlates and the Encroachment of Science on Metaphysics. Retrieved from https://www.cecilywhiteley.com/post/the-neutrality-thesis-neural-correlates-and-the-encroachment-of-science-on-metaphysics

Wilkinson, D., Kahane, G., & Savulescu, J. (2008). "Neglected personhood" and neglected questions: remarks on the moral significance of consciousness. *Am J Bioeth, 8*(9), 31-33. doi:10.1080/15265160802412486

Wolpaw, J. R., Birbaumer, N., McFarland, D. J., Pfurtscheller, G., & Vaughan, T. M. (2002). Brain-computer interfaces for communication and control. *Clin Neurophysiol, 113*(6), 767-791.

Acta Universitatis Upsaliensis

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine 1597

Editor: The Dean of the Faculty of Medicine

A doctoral dissertation from the Faculty of Medicine, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine. (Prior to January, 2005, the series was published under the title "Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine".)

Distribution: publications.uu.se
urn:nbn:se:uu:diva-392187

ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2019