UNIVERSITÁ DEGLI STUDI DI MILANO
FACOLTÁ DI MEDICINA E CHIRURGIA

DOTTORATO DI RICERCA IN FISIOLOGIA
SETTORE SCIENTIFICO DISCIPLINARE BIO-09
CICLO XXIV°

Tesi di Dottorato di Ricerca

PROBING THE BRAIN'S CAPACITY FOR CONSCIOUSNESS THROUGH THE SPATIOTEMPORAL COMPLEXITY OF THE CORTICAL ACTIVITY EVOKED BY TRANSCRANIAL MAGNETIC STIMULATION

Dottorando: Dott. Adenauer Girardi Casali

Matricola: R08233

Tutor: Prof. Maurizio Mariotti
Dipartimento di Scienze Cliniche "Luigi Sacco"

Co-Tutor: Dott. Marcello Massimini

Coordinatore: Prof. Paolo Cavallari

Anno Accademico 2010-2011

Contents

Summary v

List of Abbreviations viii

List of Figures ix

List of Publications xi

1. INTRODUCTION: CONSCIOUSNESS AND THE BRAIN 2

1.1 - Towards a Science of Consciousness 5
1.1.1 – *Empirical Findings in the Absence of Theoretical Guidance* 5
1.1.2 – *Towards a Theory of Consciousness?* 7
1.1.3 – *The Project of an Empirical Science of Consciousness* 9

1.2 – Consciousness: Brain, World and Causality 12
1.2.1 – *A Manifold of Features and the Brain's Functional Specialization* 13
1.2.2 – *The Unity of Experience and Integration in the Brain* 15
1.2.3 – *The Content of Consciousness and Causal Interactions in the Brain* 18

1.3 – TMS/hd-EEG: a Window to Causality in the Brain 19
1.3.1 – *Transcranial Magnetic Stimulation* (TMS) 21
1.3.2 – *The Combination of TMS with High-density Electroencephalography* (TMS/hd-EEG) 23

1. 4 - Plan of the Thesis 24

2. METHODS ON TMS/HD-EEG 26

2.1 – TMS/hd-EEG Recording 27
2.1.1 – *TMS Targeting* 27
2.1.2 – *hd-EEG Recording* 27
2.1.3 – *EEG Pre-processing* 28

2.2 – Does the response to TMS reflect particular properties of stimulated circuits rather than stereotypical and/or random reactions? 30
2.2.1 – *Quantifying sensitivity and specificity: the Divergence Index (DI)* 31
2.2.2 – *TMS-evoked potentials are sensitive and specific for changes in perturbation parameters* 32

2.3 – How to extract a signal of neural activation from the EEG? 37
2.3.1 – *The Forward Model* 38
2.3.2 – *Estimating Primary Currents* 39
2.3.3 – *Automatic Anatomical Classification* 42

2.4 – How to access cortical activation produced exclusively by causal interactions in the stimulated circuits? 43

2.5 – Can we quantitatively describe the pattern of cortical activation evoked by TMS? 45
2.5.1 – *Indices of Cortical Responsiveness* 46
2.5.2 – *Interpretation of Indices of Cortical Responsiveness* 48
2.5.3 – *Measuring Local Excitability* 54
2.5.4 – *Measuring Cortical Connectivity* 55

3. CAUSAL RESPONSES OF THE THALAMOCORTICAL SYSTEM TO TRANSCRANIAL MAGNETIC STIMULATION 58

3.1 – Alert Wakefulness **58**
3.1.1 – *The Natural Frequencies of Corticothalamic Circuits* 58
3.1.2 – *The Spatiotemporal Complexity of Cortical Activation Evoked by TMS during Wakefulness* 65

3.2 – Sleep **67**
3.2.1 – *Slow-wave Sleep* 67
3.2.2 – *REM Sleep* 70

3.3 – Anesthesia **72**
3.3.1 – *Experimental Protocol* 73
3.3.2 – *Results* 73

3.4 – Disorders of Consciousness **78**
3.4.1 – *TMS/hd-EEG in DOC Patients: Experimental Protocol* 78
3.4.2 – *Data Analysis and Results* 79

4. THE PERTURBATIONAL COMPLEXITY INDEX **89**

4.1 – Calculating the complexity of the cortical activity evoked by TMS **89**
4.1.1 – *The Lempel-Ziv Algorithm for Binary Matrices* 92
4.1.2 – *The Lempel-Ziv Perturbational Complexity Index* (PCI_{LZ}) 93
4.1.3 – *The Transposed Lempel-Ziv Perturbational Complexity Index* (PCI^{T}_{LZ}) 94
4.1.4 – *The Conditional Entropy Perturbational Complexity Index* (PCI_{CE}) 95

4.2 – Testing PCI **96**
4.2.1 – *PCI allowed the discrimination of alert wakefulness from NREM sleep and anesthesia-LOC, independent of the strength and the extent of cortical activation* 96
4.2.2 – *PCI allowed a clear-cut distinction between conscious and unconscious healthy subjects, independent of the stimulated area and stimulation intensity* 99
4.2.3 – *The discrimination of PCI between consciousness and unconsciousness did not depend on the particular method used to calculate complexity* 99
4.2.4 – *The PCI time-course revealed a reproducible time-scale related to the emergence of complex TMS-evoked cortical activation during alert wakefulness* 101
4.2.5 – *PCI increased progressively, in parallel with the level of consciousness* 103
4.2.6 – *PCI is a measure of the spatiotemporal complexity of the cortical activity evoked by TMS* 106

5. COMPLEXITY AND THE BRAIN'S CAPACITY FOR CONSCIOUSNESS **109**

5.1 – Possible Physiological Mechanisms Underlying Changes in the Spatiotemporal Complexity of Cortical Activity **112**

5.2 – Limitations of the Study and Future Developments **117**

5.3 – The Brain's Capacity for Consciousness and the Science of the Mind **119**
5.3.1 – *The Search for the Contents of Consciousness* 119
5.3.2 – *The Objective Character of Intentionality* 122
5.3.3 – *PCI and the Science of the Mind* 126

Appendix **130**

References **137**

Acknowledgements **156**

Summary

Neuroscience faces the challenging task of developing and implementing objective measures of consciousness that can be applied to patients who are unable to interact with their external environment. The standard clinical assessment of these patients relies heavily on the subjective distinction between voluntary and involuntary or reflexive movements and electrophysiological and neuroimaging protocols have been recently developed to improve diagnosis and probe for signs of awareness. However, because the ability to unambiguously infer the capacity for consciousness through these novel techniques is determined ultimately not by consciousness itself but the awareness of a specific stimulus, their use to diagnose consciousness at the single-patient level is challenged by difficulties related to the application and interpretation of results. Clearly, the diagnosis of brain-injured patients could benefit greatly from a scientific investigation of the bases of consciousness that could give evidence not only of the neural correlates of particular conscious perceptions, but mainly of the conditions necessary for supporting all and any conscious experience.

This thesis addresses the possibility for investigating the brain's capacity for consciousness following a path that has not yet been explored. General considerations about what constitutes the content of consciousness and the hypothesis that the brain must be involved in sustaining conscious experience led us to hypothesize that consciousness depends on the brain's capacity to sustain causal interactions between different areas of the thalamocortical system. To investigate this hypothesis, one should employ a perturbational approach, directly stimulating the cerebral cortex, avoiding possible subcortical filtering or gating, and recording the neural effects of the perturbation with the appropriate temporal resolution. Today, such a perturbational approach to explore causality in the thalamocortical system can be implemented non-invasively in humans, thanks to the combination of navigated transcranial magnetic stimulation (TMS) and high-density electroencephalography (hd-EEG).

In a series of recent papers we developed standardized and data-driven procedures to TMS/hd-EEG signal processing, demonstrating that this technique allows exploring particular properties of stimulated circuits, rather than stereotypical and/or random reactions (Casarotto et. al, 2010), such as natural frequencies of cortical oscillation (Rosanova et al., 2009;

Pigorini et al., 2011), thresholds for cortical activation, cortico–cortical delays and patterns of neural connectivity (Casali et al., 2010). We applied these methods in the study of the cortical activation evoked by TMS in healthy subjects during alert wakefulness (Rosanova et al., 2009; Huber et al., 2012; Johnson et al., 2011) and anesthesia (Ferrarelli et al., 2010), confirming results observed previously during slow-wave sleep (Massimini et al., 2005, 2007, 2009). In all cases, what fundamentally distinguished TMS-evoked responses when consciousness was clearly present from responses in conditions in which consciousness was unambiguously absent was an overall reduction of the spatiotemporal complexity of the cortical activity induced by the TMS during loss of consciousness. Moreover, performing TMS/hd-EEG measurements at the bedside of brain-injured, non-communicating patients, who gradually recovered consciousness from the vegetative state (VS), we observed that the spatiotemporal complexity of cortical activity consistently increased during the recovery of cognitive function underlying the evolution from VS to minimally conscious state (MCS) (Rosanova et al., 2012).

Therefore, aiming at an objective evaluation of the brain's capacity for consciousness, we developed and tested a feasible measure of spatiotemporal complexity, the Perturbational Complexity Index (PCI), which is high only if many regions of the cerebral cortex react to the initial perturbation quickly and in different ways (Casali et al., 2012). Remarkably, in a total of 116 TMS sessions collected from 19 healthy subjects and 17 brain-injured patients, we invariably found high PCI values in conditions in which consciousness was clearly present and low PCI values in conditions in which consciousness was unambiguously reduced. This difference was able to reliably discriminate between conscious and unconscious healthy subjects, producing disjoint distributions that were independent of the stimulation parameters, the strength and the extent of the cortical activation. Moreover, PCI was able to detect progressive changes in consciousness, such as those that occur while a subject is falling asleep, and to discriminate between ambiguous consciousness levels in patients suffering from disorders of consciousness (MCS) from both lower (VS, sleep/anesthesia-LOC) and higher (LIS, healthy wakefulness) levels of consciousness.

The spatiotemporal complexity of the cortical activity evoked by TMS is a single number that can be calculated at the bedside with little *a priori* information. Because this measure aims at the brain's capacity for consciousness, instead of behavioral or neural correlations of conscious perception, this technique does not depend on the willingness or

ability of the patient to engage in assessment protocols and can be employed bypassing sensory pathways and subcortical structures to directly probe the thalamocortical system. Our results support PCI as an appropriate tool to approximate an objective measure of the neural correlate of consciousness with the potential to assist the diagnosis and prognosis in brain-injured patients and with unique theoretical implications to a science of consciousness.

List of Abbreviations

AEP : Auditory Evoked Potential
BA : Brodmann's Area
bPL : Broad-band Phase-locking
CRS-R : Coma Recovery Scale-Revised
CT : Computerized Tomography
DI : Divergence Index
DOC : Disorders of Consciousness
EEG : Electroencephalography
EMCS : Emergence from Minimally Conscious State
EOG : Electrooculogram
ERP: Event-related Potential
ERSP : Event-related Spectral Perturbation
fMRI : Functional Magnetic Resonance Imaging
gERSP : Global Event-related Spectral Perturbation
GMFP : Global Mean Field Power
hd-EEG : high-density Electroencephalography
LIS : Locked-in Syndrome
LOC : Loss of Consciousness
MCS : Minimally Conscious State
MEG : Magnetoencephalography
MNI : Montreal Neurological Institute
MRI : Magnetig Resonance Imaging
NBS : Navigated Brain Stimulation
NCC : Neural Correlates of Consciousness
NREM: Non-rapid Eye Movement
NSET : Nearest-sensor EEG Threshold
OAA/S : Observer's Assessment of Alertness/Sedation
PCI : Perturbational Complexity Index
PCI_{CE} : Conditional Entropy Perturbational Complexity Index
PCI_{LZ} : Lempel-Ziv Perturbational Complexity Index
PET : Positron Emission Tomography
REM : Rapid Eye Movement
ROC : Receiver Operating Characteristic
SCD : Significant Current Density
SCS : Significant Current Scattering
SEP : Somatosensory Evoked Potential
SPM : Statistical Parametric Mapping
SS : Statistically Significant Sources
TMS : Transcranial Magnetic Stimulation
VS : Vegetative State
WMN : Weighted Minimum Norm

List of Figures

Figure 2.1: Single-subject TMS-evoked potentials. 29
Figure 2.2: Non-parametric statistical procedure to perform single-subject pairwise comparisons between TMS-evoked potentials 33
Figure 2.3: Results of pairwise comparisons between TMS-evoked potentials of a representative subject. 34
Figure 2.4: Divergence Index of all pairwise comparisons between TMS-evoked potentials. 36
Figure 2.5: Accuracy of source reconstruction after non-parametric statistics 45
Figure 2.6: Synthetic indices of cortical responsiveness to TMS in a single subject at different stimulation intensities 50
Figure 2.7: Time course of SCD, bPL and SCS 51
Figure 2.8: Spatial distribution of SCD, bPL and SCS 52
Figure 2.9: Instantaneous spatial distribution of SCD, bPL and SCS 53
Figure 2.10: Caracterizing Cortical Excitability 55
Figure 2.11: Caracterizing Cortical Connectivity 57
Figure 3.1: TMS induced global EEG oscillations that are specific for the stimulated site 60
Figure 3.2: The specificity of natural frequencies was preserved with different stimulation intensities 61
Figure 3.3: The specificity of the natural frequency was reproducible across subjects 62
Figure 3.4: The natural frequency is a local property of individual thalamocortical modules 64
Figure 3.5: TMS evoked a balanced, long-range differentiated pattern of activation during wakefulness 66
Figure 3.6: The balanced, long-range, differentiated pattern of activation observed in wakefulness was impaired during slow-wave sleep 68
Figure 3.7: The ability of thalamocortical circuits to produce differentiated responses was impaired during slow-wave sleep 70
Figure 3.8: The widespread differentiated pattern of TMS-evoked activation observed in wakefulness was gradually impaired upon falling into NREM sleep and partially recovered during REM sleep 71
Figure 3.9: The balanced, long-range, differentiated pattern of activation observed in wakefulness was impaired during anesthesia-induced LOC 74
Figure 3.10. TMS during anesthesia evoked a large positive–negative wave in the stimulation site but little activation in distant areas 75
Figure 3.11. TMS-evoked brain responses gradually changed while transitioning from high (OAA/S = 5) to low (OAA/S = 1) levels of vigilance 76
Figure 3.12: Cortical connectivity (SCS), but not reactivity (SCD), captured cortical changes during LOC 77
Figure 3.13: Examples of TMS-evoked cortical responses of patients who evolved from coma into different clinical states 81
Figure 3.14. TMS-evoked cortical responses in Group I patients 82
Figure 3.15. Clinical evaluation and TMS-evoked cortical responses in Group II patients 84
Figure 3.16. EEG spectra showed evident changes from MCS to EMCS but not from VS to MCS 86
Figure 3.17. Effective connectivity for all patients and TMS/EEG measurements 87
Figure 4.1: Significant sources (SS) binary matrices manifested the reduction of the spatiotemporal complexity of TMS-evoked cortical activation during slow-wave sleep 91

Figure 4.2: Complexity of the cortical activity evoked by TMS (PCI_{LZ}) but not reactivity (SCD) or connectivity (SCS) unambiguously discriminated alert wakefulness from NREM sleep and anesthesia-LOC 98
Figure 4.3: PCI_{LZ} discriminated between conscious and unconscious healthy subjects, independently of the stimulated area and the stimulation intensity 100
Figure 4.4: The discrimination between conscious and unconscious healthy subjects was preserved using the PCI calculated based on conditional entropy (PCI_{CE}). 101
Figure 4.5: The complexity time-course indicates a reproducible time-scale related to the emergence of complex TMS-evoked cortical activations during alert wakefulness 102
Figure 4.6: Perturbational complexity was progressively reduced during the transition to NREM sleep and partially recovered during REM sleep 104
Figure 4.7: PCI_{LZ} increased progressively in parallel with the level of consciousness 105
Figure 4.8: PCI discriminated intermediate levels of consciousness (MCS and EMCS) from both lower (VS) and higher (LIS) levels of consciousness in brain-injured patients 106
Figure 4.9: PCI^{T}_{LZ} and PCI_{LZ} are strongly correlated 107
Figure A.1: Diagram of the algorithm used to calculate the Lempel-Ziv complexity of a binary matrix. 130
Figure A.2: Source Entropy (H) and Signal-to-noise ratio (SNR) are displayed for all sessions in which PCI was calculated. 131
Figure A.3: PCI^{T}_{LZ} discriminated between conscious and unconscious healthy subjects, independently of the stimulated area and the stimulation intensity. 132
Figure A.4: PCI^{T}_{LZ} increased progressively in parallel with the level of consciousness 132

List of Publications

Casali, A. G., Casarotto, S., Rosanova, M., Mariotti, M. and Massimini, M. "General indices to characterize the electrical response of the cerebral cortex to TMS". *NeuroImage*, **49**: 1459-68, 2010.

Casali, A. G., Rosanova, M., Gosseries, O., Ferrarelli, F., Casarotto, S., D'Oliveira, T., Casali, K. R., Boly, M., Laureys, S., Tononi, G. and Massimini, M.,. "Gauging brain complexity in wakefulness, sleep, anesthesia and coma" (submitted for publication), 2012.

Casarotto, S., Romero Lauro, L. J., Bellina, V., **Casali, A. G.**, Rosanova, M., Pigorini, A., Defendi, S., Mariotti, M. and Massimini, M., "EEG Responses to TMS Are Sensitive to Changes in the Perturbation Parameters and Repeatable over Time". *Plos ONE*, **5**:e10281, 2010.

Ferrarelli, F., Massimini, M., Sarasso, S., **Casali, A. G**., Riedner, B. A., Angelini, G., Tononi, G. and Pearce, R. A., "Breakdown in cortical effective connectivity during midazolam-induced loss of consciousness". *Proc. Natl. Acad. Sci. USA*, **107**: 2681-86, 2010.

Huber, R., Mäki, H., Rosanova, M., Casarotto, S., Canali, P., **Casali A. G.,** Tononi, G. and Massimini, M., "Human cortical excitability increases with time awake", *Cereb. Cortex* (in press), 2012.

Johnson, J., Kundu, B., **Casali, A. G**. and Postle, B., "Effects of Physiological State on Temporal and Spectral Properties of the TMS-Evoked Response" (submitted for publication), 2011.

Massimini, M., Boly, M., **Casali, A. G.**, Rosanova, M. and Tononi, G., "A perturbational approach for evaluating the brain's capacity for consciousness". *Progress Brain Res.*, **177**: 201-14, 2009.

Pigorini, A., **Casali, A. G**., Casarotto, S., Ferrarelli, F., Baselli, G., Mariotti, M., Massimini, M. and Rosanova, M., "Time-frequency spectral analysis of TMS-evoked EEG oscillations by means of Hilbert-Huang transform", *J. Neurosci. Methods* **198**:236-45, 2011.

Rosanova, M., Gosseries, O., Casarotto, S., Boly, M., **Casali A. G**., Bruno, M., Mariotti, M., Boveroux, P., Tononi, G., Laureys, S. and Massimini, M., "Recovery of cortical effective connectivity and recovery of consciousness in vegetative patients". *Brain* (in press), 2012.

Rosanova, M., **Casali, A. G.**, Bellina, V., Resta, F., Mariotti, M. and Massimini, M., "Natural Frequencies of Human Corticothalamic Circuits". *J. Neurosci.*, **29**:7679-85, 2009.

"*One of the amazing characteristics of nature is the variety of interpretational schemes which are possible.*"

Richard Feynman

Chapter 1.

Introduction: Consciousness and the Brain

> "*Her mind still rising and falling with the sea, the taste and smell that places have after long absence possessing her, the candles wavering in her eyes, she had lost herself and gone under. It was a wonderful night, starlit; the waves sounded as they went upstairs; the moon surprised them, enormous, pale, as they passed the staircase window*." (Virginia Woolf, "To the Lighthouse")

All our experience occurs while we are aware. What we think, desire, feel, dream, imagine, believe and doubt, all objects, properties or feelings, only have meaning for us as we are able to refer to them consciously. Consciousness is, as stated by Nagel (1974), "*what it is like to be something*" and since we can not be anything out of it, it is no wonder that the question of its nature has covered the entire history of Western philosophy, is in the center of religion and art, and still remains one of the greatest mysteries of mankind.

For the nature of consciousness, a subject previously considered outside the scope of empirical science and relegated to purely speculative areas of human thought, today is the center of a discussion involving the medical diagnosis of extreme conditions and ethical issues that deal with the treatment, life and death of patients suffering from disorders of consciousness (DOC). Chances of recovery for critically ill patients, diagnosed in the vegetative state (VS), are close to zero three months after nontraumatic brain damage or one year after a traumatic lesion (Royal College of Physicians, 1994) and, in these cases, withhold or withdraw of life support may be ethically justified (Andrews, 2004; Laureys and Boly, 2007). However, the diagnosis of such severe conditions is based on clinical behavior that rely heavily on subjective interpretations: VS, characterized by the return of arousal with no

evidence of awareness, is diagnosed when the patient, who is unable to communicate functionally, has no voluntary movements but only reflexive or automatic ones. On the contrary, patients who are still incapable of functional communication but show signs of voluntary, reproducible movements, even though inconsistent, meet the criteria for the minimally conscious state (MCS) (Giacino et al., 2002). The prognosis of MCS patients is significantly more favorable than VS patients (Giacino, 2004) and currently no consensus on the duration of this condition has been established in support of end-of-life decisions. Therefore, such important and difficult decisions, as well as other issues related to quality of life of brain-injured patients (Boly et al., 2008), end up being based on the subtle distinction, from the clinical point of view, between what can be considered as a voluntary movement and what should be classified as an involuntary reflexive movement.

There is currently no consensus or convincing scientific evidence allowing for the establishment of methods to differentiate, in an unambiguous and objective way, automatic movements from those that show signs of consciousness. For instance, in a recent study, MCS patients with visual pursuit did not track the moving objects frequently used in routine examinations, but presented unambiguous signals of visual tracking when exposed to moving mirrors (Vanhaudenhuyse et al., 2008). Neurobehavioral rating scales such as the Coma Recovery Scale-Revised (CRS-R) were developed specifically to better differentiate MCS from VS (Giacino et al., 2004) and recent studies comparing the clinical consensus diagnosis to diagnoses derived from rating scales assessments, performed by research staff, indicate that the rate of misdiagnosis of VS has not substantially changed in the last decade: more than 40% of patients diagnosed with VS based on the clinical consensus were found to be in MCS following standardized assessments (Schnakers, et al., 2009; Andrews et al.,1996). In fact, well-documented cases prove that if voluntary behavior, when present, is sufficient for the inference that the person is conscious, its absence, by contrast, does not necessarily imply unconsciousness (Boly et al., 2007; Laureys et al., 2004; Monti et al., 2009; Owen and Coleman, 2008). Motor impairments, as observed in the locked-in syndrome (LIS) where patients may be fully aware but unable to respond (Laureys et al., 2004; Schnakers et al., 2009b), fluctuating arousal level, or even ambiguous and rapidly habituating responses may impair behavioural assessment in brain-injured patients (Gill-Thwaites, 2006; Giacino et al., 2009).

For these reasons, electrophysiological (Kotchoubey et al., 2005, Fellinger et al., 2011) and neuroimaging protocols (Owen and Coleman, 2008; Monti et al., 2009) have been developed to improve diagnosis and probe for signs of awareness. Studies of functional magnetic resonance imaging (fMRI) have shown evidence of conscious perception in patients diagnosed as being in the vegetative state, when they were exposed to sensory stimuli in a hierarchical order, beginning with the simplest form of stimulus processing and then progressing sequentially through more complex cognitive functions (Owen et al., 2006, Monti et al., 2010). In addition, electrophysiological studies recently demonstrated that the recovery of residual cognitive function in brain-injured patients can also be accessed by both early and late event-related potentials (ERPs) (Schnakers et al., 2008, Faugeras et al., 2011, Kotchoubey et al., 2005, Boly et al., 2011).

Although protocols employing fMRI and ERP have important clinical and scientific implications to the diagnosis and prognosis in brain-injured patients, the acquisition, analysis and interpretation of results are not without their difficulties. For instance, in particular circumstances, interpretation of fMRI activations may be prevented by alterations in the normal coupling of neuronal activity to the local haemodynamics in brain-injured patients (Hamzei et al. 2003; Rossini et al. 2004). In addition, late ERP components such as P3b can be absent in patients who show behavioral signs of consciousness (Holler et al., 2011; Monti et al., 2010; Bekinschtein et al., 2009; Bardin et al., 2011), and some studies suggest that early activity generated at pre-attentive level can not reliably distinguish conscious from unconsciousness in healthy subjects (Simpson et al., 2002) or MCS from VS in brain injured patients (Fischer et al., 2010). Finally, studies have shown that aspects of human cognition, including those related to semantic processing, can be present even in the absence of consciousness (Bonebakker et al., 1996, Dehaene et al., 1998). This may also prevent the correct interpretation of the results: a “normal” neural response observed in patients unable to communicate effectively does not necessarily imply evidence of awareness (Davis et al., 2007), and false negatives are very common even in healthy volunteers (Owen and Coleman, 2008; Bardin et al., 2011).

Apart from the difficulties in the application and interpretation of results, the ability to infer unambiguously the capacity for consciousness through the methods mentioned above is determined ultimately not by consciousness itself, but the awareness of a specific stimulus. Therefore, false negatives can also be obtained in patients who, even after partial recovery of

high-order cognitive functions, are not aware of the stimulus due either to a fluctuation in the level of consciousness or blockade of ascending pathways, but also by attention problems, given the complex paradigms required for unambiguous interpretation of positive results.

It is clear, therefore, that the diagnosis of brain-injured patients could benefit greatly from a better understanding of what is indeed relevant to consciousness. What is required, ultimately, is a scientific investigation of the bases of consciousness that gives evidence not only of the neural correlates of conscious perception of a particular stimulus, but mainly of the conditions necessary for supporting all and any conscious experience. In principle, it is expected that methods can be developed to investigate the brain's capacity for consciousness independently of the subject's willingness or ability to engage in assessment protocols, the integrity of sensory pathways or motor functions. Such methods could therefore avoid the problems of interpretation associated with both the relationship between specific brain activations and specific contents of consciousness, as well as between specific contents of consciousness and consciousness itself. Current methods depend on neural correlates of specific contents of consciousness only because, scientifically, consciousness itself is very poorly understood.

1.1 - Towards a Science of Consciousness

Should a proper scientific approach to consciousness start empirically in the absence of minimal guidance to what really matters for consciousness or from theoretical considerations that reduce consciousness to structures or properties of the brain?

1.1.1 – *Empirical Findings in the Absence of Theoretical Guidance*

Motivated by the important advances observed in techniques of neuroimaging and neurophysiological records, part of the research on the neural basis of consciousness is developed in a purely empirical way. It is hoped that monitoring of neural activation in the presence of sensory stimuli can reveal brain structures or properties that are essentially associated with the presence of consciousness itself. If, on the one hand, stimuli that remain unconscious are still able to affect the brain at different levels (Kouider and Dehaene, 2007;

Van den Bussche et al., 2009b), several brain processes and structures at high levels in the hierarchy of sensory information processing seem to be involved only in conscious perceptions. fMRI studies on visual perception have shown increased activation during conscious perception in higher areas such as fusiform gyrus (Haynes et al., 2005; Polonsky et al., 2000; Williams et al.,2008), and associative cortical regions such as the inferior parietal and prefrontal cortices (Dehaene et al., 2001b, Haynes et al., 2005b), while differences of the same magnitude were not observed in the primary areas. Several electro and magneto-encephalographic studies indicate a correlation of conscious visual perception with a late component (> 300 ms) of ERPs (Babiloni et al., 2006; Del Cul et al., 2007; Fernandez-Duque et al., 2003; Koivisto et al., 2008; Lamy et al., 2009) that seems to be associated with activation of areas such as hippocampus, parietal and frontal associative cortices (Halgren et al., 1998; Mantini et al., 2009). Similar results were obtained with other types of sensory stimulation, such as hearing (Bekinschtein et al., 2009a; Diekhof et al., 2009; Sadaghiani et al., 2009), where conscious awareness differed from unconscious by the presence of late potentials in associative areas. Conscious perceptions were also accompanied by increased synchronization of the phase of gamma oscillations between distant regions (Doesburg et al., 2009; Melloni et al., 2007; Rodriguez et al., 1999; Schurger et al., 2006; Wyart and Tallon-Baudry, 2009). Intracranial recordings during visual paradigms have confirmed these results and also indicated the presence of a marked increase of causal interaction between cortical areas during conscious perception (Gaillard et al., 2009).

These and other similar studies open up a new perspective in the investigation of the relationship between brain function and conscious perception. However, because conscious perceptions occur in a subject who is already conscious, attending exclusively to the neural correlates of conscious perceptions could not possibly give us all conditions for consciousness. Indeed, in the absence of a minimal guidance to what really matters for consciousness, advances towards the neural basis of the phenomenon of conscious experience could not but progress slowly and in ambiguous directions. Those aspects of brain activity suggested by these studies as measurable determinants of the level of consciousness, such as high levels of cortical depolarization with gamma band oscillations (Llinás et al., 1998), long-range synchronization of fast-frequency activity (Engel and Singer, 2001; Tallon-Baudry, 2004) and activation of a widespread fronto-parietal network (Dehaene et al., 2003) were not found to correlate in all cases with the level of consciousness. For instance, widespread cerebral responses to sensory stimulation were observed during anesthesia-induced loss of

consciousness (LOC) and non-rapid eye movement (NREM) sleep (Kakigi et al., 2003; Kroeger and Amzica, 2007). Gamma activity and synchrony was found low in rapid eye movement (REM) sleep - when subjective experience is usually vivid - and high in anaesthesia (Vanderwolf, 2000). Also, intracranial recordings show that gamma activity (Destexhe et al., 2007) and gamma-coherence (Bullock et al., 1995) persist during slow-wave sleep. Brain metabolism and fast rhythms can be high when consciousness is abolished by some anesthetic agents (Imas et al., 2005; Långsjö et al., 2005, Maksimow et al., 2006) or when consciousness is lost during the tonic phase of generalized tonic-clonic seizures (Blumenfeld, 2005; Blumenfeld et al., 2003). Together, these results support the claim that empirical progress in identifying the neural correlates of consciousness (NCC) can not dispense with theoretical considerations but should be guided by what is really relevant to consciousness itself.

1.1.2 – *Towards a Theory of Consciousness?*

Aiming to establish, from first principles or based on empirical findings, a neuroscientific theory of consciousness, there is a recent profusion of a variety of theoretical formulations, emerging with different assumptions and from different approaches to specific experimental results. One reads, for example, "*consciousness corresponds to the capacity to integrate information*" (Tononi, 2004); "*consciousness is the gateway to the brain*" (Baars, 2005), "*the global availability of information through a global neural workspace is what we subjectively experience as the conscious state*" (Dehaene and Naccache, 2001); "*a subjective state is a transient neuronal assembly of varying degree, recruited by the activation of a hard-wired hub, or network of neurons*" (Greenfield and Collins, 2005). Also, necessary and sufficient conditions for consciousness are affirmed to be "*coherent semisynchronous oscillations, probably in the 40-70 Hz range*" (Crick and Koch, 1990); "*the formation of dynamic links mediated by synchrony over multiple frequency bands*" (Varela, 2001); "*a large cluster of neuronal groups that together constitute, on a time scale of hundreds of milliseconds, a unified neural process of high complexity - a 'dynamic core'*" (Tononi and Edelman, 1998); or even that they "*may be expressed by only a small set of neurons, in particular those that project from the back of cortex to those parts of the front of cortex that are not purely motor and that receive feedback from there*" (Crick and Koch, 2003).

Contrary to the fundamental importance of empirical research in understanding the role of the brain in the phenomenon of conscious experience, the relevance of these theoretical formulations for what is intended as a science of consciousness is questionable. These so-called "theories of consciousness" frequently start from the mistaken idea that in order to make science of consciousness and fight against the old dualistic perspective it is necessary, first of all, to **define** consciousness in terms of brain structural or functional characteristics. Thus, in the hurry to find a definition of consciousness, these formalisms often incur into argumentative fallacies, such as the confusion between ontological and causal reductions (Searle, 1987; 2004; Velmans, 2000). Moreover, even those propositions that start correctly by approaching the problem from some phenomenological considerations - as the "Information Integration Theory of Consciousness" (Tononi, 2008) - end ultimately restricted to partial results of such phenomenology, ignoring some well established aspects of intentionality and admitting highly controversial ones.

The skepticism that history and philosophy of science from the recent past have taught us about the possibility of finding *a priori* principles of nature, demands the careful search, wherever possible, for concrete empirical results before establishing the foundation for a science still in its infancy in its field of study. With respect to theoretical considerations of consciousness, it has long been known that reductionism, whether in terms of type-type materialism or of functionalism, carries the old burden of explaining what characterizes consciousness at the phenomenological level, i.e., the qualitative aspects of experience and the so called **intentionality** of consciousness: the capacity of the conscious mind to represent, to stand for things and states of affairs, to be directed towards things outside of it (Brentano, 1874). In particular, a scientific theory of consciousness must face the important distinction between syntax and semantics (Searle, 2004): there are two things going on in the conscious subject, one is the neural states of the brain when the subject thinks, the other is the meaning that is attached to these conscious states. And when we attend to the problem of meaning, aiming at a "Theory of Reference", we are faced with an "open texture" underlying notions such as "synonymy", "proof", "confirmation", where all attempts to formalize meaning seem to presuppose the notion of reference that should be explained by them (Putnam, 1988; 1999). Knowing the meaning of words depends on grasping the manner in which they are used, and this in turn can not be founded in definitions and logical axioms (Wittgenstein, 1953): the mind is not a computer (Edelman, 2006; Searle, 2004; Putnam, 1999). The nature of such peculiar relation to the objects of consciousness and the distinction of this "open" type of

representation from other formal, computer-like, representations are some of the central questions to be addressed by any theory that intends to be a theory of consciousness. It is not by just redefining consciousness in terms of information processing in the brain that one will be able to scientifically explain something else than information processing in the brain itself[1] (Gallagher and Zahavi, 2008).

The call for a hasty reductionism, even if misleading, is often seen as positive from a pragmatic point of view: one tends to accept a poor argument just because "it works". However, such theories of consciousness are characterized by the fact that the choice of a reductionist paradigm, with all its intrinsic charge, has not resulted in any real advance in the implementation of empirical measures correlated with consciousness. Indices of neural activation, proposed by such theoretical frameworks, where testable, have not had until now unambiguous experimental corroboration, or, in the majority of cases, are still far from being applicable to the human brain, despite all the advances in neuroimaging and signal processing techniques (Seth et al., 2006).

1.1.3 – *The Project of an Empirical Science of Consciousness*

We can then say that much of the research currently being developed in the context of a science of consciousness is divided into two approaches to the problem that share a common characteristic. On the one hand, the empirical scientist searches for NCCs but with no phenomenological guide for the investigation. Under the supposition that everything that is fundamentally relevant to consciousness is part of the subjective realm, accessible only to introspection and outside the scope of objective science, the scientist must be satisfied with a "groping blindly", with no regard for anything that relates to the phenomenology of conscious experience, which is left to philosophy or psychology. This restriction seems to leave no possibility of a proper science of consciousness. On the other hand there are those who recognize the importance of a theoretical foundation, without which science itself is not

[1] Quoting Hilary Putnam, "*the very idea of a theoretical identification presupposes that the concepts to be reduced are already under some kind of scientific control (recall the case of optics, or of thermodynamics). To introduce a set of concepts that at present figure in no laws* [i.e., the contents of consciousness]*, and then immediately to begin talking of searching for theoretical identifications of these "narrow contents" with computational states of the brain (which, as we noted earlier, also have not been defined, since we have the problem of what formalism is being envisaged when one talks of "computational states" here) is to engage in a fantasy of theoretical identification. It is to mistake a piece of science fiction for an outline of a scientific theory that it only remains for future research to fill in.*" (Putnam, 1997, p. 37).

possible, but agree that this theoretical possibility must be based on the definition of consciousness in terms reducible to some property or structure of the brain. The intention is to establish a theory of consciousness without having to consider the compatibility of its premises with the phenomenological properties of conscious experience. The objective measures proposed are, therefore, theoretically based not on what is usually meant by "consciousness", but in what theories themselves assume *a priori* that consciousness is.

In both cases, when consciousness is the object of study, it is often ignored that the most fundamental objective of science is precisely that of, from phenomenological considerations which set a reference to the object in question, search for the essential characteristics behind what is apparent. When Isaac Newton, in his "*Treatise of the reflections, refractions, inflections and colours of light*", searched for what is essentially part of the luminous phenomenon, he did not intend in any moment to ignore the features that were observed in the phenomenal experience of light and color. Newton did not describe the spectral decomposition of the light with eyes closed, fearing that a subjective character intrinsic to the experience of color could distort scientific objectivity. Similarly, Newton did not open the field for scientific developments in optics that followed by starting from a definition that would *a priori* reduce light to some other natural entity, without first carefully analyzing the possibility that this definition would explain the properties of the luminous phenomenon. Science could never ignore the appearances: it is exactly by trying to explain what appearances are that one can find what is essential in nature. Now, if conscious experience is a particularly difficult subject because it seems to be, at least at first sight, primarily related to phenomenological aspects (Zeeman, 2005), this is another reason for the need of including considerations of its phenomenal features in the scientific investigation of its basis.

But then, should we not have to base our investigation on pure introspection? And wouldn't attention to considerations of what it is like to be conscious represent an inescapable obstacle to the progress of science towards understanding what really matters for consciousness? The present thesis is the result of our efforts to respond negatively to both questions. Following a path that we believe has not yet been explored, we try here to advance toward an empirical science of consciousness, focusing on the full meaning of these words. A **science** because the object studied here will be treated according to the scientific method and tools, avoiding speculative arguments without theoretical or empirical foundation but based,

ultimately, only on the results of scientific experiments dealing with quantifiable objective properties. The investigation is said **empirical** because we will not propose here the development of a theoretical framework, based on *a priori* principles or on considerations still without experimental evidence, but an objective measure, based on empirical data, with the potential to be related to consciousness and that can be tested in various conditions of relevance. Finally, we can say that the ultimate object of this investigation is the **consciousness**, because our efforts toward an objective measure with the potential to assist in clinical diagnosis of DOC will not ignore what is widely described as that which characterizes the conscious experience. We will set a reference to our object of study exactly by focusing on the intentionality of consciousness, the fundamental directedness towards what is experienced: what it is like to be able to perceive something, to refer to something, to think about what we think.

For this purpose, it is imperative to establish a dialogue with other areas of knowledge. Throughout the twentieth century, philosophers have struggled to clarify several relevant aspects in the characterization of the concept of consciousness, developing a phenomenology that points to objective conditions dissociated from introspection that not only assume but require scientific research. Techniques of signal acquisition and processing were developed at the interface between neuroscience, engineering, physics and applied mathematics, and, with the emergence of new computational methods, new possibilities for acquisition and analysis of biological data have been opened. Finally, neurophysiology has advanced in the understanding of structures and mechanisms responsible for information processing in the brain. Our work emerges from this complex interface which is the natural environment for a scientific investigation of consciousness. By paying attention to simple and well established phenomenological considerations of what is involved in conscious experience, we will show that the natural assumption that the brain, especially the human thalamocortical system, must have at least a role of serving as a necessary means to access the content of consciousness, will lead us to the implementation of a noninvasive measure of neural activation with the potential to be correlated to the brain's capacity for consciousness and to assist the diagnosis and prognosis in brain-injured patients.

1.2 – Consciousness: Brain, World and Causality

It is thus not possible to start our work by stating the ultimate nature of conscious experience. We do not know if consciousness is made of some strange substance, a *res cogitans*, as would Descartes (1641), or a "nonmaterial entity", as suggested by Eccles (1980), maybe an exotic holistic property, as imagined by Sperry (1985), possibly emerging from the brain, as would Libet (1996); or if it really has nothing of strange in it, being only a "*motion in the brain*" as stated by Hobbes (1651), a type of localized memory as proposed by Dennett (1991); or just another brain function, a "*focal attention*" as suggested by James (1890), being perhaps nothing more than a specific way of processing information that can be transferred to a silicon brain. We can not even say whether everything that is relevant to consciousness is located within the human mind, as already suggested by Aristotle, or whether conscious experience is fundamentally a type of interaction with the world that points away from us, as advocated by Heidegger (1927).

Nevertheless, even without a complete definition, we all know what is referred to when the word "consciousness" is used. We say that a person is conscious when, for example, she/he is awake, in a special state of attention, able to perceive and react properly to objects in the world, capable of formulating thoughts and using words correctly to report such objects, sensations and thoughts to other conscious subjects. On the other hand, we say that a person is unconscious when, for example, she/he is in a deep sleep, a state in which one is not able to react properly to objects that are presented, nor shows signs of formulating thoughts. Consciousness is a state of "awareness" in which things are experimented, in which we are directed toward things: we think of something, wish something, imagine something, believe in something. Without a thought content, there is no thought, without a experienced content there is no experience: all consciousness is consciousness of something. Therefore, when we say that a subject has conscious experiences, as opposed to automatic unconscious affections, we are affirming that there is something that it is like **for** the subject, her/his peculiar point of view, full of **intentional content**: the things thought, perceived, the facts remembered, the images dreamt, others just imagined or believed; they are all objects **for**, "represented by" or "revealed to", the conscious subject.

It is natural to assume, from the perspective of contemporary neuroscience, that if something allows us the access to such contents thought, desired, imagined, pursued, if something is responsible for the type of interaction that occurs between the subject and the

object of consciousness, this vehicle to the content of experience must depend on the brain. Whatever the ultimate nature of conscious experience may be, to the extent that consciousness is related to a content, the brain, the human organ that gives access to the objects of the world outside us and is the actual residence of our memories and feelings, must have a key role in supporting our experience.

If we want to characterize the brain's role in supporting conscious experience we should then ask for the properties of the intentional objects present to consciousness. What is the content of our experiences? The precise nature of the intentional objects, whether constituted exclusively by internal representations or fundamentally by external real objects, is a matter of intense contemporary debate (Menary, 2010; Rowlands, 2010; Searle, 2004; Putnam, 1999). Nevertheless, it will be enough for our purposes to consider two well-accepted and long described phenomenological aspects of the contents of consciousness. When we think of something, when we perceive something, the object of our awareness is at once a unity and a multiplicity (Kant, 1787; Husserl, 1913; Merleau-Ponty, 1945): in thought, imagination, perception things and facts are unveiled to us, in each moment, as units composed of a multiplicity of features. Being aware is to be directed towards a world of colors, sounds, tastes, smells, memories, emotions, all united under one scheme. Our experiences take place as in the fragment of Virginia Woolf's novel, in which the sense of that "*wonderful night, starlit*" is inseparable from Lily Briscoe's memories, from the "*enormous, pale*" moon she sees in the sky, the sound of the waves she hears, her steps going upstairs, and all the emotional content of her mind that rises and falls with the sea.

1.2.1 – *A Manifold of Features and the Brain's Functional Specialization*

If experience and cognition take place through a manifold of simple features - texture, position in time and space, images, sounds, smells, tastes, sensations such as heat, cold, pain and pleasure - if we are able to access the objects of intentionality through their multiple features, then this multiplicity is to be given to us as such by independent sensitive modalities[2]. Now, if we can access these different modalities through the brain, then the brain

[2] This is true even if one assumes a "natural realistic" scenario (Putnam, 1999), where "sensations" (or the so called "secondary qualities", such as color) are considered as a set of *idealized abstractions* from the perceived **real objects** (see also James, 1912; Austin, 1962), instead of a set of "properties" of the **mental representations** which are causally (but not cognitively) related to the perceived objects (Russell, 1914). Whether considered as a passive affection or an active ability of the mind, conscious perception unveil to us objects (inside the mind, or

must somehow receive them, or process them, independently. Actually, since the 50s it is known, mainly from anatomical considerations and studies of cortical lesions that cause specific problems of perception, that each sensory modality in the brain is mediated by a distinct sensory system with a largely independent processing: a well established principle of the organization of the brain is that it is structured in different specialized functional modules (Kolb and Whishaw, 1990). For example, while the area responsible for processing somatosensory information is located in the postcentral gyrus, the visual cortex is located in the occipital lobe near the calcarine fissure, the auditory cortex near the Heschl's gyrus, the emotion content of perception is processed by the Amygdala, and the Hippocampus, in the temporal horn of the lateral ventricle, is responsible for the storing of information in long-term memory.

When we focus on the processing within a single specific sense modalility, we observe a further specialization: sensory and motor areas are functionally divided into primary, secondary and tertiary regions, depending on the distance, in terms of information processing, to the peripheral sensory and motor pathways. A complete retinotopic mapping, for example, is present in the primary area (V1) of the visual cortex, with individual cells firing selectively to color, contours and contrasts; in area V2, there are cells sensitive to illusory contours, and others capable of coding for differences such as those that determine if the image belongs to the main figure or to the background (Qiu and von der Heydt, 2005); information such as color and orientation of objects are decisive in the firing of some V4 cells, while specific neurons in area V5 are sensitive to movement in specific directions. Such functional specialization may be seen even within a single primary sensory region. The primary somatosensory cortex, for example, has topographic maps of the skin in four Brodmann's areas (BAs): 3a, 3b, 1 and 2. Basic processing of tactile information takes place in area 3, while in areas 1 and 2 more complex processing is associated with neurons with larger and more stable receptive fields. In area 2, for example, there are neurons that respond by integrating information from tactile sensors and sensors sensitive to the position of the members.

Recent studies using fMRI in natural conditions, while subjects watched a movie, have confirmed this functional specialization of the brain (Bartels and Zeki, 2004; 2005). It was observed that different areas, even within the visual cortex, responded to stimuli with

outside in the world) that are composed of a multiplicity of independent aspects, such that a conscious brain must be capable to individualize them in cognition.

different time courses. The activity time-course in each area could serve as a temporal fingerprint to be used to identify specialized cortical subdivisions, which corresponded anatomically across subjects with highly area-specific inter-subject correlations. In this scenario, our access to the multiplicity of experience is made possible by the brain through a series of parallel relays in which the information related to each feature such as color, motion, sound or smell is processed independently and simultaneously by functional modules.

1.2.2 – *The Unity of Experience and Integration in the Brain*

But if it is through such parallel processing that the brain begins to process what is received through the senses, it is also clear that our conscious experience is not formed only by a pure fragmented multiplicity of sensations without any unity. Instead, in any experience, however simple it may be, all the features that initially appear to be processed in parallel are integrated into a whole that we associate with the content of consciousness. If, by hypothesis, the brain is our means of access to such content, it is required therefore that somehow the information contained in primary sensory areas is combined and processed to be integrated to form a multi-modal unit. Indeed, a second organizational principle of the brain, responsible in part for such synthesis of modalities, is the hierarchical structure of its functional systems in stages of information processing (Kandel et al., 2000). It is well known, for instance, that neurons in the lateral geniculate nucleus of the thalamus responds to light in specific regions of the visual field. These cells converge on cells in the primary visual cortex which in turn fire more selectively, only when a specific combination of thalamic inputs is active. Information about color, shape, and depth are added to processing in subsequent steps until, in areas of higher-level processing, neurons are able to respond to highly complex information, such as faces of familiar persons (Quiroga et al., 2005). A similar process is observed in the other sensory areas, so that each primary area serves as the input for higher-order adjacent areas, where information is processed and refined. The higher the processing level, the more abstract is the encoded information.

This abstraction follows from unimodal association areas, up to a multi-modal integration. Several studies of lesions in the frontal, parietal and temporal lobes indicate that specific regions of the cortex are related to specific cognitive functions that depend on information from different sensory and motor modalities, including memory and emotional

centers (Etcoff et al., 1991; Damasio, 1994; Khan et al., 2011; Vallar, 2007). The limbic association area, in the medial edge of the cerebral hemisphere, integrates emotion and memory; regions of the angular gyrus are related to language and integrate both visual and somatosensory inputs; the posterior association area, between parietal, temporal and occipital lobes, integrates information from several sensory modalities enabling the perception of complex objects. Finally, association areas in the frontal lobe integrate information from both sensory and motor areas as well as from the remaining association cortex and are responsible for complex tasks such as contextual assessment and planning of the most appropriate behavior for a specific circumstance.

We can then form a first image of how the content of experience is made available to us by the brain. A particular object gives us a series of sensations, each one related to a specific feature. Information about these features is processed in the primary sensory areas, and then follow in "bottom-up" circuits to areas that refine it and integrate it: the farther this information is from the pure multiplicity present in the initial sense-data, the closer it is to what we experience as the content of our thinking, an object or integrated scene composed of a multitude of features.

This bottom-up hierarchy, however, could hardly be enough to provide us with the entire contents of our experiences. To be conscious is to be immersed in a world of content, where things perceived, thought, remembered are revealed as unified objects. This immersion, however, is not just a passive reception, but always takes place as a type of engagement. Phenomenology has long discussed the intentional character of experiences and cognition: when we are conscious we are **directed towards** something or some state of affairs. The world - sensuous, remembered, imagined, thought - does not just affect us, it is **represented** by us or **disclosed** to us: it has **meaning**. In particular, when we are aware of an object, we may have access not only to the object itself, but also to what it is like to be aware of it (Nagel, 1974). We are therefore "engaged" in our thoughts and perceptions, and this intentional engagement is, arguably, what characterizes conscious perception at the phenomenological level, distinguishing it from unconscious affection. Through this intentional character of consciousness, the world revealed to us is a dynamic world in which thoughts and perceptions follow each other, influencing each other reciprocally in a temporal unit (James, 1912). Something we see makes us think or imagine something else, which in turn brings us memories and emotions, in a way that these lead us to make plans and guide our actions, and can even change the way we perceive the world: we guide perception with

action and attention, we color the world with our emotions. And so the experience seems to take place through an intricate network of connections between the conscious subject and the objects of consciousness. In this scenario, it is difficult to imagine how the entanglement required by intentionality, this directedness towards its contents, can be made possible only through bottom-up thalamocortical circuits. The complete unit that we experience in consciousness seems to require that what is at the top of hierarchy of processing, which are supposed to refer to the things we think or perceive, can affect the perception and thinking in successive moments, allowing our access to the "*what-it-is-like-ness*" of its awareness. Moreover, even if it was possible to sustain this intricate relationship with a processing structure exclusively of the bottom-up type, certainly a restriction to such feedforward circuits would be a much more expensive way to program the interaction with the world. It is natural to assume, instead, that those cells of the primary areas of the cortex are not used only as relays of sensory maps, sending information to higher stages, but that they can also be called again by these same higher stages, so that the information processing required for experience can be more effective with a smaller number of basic units. These feedback loops would then be required to form the object of consciousness precisely for making available the means by which the above can influence what is below in the level of the hierarchy of information processing, and this would be related to the intentional unit that we observe in the object of consciousness.

Anatomically, it is well known that sensory areas such as the visual cortex have a massive network of feedback loops (Salin and Bullier, 1995). Only recently, however, the key role of such circuits in visual perception has been shown. Studies of reversible inactivation of area V5 in monkeys demonstrated that cells from lower areas in the hierarchy of processing of visual information (V1, V2 and V3) are substantially reduced in response to visual stimuli when upper areas are temporarily inactive (Hupé et al., 1998). The feedback from upper areas of the visual cortex to primary areas would facilitate the identification and response to moving objects in the receptive field, increasing the removal of what belongs to the background stimulus. Further evidence of the role of such circuits in perception comes from studies of visual tasks in which longer delays between stimulus and response of primary cells are obtained. For example, during tasks of visual search, neurons in the primary cortex of monkeys fire with greater intensity if the object they search for is in their receptive fields. These more intense responses are however delayed approximately 200 ms with respect to initial activations to the onset of the search display, suggesting a feedback action (Gottlieb et

al., 1998; Motter, 1994) and indicating that backward connections are necessary for recurrent interactions among levels of cortical hierarchies (Garrido et al., 2007). Thus, cells from primary areas function as simple detectors of specific sensory aspects when they are in the early stages of perception (<80 ms), but in longer latencies they participate, through feedback connections from higher areas, of more complex processing systems involved in experience (Harrison et al., 2007; Huang et al. 2007; Lamme e Roelfsema, 2000).

Some studies suggest that the mechanism of such feedback loops can extend beyond the corticocortical circuits, involving the thalamus in re-entry processes (Guillery and Sherman, 2002). So, higher-level sensory mechanisms are seen modulating the thalamic circuitry in ways that optimize abstraction of a meaningful representation of the external world, in a scenario in which conscious perception involve a dynamic interplay between the thalamus, lower and higher-order cortical areas (Cudeiro and Silito, 2006; Silito et al. 2006). Several other studies show the importance of feedback circuits for conscious perception, influencing attention in visual (Moran and Desimone, 1985) and auditory (Winkowski and Knudsen, 2006) processing, as well as memory retrieval (Tomita et al., 1999) and the ability of the brain to provide objects of the imagination (Kreiman et al., 2000) for both visual (Guariglia et al., 1993; Kosslyn et al., 1999; Miyashita, 1995) and musical imagery (Kraemer et al., 2005).

1.2.3 – *The Content of Consciousness and Causal Interactions in the Brain*

The above results indicate that the brain has a complex hierarchical structure of information processing in which specialized modules are causally connected by feedforward and feedback circuits, including via corticothalamocortical re-entries. Such brain organization also appears to be required by the phenomenological observations on the characteristics of the intentional content: a multitude of features integrated into composite units, influencing our perception, guiding our action, and towards which we are directed. From these considerations, then, emerges a scenario of how the brain may sustain our access to the objects of consciousness. When consciously perceived, objects in the world should be able to cause a cascade of interactions involving the thalamus and the cortex, allowing neural activity to travel causally between several specific cortical areas: from the thalamus to primary sensory regions and from these to unimodal associative areas, moving in feedforward into the

multimodal associative cortex and from both associative areas, by feedback circuits, to sensory areas, also re-entering the system via complex corticothalamocortical projections. In this scenario, the set of causal interactions between the various specialized areas of the thalamocortical system appears as a mechanism that would allow us to access the contents of intentionality. We are then naturally led to the idea that, whether our consciousness is something that depends only on internal brain functions, whether it is an intricate way of interacting with the external world, the way to access to the contents of consciousness, without which the experience itself is empty - without which, therefore, there is no experience - depends on a certain capacity of the brain: the capacity of the thalamocortical system to sustain this complex network of causal interactions.

Such interactions clearly coordinate opposing forces: specialized functional modules with independent activities coexist with a mutual integration that tends to inhibit such independence. If the contents of consciousness are disclosed to us from this balance of opposing forces, it is natural to assume that this takes place in a situation of dynamic equilibrium. Integration can not be so strong that it could inhibit the independent processing of multiple features, since in this case the object of consciousness as an integrated **multiple** would become unavailable; on the other hand, this integration can not be so weak that the **unity** of the object is lost, leaving only a fragmented multiple. Thus, since this dynamic balance required by the content of that we experience certainly depends on neurophysiological properties of the thalamocortical system, the investigation of the ability of such a system to sustain this dynamic equilibrium of causal activations emerges, in this scenario, as a possible way to probe what could be called the brain's capacity for consciousness.

1.3 – TMS/hd-EEG: a Window to Causality in the Brain

An immediate limitation imposed on techniques that aim to reveal the nature of such causal interactions, hypothesized to be involved in the conscious processing of information, is the temporal resolution of the corresponding neural activation record. The feedforward sweep of visual perception, for example, is estimated to be completed in about 80-100 ms (Nowak and Bullier, 1998; Schmolesky et al., 1998), while recurrent connections operate up to a latency of around 300 ms (Lamme and Roelfsema, 2000; Roelfsema et al., 1998). Therefore,

under the demand of a time scale of a few milliseconds, electroencephalography (EEG) and magnetoencephalography (MEG) appear as the best methods available to explore, in a non-invasive way, the causal processes involved in consciousness.

Some conceivable electrophysiological measures of causality have been proposed in connection with theoretical perspectives of consciousness (Seth et al., 2006). Tononi and Edelman (1998), for example, focused on the relevance of causal processes to consciousness, such as those involved in thalamocortical re-entry, and introduced a measurement of complexity which is sensitive to the balance between the degrees of segregation and integration of a physical system. Neural Complexity, as it is called, is based on the amount of information shared by the parts of the system. It is calculated by observing the spontaneous activity of parts of the system and estimating the sum of the average mutual information across all bipartitions of the system. In addition to depend on the calculation of all bipartitions of the system, a task computationally prohibitive for large networks, this measure is symmetric between the parts and therefore unable to capture the causal link between brain areas, but only their statistical correlations. Seth's suggestion to introduce the required asymmetry for estimating causality is to apply Granger's statistical measure to the set of temporal series of the system (Seth, 2005). Granger Causality infers causality between signals from the ability to predict a signal from another one: if a signal causes a second signal, then past values of the first should contain information that helps predict the second, above and beyond the information contained in past values of the second signal alone. This measure, however, is also difficult to apply directly to the brain, since Granger causality depends on the correct estimation of a multivariate regression model of the thalamocortical system and the number of parameters that should be estimated in these models grows in proportion to the square of the number of elements in the network. In addition to this difficulty of implementation, the application of Granger Causality is not able to overcome the general problem of inferring causality from the spontaneous activity of a system. What we actually observe with the measure of Granger is the ability to statistically predict a signal in a future time from another signal in a previous time. However, if the fact that one signal causes another implies the ability to better predict the second signal by using the information contained in the first one, the reverse is not true: we can not always infer causality from statistical prediction. For example, applying Granger Causality to the time-varying, complex correlations among retinal neurons that are responding to a rich visual scene may lead one to the conclusion that activity of some retina neurons is causally determined by the activity of

other retina neurons. However, it is enough to perturb a few retinal elements and to record from the rest of the cells to realize that, to a large extent, the retina is actually composed of segregated modules that do not interact with each other and the activity that seems to be due to a causal influence of a part of the system over another one may actually be caused by a single common input external to the system.

How should we then proceed to measure causal interactions in thalamocortical system? In general, inferring the causal influence of one system on another depends on more than just a passive observation of the activity of the parties involved, requiring also an adequate control of other environmental variables, presumably relevant to the dynamics of the systems studied. This control is usually obtained by a perturbational approach. Let A and B be two systems immersed in an environment E. If one induces a controlled activity in B, keeping E invariant, and observes a subsequent significant response in A which is not seen when B is not stimulated, one have only then justified reasons to infer that B was able to causally influence A. Thus, the best way to access causal activation inside the thalamocortical system depends on the possibility of stimulating in a controlled way a specific area of the brain and to register the subsequent response of the rest of the system. Averaging out environmental variations by applying statistical measures to the response of the system, one could then safely infer causal influences of the perturbed area in the responding cortical regions. In order to accomplish this, one should employ a technique that directly stimulates the cerebral cortex, avoiding possible subcortical filtering or gating, and record the neural effects of the perturbation with the appropriate temporal resolution. Today, such a perturbational approach to explore causality in the thalamocortical system can be implemented non-invasively in humans (Paus, 2005), thanks to the development of a novel electrophysiological technique based on the combination of navigated transcranial magnetic stimulation (TMS) and high-density electroencephalography (hd-EEG).

1.3.1 – *Transcranial Magnetic Stimulation* (TMS)

TMS is the application of a short magnetic pulse to the cortex, resulting in the induction of an electric field on the surface of the cortex and subsequent neural activity or changes in resting potentials (Walsh and Pascual-Leone, 2003). In TMS, intense current (8 kA) resulting from the discharge of a capacitor passes through a coil, producing a strong

magnetic pulse (2-3 T) with a short duration (about 1 ms and a rise time of approximately 100-200 μs). A figure-eight coil-shape is normally used for its ability to produce a more focused magnetic pulse due to the sum of the magnetic fields produced by the reverse coils inside the target area and their cancellation outside it. The volume stimulated under a standard figure-eight-shape coil is estimated to comprise an area of up to 12 cm^2 on the cortical surface (20 mm below the coil), and to be reduced to zero within 2 to 3 cm of cortical depth (Barker, 1999). When the orientation of the electric field lines in relation to the axons forms a field gradient across the surface of the neuron, the induced current is able to activate the neuron (Ruohonen and Ilmoniemi, 1999). Through this type of mechanism, TMS induces neural activation mainly in interneurons parallel to the cortical surface or pyramidal neurons bended with respect to the coil's orthogonal direction, without being able to distinguish between inhibitory or excitatory neurons, nor between orthodromic and antidromic direction of stimulation.

TMS is used in studies that range from the cortical physiology and brain plasticity to therapeutic applications (Hallett, 2000). For our purposes, the importance of TMS is to allow for the direct stimulation, in a controlled and non-invasive way, of specific functional modules of the cortical surface. Several studies indicate the functional specificity of the primary effects of TMS. For example, TMS can produce muscle twitches when applied to the motor cortex (Barker et al., 1985), phosphenes when applied to the visual cortex (Kastner et al., 1998; Meyer et al., 1991) and speech arrest when applied to the frontal cortex (Pascual-Leone et al., 1991). Furthermore, studies combining TMS with positron emission tomography (PET) and fMRI reinforce the TMS focality by reporting that the neural activation induced by TMS is initially restricted only to areas located immediately below the coil, and secondarily spreads to those brain regions that are connected with the stimulation site (Paus et al., 1997; 2001, Siebner et al., 1998).

Also, as TMS stimulates the cerebral cortex directly, it can activate cortical neurons with a wide range of stimulation intensities, without being constrained by the physiology of peripheral receptors and nerves, providing full excitability profiles from threshold to saturation (Kähkönen et al., 2005b; 2005c; Komssi et al., 2004). Moreover, TMS does not depend on the integrity/status of sensory and motor systems and can be applied to any patient (de-afferentated, paralyzed, unconscious) and to any cortical area (primary and associative). Finally, by integrating TMS with MRI-guided infra-red navigation systems, which employs a 3D infrared Tracking Position Sensor Unit to map the positions of TMS coil and subject's

head within the reference space of individual structural MRI, it is also possible to render the perturbation controllable and reproducible, in most cortical regions.

1.3.2 – *The Combination of TMS with High-density Electroencephalography* (TMS/hd-EEG)

Above the direct stimulation of specific regions of the cortex in a non-invasive, reproducible and controlled way, the investigation of the patterns of causal activation in the thalamocortical system also depends on the possibility to record the response of this system to the stimulus in adequate time resolution. The development of multichannel TMS-compatible EEG amplifiers (Virtanen et al., 1999) has recently opened the possibility of combining TMS with hd-EEG, thus recording the electrical response of the human brain to the direct cortical stimulation with excellent temporal resolution (Iramina et al., 2003; Komssi and Kähkönen, 2006; Thut et al., 2005). The TMS-compatible EEG amplifier gates the TMS artefact and prevents saturation by means of a proprietary sample-and-hold circuit that keeps the analog output of the amplifier constant between 100 μs before and 2 ms after the stimulus (Virtanen et al., 1999). This device guarantees total absence of TMS-induced magnetic artefacts in most EEG recordings and artefact-free EEG recordings from 8 ms after stimulus, in all cases.

It is worth highlighting some of the specific advantages that TMS/hd-EEG may offer as a tool to probe the brain (Massimini et al., 2009):

1. TMS-evoked activations are intrinsically causal. Thus, unlike methods based on temporal correlations, TMS/hd-EEG immediately captures the ability of different elements of a system to affect each other.

2. TMS/hd-EEG bypasses sensory pathways and subcortical structures to probe directly the thalamocortical system. Therefore, unlike peripherally evoked potentials and evoked motor activations, TMS/hd-EEG does not depend on the integrity of sensory and motor systems and can access any patient (deafferentated or paralysed). Moreover, with TMS one can stimulate most cortical areas (including associative cortices) employing several different parameters (intensity, angle, current direction), thus probing a vast repertoire of possible responses, above and beyond observable ongoing brain states.

3. TMS-evoked potentials can be recorded with millisecond resolution, a time scale that is adequate to capture effective synaptic interactions among neurons.

4. TMS/hd-EEG does not require the subject to be involved in a task and the observed activations are not affected either by the willingness of the patient to participate or by his effort and performance. Hence, this approach is well suited to assess the objective capacity of thalamocortical circuits independently on behaviour.

5. TMS/hd-EEG can be made portable in order to overcome the logistical and economic hurdles that may separate severely brain-injured patients from advanced imaging facilities.

1. 4 - Plan of the Thesis

Phenomenological considerations on the nature of the intentional content of consciousness and the neurophysiological findings that suggest how this content is processed by the brain indicate that the support of conscious experience depends on the ability of the thalamocortical system to process information causally, integrating various specialized areas of the cortex at different moments. This study aims to apply a perturb-and-measure approach to the human thalamocortical system, through the combination of TMS and hd-EEG, to investigate the characteristics of these causal activations allegedly involved in conscious experience, for the introduction of a feasible empirical measure which allows for the discrimination between conscious and unconscious subjects.

A first step in this direction involves the development of standardized and data-driven procedures which are capable of describing the causal responses to TMS in the brain. In Chapter 2 we expose such methods to extract, from the raw EEG data, and passing through source activity reconstruction and non-parametric statistical analysis, a limited set of informative indices which are capable to quantitatively characterize different properties of the stimulated neuronal circuits, such as threshold for activation, cortico–cortical delays and patterns of neural connectivity.

Chapter 3 reports our results obtained from applying these methods in the study of the cortical activation evoked by TMS across subjects under different conditions: healthy subjects during alert wakefulness, slow-wave sleep, REM sleep and anesthesia; conscious locked-in patients and brain-injured, non-communicating patients who recovered from coma into different clinical states. In all cases, what fundamentally distinguished TMS-evoked responses when consciousness was unambiguously present from responses in conditions in

which consciousness was unambiguously reduced was an overall reduction of the complexity of the spatiotemporal cortical activation induced by the perturbation during loss of consciousness. Moreover, this spatiotemporal complexity consistently increased during the recovery of cognitive function underlying the evolution from vegetative state to minimally conscious state in brain-injured patients.

Motivated by these results, we introduce in Chapter 4 a feasible measure of the spatiotemporal complexity of the cortical activity evoked by TMS, called Perturbational Complexity Index (PCI). PCI is high only if many regions of the cerebral cortex react to the initial perturbation quickly and in a differentiated way. Remarkably, in a total of 116 TMS sessions collected from 19 healthy subjects and 17 brain injured patients, we invariably found high PCI values in conditions in which consciousness was unambiguously present (alert wakefulness in healthy subjects and locked-in patients) and low PCI values in conditions in which consciousness was clearly reduced (sleep, anaesthesia and vegetative state). This difference allowed for the reliable and robust discrimination between conscious and unconscious healthy subjects, producing disjoint distributions that were independent of the stimulation parameters, the strength and the extent of the cortical activation. Moreover, when PCI was measured in brain-injured patients who recovered consciousness from the vegetative state, it increased progressively in parallel with the level of consciousness.

Chapter 5 is devoted to a discussion of possible mechanisms underlying the loss of the thalamocortical capacity to sustain spatiotemporal complex patterns of causal activations during LOC. We will also discuss the limitations and advantages of PCI as well as its potential application toward developing a science of consciousness.

Chapter 2.
Methods on TMS/hd-EEG

The combination of transcranial magnetic stimulation with high-density electroencephalography is potentially capable of providing direct access to causal processes in the brain which are supposedly relevant to an empirical investigation of the nature of consciousness. In such a project, the application of TMS/hd-EEG to the thalamocortical system aims mainly at the study of neural activity that is not produced randomly, nor results from a common input ruling segregated functional modules, which is not a stereotyped response to stimulation and neither express mere temporal correlation, but that originates exclusively from the causal interaction between the parts of the neural circuitry involved. Therefore, the exploitation of this potential of TMS/hd-EEG depends on developing methods of analysis that, using the EEG signal produced in response to TMS, result in a quantitative description of neural activation arising exclusively from causal interactions between the parts of the thalamocortical system.

In this chapter, after describing the procedures involved in TMS/hd-EEG recording, we answer four questions that lead us to a standardized and data-driven signal processing method to qualify and quantify neural activation of causal origin within the thalamocortical system: 1) Does the EEG response to TMS reflect particular properties of stimulated circuits rather than stereotypical and/or random reactions? 2) How to extract a signal of neural activation from the EEG? 3) How to access the components of such a response that are produced exclusively by causal interactions in the stimulated circuits? 4) Can we quantify relevant characteristics of such circuits to describe the cortical activation evoked by TMS?

2.1 – TMS/hd-EEG Recording

We begin by describing the general procedure employed in the recording of the EEG potentials evoked by TMS, which will then be the object of our attention in this and in the following chapters.

2.1.1 – *TMS Targeting*

A Focal Bipulse 8-Coil (mean/outer winding diameter ca. 50/70 mm, biphasic pulse shape, pulse length ca. 280 μs, focal area of the stimulation hot spot 0.68 cm^2) driven by a Mobile Stimulator Unit (Eximia TMS Stimulator, Nexstim Ltd., Helsinki, Finland) was targeted to specific cortical locations using a Navigated Brain Stimulation (NBS) system (Nexstim Ltd., Helsinki, Finland). Structural MRI images at 1 mm^3 spatial resolution were acquired with a 1T Philips scanner from all subjects enrolled in the studies. The NBS system employs a 3D infrared Tracking Position Sensor Unit to map the positions of TMS coil and subject's head within the reference space of individual structural MRI in order to precisely identify the TMS stimulation target. Optimal alignment between MRI fiducials and digitized scalp landmarks (nasion, left and right tragus) was verified prior to all experiments. NBS also calculates on-line the distribution and intensity of the intracranial electric field induced by TMS, using a best-fitting spherical model of subjects' head and brain and taking into account the exact shape, 3D position and orientation of the coil. TMS hot spot (i.e. 98% of the maximum stimulating electric field calculated at individually-determined depth) was kept fixed to the stimulation target with the current perpendicular to its main axis. The reproducibility of the stimulation coordinates across sessions was guaranteed by an aiming device that indicated in real-time any deviation from the desired target greater than 3 mm.

2.1.2 – *hd-EEG Recording*

TMS-evoked potentials were recorded by a TMS-compatible 60-channel EEG amplifier (Nexstim Ltd., Helsinki, Finland). Impedance at all electrodes was kept below 5 kΩ. The EEG signals, referenced to an additional electrode on the forehead, were filtered (0.1-500 Hz) and sampled at 1,450 Hz with 16 bit resolution. At the end of each experiment, a pen

visible to the infrared camera was used to digitise the EEG electrode positions on the subject's head. Two extra electrodes were used to record the vertical electrooculogram (EOG).

During EEG recording, subject's perception of the clicks produced by TMS coil's discharge was eliminated by means of inserted earplugs continuously playing a masking noise (always below 90 dB). A thin layer of foam was placed between coil and scalp (resulting in less than 1 mm thickness when coil was pressed against the head) in order to attenuate bone conduction. As previously demonstrated, this procedure effectively prevented any contamination of EEG signals by auditory potentials elicited by TMS-associated clicks (Massimini et al., 2005; 2007).

2.1.3 – *EEG Pre-processing*

Trials containing activity from other sources than neural were automatically rejected if EOG exceeded 70 μV (ocular activity) and/ or absolute power of EEG channel F8 in the fast beta range (N25 Hz) exceeded 0.9 μV^2/Hz (van de Velde et al., 1998) (indicating activity of fronto-temporal muscles). After averaging, channels with bad signal quality or large residual artifacts were excluded from further analysis. In all TMS sessions we retained at least 50 good channels, with an inter-session variation of at most 4 channels in each subject, thus ensuring an estimation of cortical generators that was reliable and comparable across sessions (Laarne et al., 2000). Before source modeling, signals were low pass filtered (80 Hz), down-sampled to 362.5 Hz and re-referenced to the common average reference. Figure 2.1A exhibits an example of pre-processed single-subject TMS-evoked potentials after stimulation of BA19.

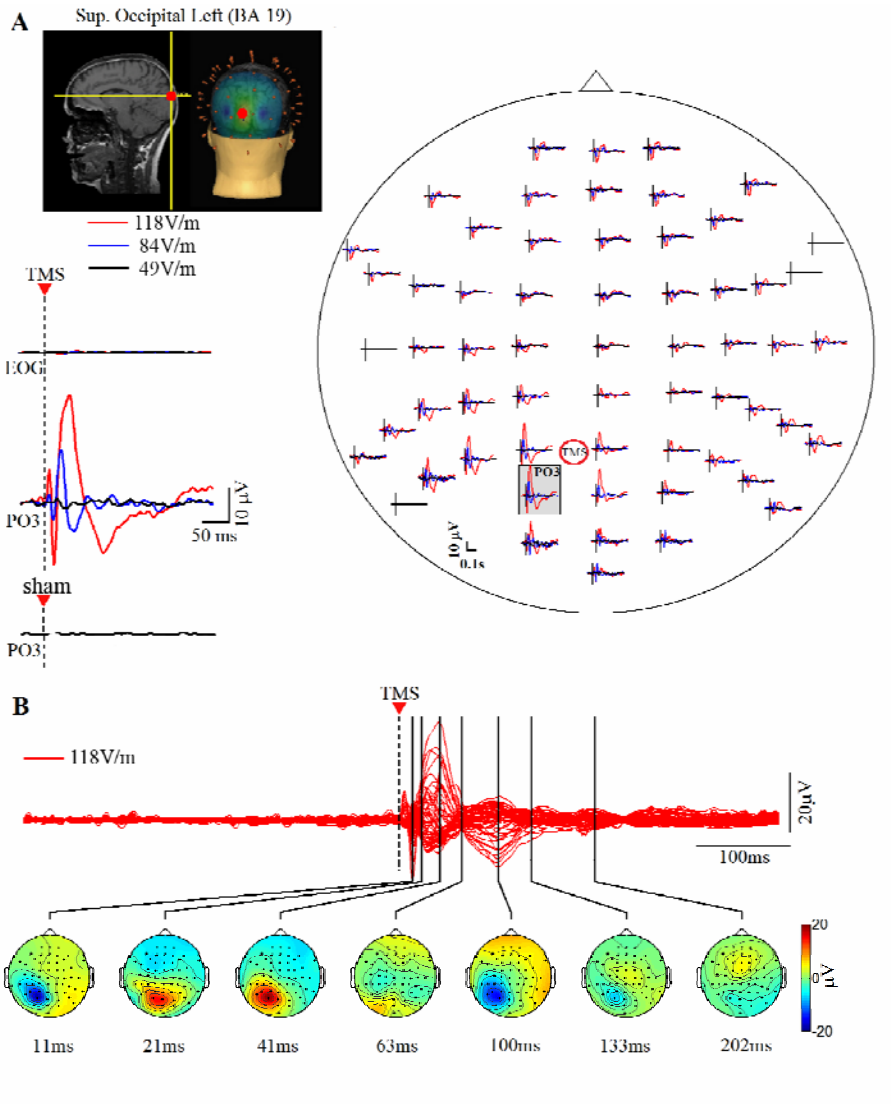

Figure 2.1: Single-subject TMS-evoked potentials. A) Top Left: Location of the TMS stimulation target (left superior occipital lobule, BA 19) is shown on individual MRI, together with the approximate distribution of the electric field induced by TMS pulses in the cortex. The orange spots represent the digitized positions of EEG electrodes on the scalp. Right: Averaged TMS-evoked potentials recorded from 60 electrodes on the scalp at three different intensities of stimulation (red 118V/m, blue 84V/m, black 49V/m). Approximate location of the TMS target is reported with a red circle. Bottom left: zoom on averaged EOG activity, TMS-evoked potentials recorded on PO3 lead, and also during a sham TMS session, when the TMS coil was discharged while separated from the scalp by a 4 cm Plexiglas cube. **B)** Superimposition of the averaged TMS-evoked potentials recorded from all channels (Butterfly plot) at a stimulation intensity of 118V/m. Color-coded instantaneous topographic distribution of the potential on the scalp is displayed for 7 relevant time samples after TMS pulse delivery.

2.2 – Does the response to TMS reflect particular properties of stimulated circuits rather than stereotypical and/or random reactions?

A first step to correctly interpret the nature of TMS-evoked potentials is to establish to what extent responses to TMS are non-random or specific when stimulation parameters and other environmental factors are kept constant, and to what extent they are non-stereotypical or sensitive to variations of stimulation parameters. Unlike sensory stimulation, TMS activates simultaneously a rather large cortical volume containing both inhibitory and excitatory fibers, possibly belonging to different functional subsystems. Thus, it is possible that different TMS perturbations may result in EEG responses that engage many different circuits and that are largely overlapping. In addition, TMS not only perturbs cortical neurons directly but may also activate the brain indirectly, due to the stimulation of scalp nerves and to the click sound associated with the coil's discharge, when unmasked, over the subject's head. For this reason, it is also conceivable that differences in the brain's reaction to different cortical perturbation may be partially obliterated by an invariant event-related potential triggered by an unwanted somatosensory or/and acoustic stimulation. Altogether, these factors may significantly hamper the sensitivity of TMS-evoked potentials. On the other hand, due to the complexity of the technique, TMS-evoked potentials may also lack specificity, by showing accidental changes related to stimulation/recordings errors. In fact, stimulating directly the cortical surface involves the control of several factors, since a large number of cortical locations can be arbitrarily selected and perturbed, each one with several stimulation parameters (e.g. intensity, time-course, and orientation of the magnetic field). Thus, a lack of precise control of these parameters across subsequent TMS/hd-EEG sessions may result in large measurement errors and in an apparent modulation of cortical responsiveness. Similarly, other factors, such as EEG sensors positioning, coil temperature, calibration of amplifiers, etc., may, if not adequately controlled, affect the specificity of TMS-evoked potentials.

Separate experimental evidences have suggested that TMS-evoked potentials have a certain degree of sensitivity to changes in stimulation parameters, such as location (Komssi et al., 2002; Kähkönen et al., 2004), intensity (Komssi et al., 2004; Kähkönen et al., 2005b) and direction of the induced current with respect to the cortical surface (Bonato et al., 2006). Moreover, a few works have demonstrated that TMS-evoked potentials can also detect changes in the state of cortical circuits, such as the ones induced by alcohol intake (Kähkönen, 2005), by falling asleep (Massimini et al., 2005; 2007) and by induction of

cortical potentiation with repetitive TMS (Esser at al., 2006). Specificity has been evaluated at the group level and the amplitude and latency of selected components of TMS-evoked potentials tend to be stable over time when stimulation parameters are constant (Lioumis et al., 2009).

But in actual fact, deciding whether a change in the EEG response to TMS is biologically relevant, or not, requires a systematic quantification of the trade-off between sensitivity and specificity. In Casarotto et al (2010) we addressed this question by performing a statistical joint evaluation of the sensitivity and specificity of TMS/hd-EEG measures. In order to test for sensitivity, subjects were submitted to different, randomly ordered, TMS sessions in the same day (day1), varying only one stimulation parameter at a time (either stimulation site, or stimulation intensity, or angle of the TMS-induced current). TMS-evoked potentials were considered sensitive to the extent that they changed when stimulation parameters were changed. In order to evaluate specificity, a subset of TMS sessions was repeated later in the same day (day1) as well as one week later (day8), without changing any stimulation parameter. TMS-evoked potentials were considered specific to the extent that they did not change over time when stimulation parameters were kept constant (thus, in the present context specificity is related to test-retest reproducibility).

2.2.1 – *Quantifying sensitivity and specificity: the Divergence Index (DI)*

By performing single-subject pairwise comparisons of the TMS-evoked potentials we could evaluate sensitivity when the sessions compared have different simulation parameters - “change comparisons” (C) - and specificity when they have identical stimulation parameters - “no change comparisons” (NC). In order to quantify the diversity of TMS-evoked responses in C and NC comparisons, we applied non-parametric statistics and computed the percentage of spatial-temporal samples that differed significantly between two sessions of TMS-evoked potentials (Divergence Index – DI). At first, a Wilcoxon rank-sum test was applied to check that the baselines (250 ms pre-stimulus) of the single trials, contributing to the two TMS-evoked potentials to be compared, had the same distribution. In case of a negative result, the most deviated trials were removed and the test was repeated until the baseline distributions of the two groups of trials were statistically equivalent ($p>0.05$). At this point, we could test the null hypothesis that two sets of TMS-evoked potentials are equivalent. If this is the case,

"mixing" together, in any random combination, the single trials collected during the two TMS/hd-EEG sessions should always result in the same TMS-evoked potential. Otherwise, the null hypothesis can be rejected. Thus, for each comparison, 1000 "mixed" TMS-evoked potentials were obtained by randomly mixing and averaging 1000 times the single trials collected in two different sessions (Figure 2.2 A, B). The set of 1000 values at each post-stimulus time sample represented the instantaneous empirical null probabilistic distribution of the voltage of the TMS-evoked potentials. In order to correct for multiple comparisons in time, we computed a single distribution for the whole time interval as follows: i) all instantaneous distributions were centralized around zero, by shifting them by an amount $\delta(t)$ (Figure 2.2C); ii) for each centralized distribution, we computed the maximum absolute value (Figure 2.2D); iii) the one-tail $(1-\alpha)100$th percentile of the distribution of the maximum absolute values was used to estimate a significance threshold G for the whole time window of interest (Figure 2.2D); iv) two boundaries were computed as $(+G+\delta(t))$ and $(-G+\delta(t))$. The temporal profile of these boundaries is modulated by $\delta(t)$, since G is a fixed threshold. The null hypothesis of equivalence between two TMS-evoked responses at each time sample t was rejected with probability of false positives α corrected for multiple comparisons when at least one of the two original potentials at that time sample lay beyond the significance threshold (Figure 2.2E). Finally, for each comparison the DI was defined as the percentage of significantly different time samples in the first 250 ms post-stimulus in all 60 EEG channels out of the total number of spatial-temporal samples. In this way, the DI was systematically calculated at the sensor level for all pair-wise comparisons (n= 92).

2.2.2 – *TMS-evoked potentials are sensitive and specific for changes in perturbation parameters*

We performed a total of 62 C comparisons with data obtained from ten right handed healthy volunteers enrolled into the study: 22 comparisons for stimulation site varying between BA6, BA7, and BA19; 20 comparison changing stimulation intensity from I% and I%+10% (I%, expressed as a percentage of maximum stimulator's output of the stimulator, was kept between 40-75% for all subjects, corresponding to an electric field between 110-120 V/m on the cortical surface); and 20 comparisons varying stimulation angle (10 comparisons 0° vs. 45° and 10 comparisons 0° vs. 90°). We also performed 3 NC comparisons for each

subject with identical sessions repeated in the same day (10 comparisons) and one week apart (20 comparisons).

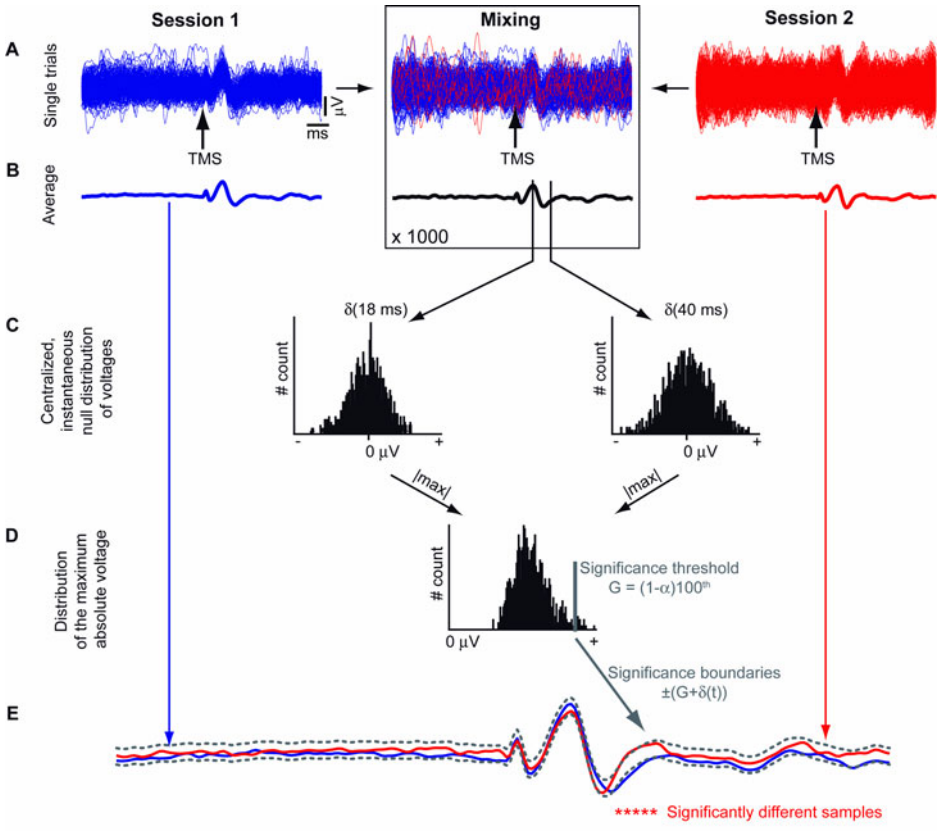

Figure 2.2: Non-parametric statistical procedure to perform single-subject pairwise comparisons between TMS-evoked potentials. Single-trial recordings from two different conditions (blue and red lines) were randomly mixed 1000 times (**A**) and averaged (**B**). Instantaneous distributions of averaged voltages were computed and centralized around zero by keeping record of the displacement δ(t) (**C**). The distribution of maximum absolute values of each centralized distribution was computed and used to define a significance threshold G as the $(1-\alpha)100^{th}$ percentile (**D**). Significance boundaries (gray dotted lines) were computed as (±G + δ(t)) and used to define the significantly different time samples (red stars) between conditions at a specific channel (**E**) (adapted from Casarotto et al., 2010).

Results of a representative subject are reported in Figure 2.3. While TMS-evoked scalp potentials and cortical currents tended to overlap in the NC comparison, they were clearly characterized by divergent spatiotemporal patterns in all the C comparisons, suggesting that the spatiotemporal characteristics of the brain response to a direct perturbation markedly depended on each and every stimulation parameter, e.g. site, intensity and angle.

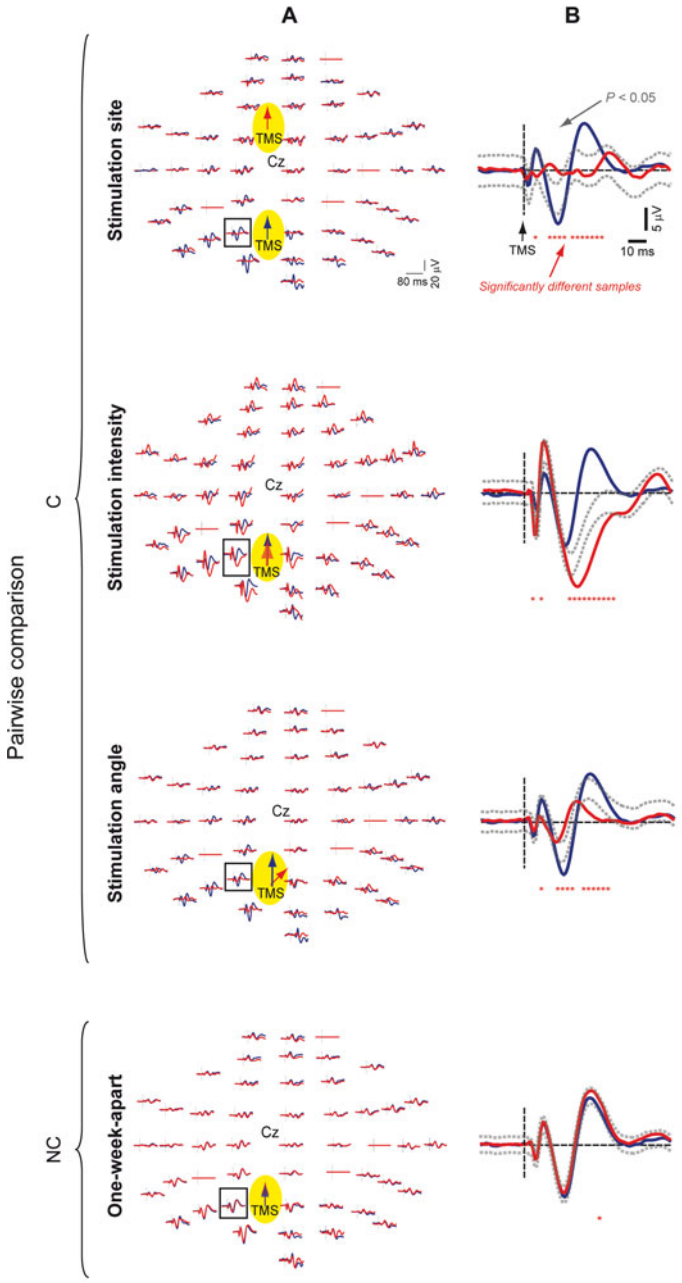

Figure 2.3: Results of pairwise comparisons between TMS-evoked potentials of a representative subject. Here, one particular TMS/hd-EEG session (stimulation of BA19 at I% intensity and 0° angle on day1) is taken as a reference (blue) and compared with four other sessions (red), where stimulation parameters are varied one at a time. Specifically, the site (BA19 vs. BA6), the intensity (I% vs. I%+10%), the angle (0 vs. 45°) and the day (day1 vs. day8) of stimulation were varied, resulting in three C comparisons and one NC comparison. For each comparison, superimposition of pairs of TMS-evoked potentials in all sensors is displayed in the panel **A**, while enlarged view of P1 channel is shown in the panel **B**, together with significance boundaries (dotted gray traces) and significantly different samples (red stars) (adapted from Casarotto et al., 2010).

Figure 2.4A summarizes the general results obtained from all subjects: each coloured dot represents the DI computed for a specific pairwise comparison.

Generally, comparing the brain responses evoked by TMS pulses delivered over different cortical sites revealed obvious differences in the space distribution and time course of voltages and currents, and the average DI of all 22 C comparisons between different stimulation sites was 11.45 ± 5.7% (range 3-19.7%).

Varying stimulation intensity resulted in amplitude and latency changes of the main TMS-evoked components, while the general topographical distribution of voltages and currents tended to be preserved. The average DI across the 20 C comparisons between different intensities of stimulation (Figure 2.4A, black dots) was 10.88 ± 6% (range: 2.31-22.2%).

When changing the angle of the TMS-induced current, the morphology of cortical responses varied in a rather unpredictable way. The average DI value was 3.92 ± 3% (range: 0.7-13.8%), with no systematic difference between 0° vs. 45° and 0° vs. 90° pairwise comparisons.

Finally, when TMS was applied with identical stimulation parameters at different times, the morphology and the spatial-temporal dynamics of TMS-evoked potentials were largely preserved. The average DI was 0.28±0.4% (range 0-1.2%) when comparing same-day recordings and 0.43±0.4% (range: 0-1.67%) for pairwise comparisons between one-week-apart sessions.

Fig 2.4B shows that the relative difference in DI values is preserved across latencies. For each type of pairwise comparison, the average DI computed at early latencies (0-60 ms) was significantly higher ($p < 0.05$) as compared to the one computed for late latencies (120-250 ms), although the relative differences among types of comparison were preserved across all time intervals ($p < 0.01$).

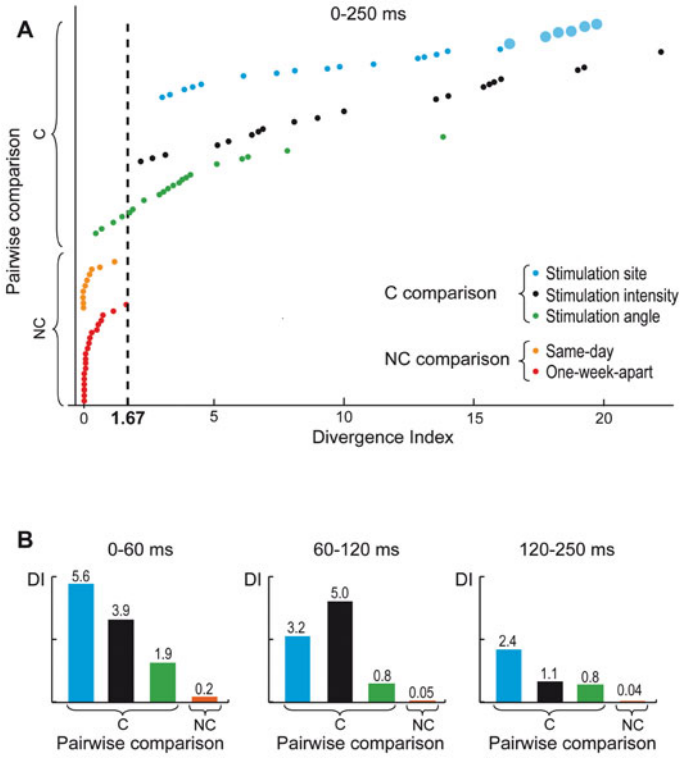

Figure 2.4: Divergence Index of all pairwise comparisons between TMS-evoked potentials. Single DI values computed over the entire post-stimulus period (250 ms) are shown (**A**) with the following color-coding: DIs of the C comparisons for changes in the stimulation site, intensity and angle are represented by cyan, black and green dots, respectively, while DIs of NC comparisons are depicted in yellow for same-day sessions and in red for one-week-apart sessions. DI values computed over different temporal windows of interest (0–60 ms, 60–120 ms, 120–250 ms) are reported in the panel **B** with the same color-coding, except for NC comparisons that are summed together and plotted in orange (adapted from Casarotto et al., 2010).

At this point, considering each DI as a threshold, we performed receiver operating characteristic (ROC) analysis and we computed the true positive rate or sensitivity rate (the number of C comparisons with DI > threshold) and the true negative rate or specificity rate (the number of NC comparisons with a DI < threshold). Plotting the ROC curve as sensitivity vs specificity, the optimal DI threshold was set in correspondence to the maximum of the Younden index (Younden, 1950), computed as [sensitivity + specificity - 1]. The optimal DI threshold according to Younden index was 1.67% and yielded a 95.1% sensitivity and 100% specificity, corresponding to an overall accuracy of 96.7%. The efficacy of DI in reliably quantifying the pairwise differences between TMS-evoked potentials was 99.1%.

These results confirm that TMS-evoked potentials, rather than being stereotypical or noisy responses, reflect, to a large extent, deterministic properties of the stimulated cortical circuits. Changing stimulation parameters almost invariably resulted in higher DIs compared to the no-change conditions. Importantly, this finding was not limited to the early latencies, and indeed DI values for the C conditions were significantly larger until 250 ms post-stimulus. This evidence demonstrates that the EEG response to TMS is primarily due to direct cortical stimulation and to the ensuing reverberation of activity in a specific network of connected elements. Indeed, changing the site of stimulation resulted in very different responses and in high DI values that were even higher when the responses triggered in areas located far away (area 19 vs. area 6) where compared. This variability in the cortical response reflects specific properties of the stimulated circuits and may be ascribed to local differences in cortical excitability (Kähkönen et al., 2005), to differences in the frequency tuning of corticothalamic modules (Kähkönen et al., 2005b) and to differences in the pattern of cortico-cortical connectivity (Ilmoniemi et al., 1997; Komssi et al., 2002).

Moreover, TMS-evoked potentials, besides being sensitive to changes, are also very stable over time. Repeating after one week a given perturbation and observing a DI >1.67% would strongly indicate that, in the mean time, some change occurred in the brain circuits. In principle, identifying a cut-off level above which one can decide whether a change in brain responsiveness occurred, or not, is crucial if one wants to use TMS/hd-EEG to track over time pathological alterations, plastic changes and therapy-induced modifications in cortical circuits

2.3 – How to extract a signal of neural activation from the EEG?

Postsynaptic potentials generated among the dendrites of pyramidal cortical neurons are believed to be the main contributors to EEG signal (Nunez and Silberstein, 2000). They typically last longer than the rapidly firing action potentials traveling along the axons of excited neurons and therefore are more likely to be synchronized. Also, the pyramidal cortical neurons are best fitted to contribute to the electric potential differences measured in the scalp because they have large dendritic trunks normally pointing perpendicularly to the cortical surface with similar spatial orientation.

The potential differences associated to these postsynaptic potentials generate intracellular post synaptic currents (called primary, or generator currents) as well as a loop of extracellular currents that flow through the volume conductor (called secondary, return, or volume currents). Both primary and secondary currents contribute to electric scalp potentials and localizing the primary electromagnetic sources of EEG activity is a major problem in any attempt to access neural activity generated exclusively by the causal effect of a neuronal group on another set of neurons. The solution of this so called **inverse problem** evolves a two-step procedure that eliminates the impact of the secondary currents flowing in the tissue. The first step takes into account geometrical and physical properties of the head in order to construct a model of the transmission of currents within the volume conductor called **the forward model**. The second step involves estimating the location of the primary currents in the cerebral cortex associated to the EEG signal.

2.3.1 – *The Forward Model*

Modeling the propagation of electromagnetic field through the different tissues of the head depends on the parametric representation of the sources and the model of the physical and geometrical properties of the head. Given that a vast class of complex currents can be expressed as sums of dipolar sources, the first task is normally achieved by modeling the primary currents as a set of equivalent dipoles representing the total activation of pyramidal cells in a given cortical region. As the electrophysiological signals typically have frequencies bellow 1kHz, the Maxwell's Equations that govern the dipolar approximation may be further simplified in the quasi-static limit. With these assumptions, the divergence free current density acting as source determine the voltage at scalp sensors by the following linear algebraic model (Baillet, 2001; Mosher et al. 1999):

$$\mathbf{v} = \mathbf{G} \cdot \mathbf{j} + \boldsymbol{\varepsilon} \tag{2.1}$$

where $\mathbf{v} = \mathbf{v}(\mathbf{x}_e,t,k)$ is the data vector containing the scalp potentials of a generic trial $k \in \{1, 2,\ldots, N_k\}$, recorded at time t from N_e electrodes located at spatial coordinates $\mathbf{x}_e$; $\mathbf{j} = \mathbf{j}(\mathbf{x}_j,t,k)$ is the corresponding source vector containing the instantaneous electrical activity of N_j dipole-like primary sources located at spatial coordinates $\mathbf{x}_j$ on a fixed grid, covering brain cortex; $\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}(\mathbf{x}_e,t,k)$ is an instantaneous additive measurement noise; and $\mathbf{G}=\mathbf{G}(\mathbf{x}_{e,},\mathbf{x}_j)$ is the

forward operator defining the electric potential propagation in head tissues and depends only on the geometrical and physical properties of the head and on the relative positions between sensors and sources.

The head model routinely used in most clinical and research applications assumes three concentric spheres with different homogeneous conductivity, representing the best-fitting spheres of inner skull, outer skull and scalp compartments extracted from individual MRIs (Zang, 1995). The solution to the field's equations determining the forward operator of the 3-spheres model is obtained by the truncation of an infinite series. A particular form to approximate the series which achieved popularity as a forward model in EEG analysis due to an equilibrated trade-off between computational resources, computational velocity and accuracy of the results is know as the BERG method (Berg and Scherg, 1994; Zang, 1995). We implemented the 3-spheres BERG method with the Brainstorm software package (freely available at: http://neuroimage.usc.edu/brainstorm), modeling the dipolar sources as a three-dimensional grid of 3004 fixed dipoles oriented normally to the cortical surface of a Montreal Neurological Institute (MNI) canonical mesh. This head model was adapted to the anatomy of each subject using the Statistical Parametric Mapping software package (SPM5, freely available at: http://www.fil.ion.bpmf.ac.uk/spm) as follows. Binary masks of skull and scalp obtained from individual MRIs were warped to the corresponding canonical meshes of the MNI atlas. Then, the inverse transformation was applied to the MNI canonical mesh of the cortex for approximating to real anatomy. Solutions to the forward operator were then obtained after EEG sensors and individual meshes were co-registered by rigid rotations and translations of digitized landmarks (nasion, left and right tragus).

2.3.2 – *Estimating Primary Currents*

Given a solution for the G operator, primary sources of EEG potentials can be obtained by the inverse solution of Eq. (2.1). However, this inverse problem is highly underdetermined ($N_j \gg N_e$) and mathematically ill-posed, since a virtually infinite number of source configurations can explain the same potential distribution on the scalp, unless specific anatomical and/or functional constraints (priors) are imposed to measurement noise and source configuration.

A straightforward constraint to the set of equations is to search, between all possible solutions, for the one that minimizes the deviations from the model, that is,

$$\hat{\mathbf{j}} = \underset{\mathbf{j}}{argmin}\left\{\left\|\mathbf{C}_e^{-1/2}(\mathbf{v}-\mathbf{G}\cdot\mathbf{j})\right\|^2\right\}, \tag{2.2}$$

where the measurement noise is assumed Gaussian with covariance $\mathbf{C}_e$, $\boldsymbol{\varepsilon} \sim N(0,\mathbf{C}_e)$.

However, this Minimal Norm approach can be too unrealistic and the classical method used in the inverse solution for MEG and EEG generalizes such a straightforward approach using the Tikhonov regularization (Tikhonov and Arsenin, 1977), minimizing a linear mixture of the residuals of data fit to the forward model and of a weighted norm of the current sources:

$$\hat{\mathbf{j}} = \underset{\mathbf{j}}{argmin}\left\{\left\|\mathbf{C}_e^{-1/2}(\mathbf{v}-\mathbf{G}\cdot\mathbf{j})\right\|^2 + \alpha\left\|\mathbf{H}\cdot\mathbf{j}\right\|^2\right\}, \tag{2.3}$$

where α is an hyperparameter expressing the relative importance of the two terms to be minimized and $\mathbf{H}$ is a regularizing operator defining some anatomical and/or functional constraints (priors) on the solution. The so called Weighted Minimum Norm (WMN) solution is given by:

$$\hat{\mathbf{j}} = \left[\mathbf{G}^{\mathbf{T}}\mathbf{C}_e^{-1}\mathbf{G} + \alpha\left(\mathbf{H}^{\mathbf{T}}\mathbf{H}\right)\right]^{-1}\mathbf{G}^{\mathbf{T}}\mathbf{C}_e^{-1}\cdot\mathbf{v} \tag{2.4}$$

The main drawback of WMN is its strictness. The solution is highly dependent on α and, although it is possible to obtain an optimal α through the inspection of a 2D plot from the first and second norms in Eq. (2.3) for all values of the hyperparameter, this can only be done for a single prior term (Brooks et al., 1999).

Alternatively, the inverse solution can be seen as an inference problem: the inference of the causes (current sources $\mathbf{j}$) from the consequences (scalp potentials $\mathbf{v}$). In the presence of noise, $\mathbf{v}$ can be considered as resulting from a stochastic process where the probability of finding a voltage distribution $p(\mathbf{v})$ is determined by the probability of a given source distribution $p(\mathbf{j})$ and the conditional probability of $\mathbf{v}$ given $\mathbf{j}$, $p(\mathbf{v}|\mathbf{j})$:

$$p(\mathbf{v}) = \sum_{\mathbf{j}} p(\mathbf{v}\mid\mathbf{j})\,p(\mathbf{j})\,. \tag{2.5}$$

The inverse solution, which requires determining the inverse conditional probability $p(\mathbf{v}|\mathbf{j})$, is then equivalent to the inference problem that can be treated using the Bayes theorem where

$$p(\mathbf{j} \mid \mathbf{v}) = \frac{p(\mathbf{v} \mid \mathbf{j})p(\mathbf{j})}{p(\mathbf{v})}. \tag{2.6}$$

That is, the probability of **j** given **v** is a normalized product of the "likelihood", $p(\mathbf{v}|\mathbf{j})$, with *a priori* information about the currents, $p(\mathbf{j})$. When the currents are assumed as Gaussian distributed, $\mathbf{j} \sim N(0,\mathbf{C}_j)$, the value **j*** that maximizes the likelihood with Gaussian noise $\boldsymbol{\varepsilon} \sim N(0,\mathbf{C}_e)$ is given by (Mattout et al., 2006)

$$\mathbf{j}^* = \left[\mathbf{G}^{\mathbf{T}}\mathbf{C}_e^{-1}\mathbf{G} + \mathbf{C}_j^{-1}\right]^{-1}\mathbf{G}^{\mathbf{T}}\mathbf{C}_e^{-1} \cdot \mathbf{v} \tag{2.7}$$

Comparing the expressions (2.4) and (2.7) we see that the solution that minimizes the weighted minimum norm for a given $\mathbf{C}_e$ and α is equivalent to the source distribution that maximizes the likelihood assuming, *a priori*, Gaussian currents with covariance $\mathbf{C}_j=(\alpha\mathbf{H}^{\mathrm{T}}\mathbf{H})^{-1}$.

Therefore, the problem (2.1) can be re-written according to a two-level hierarchical Parametric Empirical Bayesian model,

$$\begin{cases} \mathbf{v} = \mathbf{G} \cdot \mathbf{j} + \boldsymbol{\varepsilon}_e \\ \mathbf{j} = \boldsymbol{\varepsilon}_j \end{cases} \text{ with } \begin{cases} \boldsymbol{\varepsilon}_e \propto N(\mathbf{0}, \mathbf{C}_e) \\ \boldsymbol{\varepsilon}_j \propto N(\mathbf{0}, \mathbf{C}_j) \end{cases} \tag{2.8}$$

and the covariance matrices for the first level, $\mathbf{C}_e$, and for the second, $\mathbf{C}_j$, can be modeled as linear functions of covariance components of first level $\mathbf{Q}_e$ and second level $\mathbf{Q}_j$:

$$\begin{cases} \mathbf{C}_e = \sum_i \mu_i \mathbf{Q}_e^{(i)} \\ \mathbf{C}_j = \sum_i \lambda_i \mathbf{Q}_j^{(i)} \end{cases}. \tag{2.9}$$

This formulation allows the representation of an arbitrary number of priors in terms of $\mathbf{Q}_e^{(i)}$ and $\mathbf{Q}_j^{(i)}$. The prior $\mathbf{Q}_j = (\mathbf{H}^{\mathrm{T}}\mathbf{H})^{-1}$ represents the WMN framework and $\mathbf{H} = \mathbf{I}$ - corresponding to a source covariance $\mathbf{C}_j = \lambda\mathbf{I}$, where **I** is the identity matrix in the source space - is known as the maximal likelihood WMN constraint. Other priors are commonly used, as the smoothness constraint $\mathbf{Q}_j = \exp(-\mathbf{D}/2\zeta^2)$ in order to enforce correlation among

neighboring sources, where **D** is the Euclidian distance between dipoles and ζ a smoothness parameter (Mattout et al., 2006). Once priors are modeled, the unknown hyperparameters μ_i and λ_i can be jointly estimate directly from the data by applying the restricted maximum likelihood method (Friston et al., 2002, 2006; Mattout et al., 2006; Phillips et al., 2005). After $\mathbf{C}_e$ and $\mathbf{C}_j$ have been estimated, the inverse solution **j** is given by Eq. (2.7).

We applied this empirical Bayesian approach as implemented in SPM5 to estimate the distribution of electrical sources in the brain. The covariance matrix of the first level was assumed independent across EEG electrodes, with fixed variance computed from pre-stimulus recordings. The second level of the parametric empirical Bayesian approach was modeled by two priors: the maximal likelihood WMN constraint and the smoothness constraint, with a smoothness parameter set to 8 mm (Mattout et al., 2006).

The two-step method involving forward modeling and sources reconstruction has been thoroughly validated over the years, using simulated datasets (Hauk, 2004) peripheral evoked potentials (Komssi et al., 2004b) and intracranial recordings in epileptic patients (Tanaka et al., 2010). It provides the backbone of all major commercial (Curry and BESA) and freely available (SPM) softwares for source reconstruction and represents the gold-standard to compare other algorithms (Babiloni et al., 2000). In particular, the WMN constraint requires minimal a priori information about the nature of source distribution and results in a low-resolution (few centimeters) but reliable localization of primary neuronal currents, representing the tool of choice when robustness is more important than spatial resolution. Our own tests of the reliability of the inverse solution are reported with the results of the statistical analysis described in the section 2.4.

2.3.3 – *Automatic Anatomical Classification*

The inverse solution provides the instantaneous electrical activity of dipole-like primary sources $\mathbf{j}(\mathbf{x}_j,t,k)$, located at spatial coordinates $\mathbf{x}_j$ of the cortical surface as modeled by a set of vertexes $\{j = 1 \ldots N_j\}$. By applying the inverse of the transformation matrix estimated by SPM5, which maps the individual cortical surface into the MNI atlas (see section 2.3.1), to region of-interest masks provided by WFUPickAltas tool (freely available at: http://www.ansir.wfubmc.edu; Maldjian et al., 2003, 2004), it is possible to automatically

identify, for each subject, the vertexes belonging to a given cortical area according to both the Automated Anatomical Lobules and Brodmann areas classifications. Therefore, by summing the current vector $\mathbf{j}(\mathbf{x}_j,t,k)$ over these anatomically- and/or functionally-identifiable brain regions we could estimate significant anatomical and functional measures of source activity.

2.4 – How to access cortical activation produced exclusively by causal interactions in the stimulated circuits?

We saw in section 2.2 that the combination of TMS with hd-EEG allows stimulating directly the cortical surface in a controlled way and recording evoked potentials which reflect deterministic properties of the stimulated neural circuits. After source reconstruction (section 2.3), secondary currents are eliminated and single trial cortical distributions of primary dipolar currents, locket to the stimulation, are obtained. By averaging these TMS-related currents, the source activity due to uncontrolled internal and external factors is minimized revealing the spatiotemporal components of the EEG signal related to the stimulus. These averaged currents are then: 1) associated to neural sources; 2) reflecting deterministic properties of the neuronal circuits; and 3) locked to the stimulation of a given cortical area. The assumption that these currents are produced, in the region under the stimulator, directly by the magnetic stimulation, or, in distant cortical areas, by the causal effect of the activation of the stimulated cortical neurons on other neuronal groups, depends on a forth and last requirement: 4) the averaged activity in question must be absent when the stimulus is not applied.

Thus, in order to estimate the activity related to the stimulation by causal interactions between neuronal groups, after source modeling we need to assess where ($\mathbf{x}_j$) and when (t) the average cortical response

$$\mathbf{j}_0(\mathbf{x}_j,t) = \frac{\mathbf{1}}{\mathrm{N}_k}\sum_{k=\mathbf{1}}^{N_k}\mathbf{j}(\mathbf{x}_j,t,k), \tag{2.10}$$

is significantly different from pre-stimulus source activity. Due to the large number (3004) of cortical sources, assessing significance involved controlling for the risk of false positives. We controlled false-positives through the familywise error rate based on global maximum

distributions, by applying a non-parametric permutation-based statistical procedure (Pantazis et al., 2003). Under the null hypothesis of no TMS-evoked response in the average and assuming that noise is equally distributed before and after TMS pulses, a random permutation of pre- and post-stimulus periods in a subset of trials should result in approximately the same mean activity as in the original dataset. Provided that an equal number of samples T is collected before and after TMS pulses (2T+1 time samples in each trial), it is then possible to generate $N_p < 2^{Nk}$ sets, each set resulting from N_k binary decisions, one for each trial, about swapping the pre- and post-stimulus periods. For each permuted set *p*, the corresponding average cortical response $\mathbf{j}_p(\mathbf{x}_j,t)$ was then computed for comparison with the original average $\mathbf{j}_0(\mathbf{x}_j,t)$. Before comparison, average responses at each source location were normalized by subtracting the mean pre-stimulus value and dividing by pre-stimulus variance (normalized averages $\mathbf{j}_0^n(\mathbf{x}_j,t)$ and $\mathbf{j}_p^n(\mathbf{x}_j,t)$), in order to allow for an equal weighting of sources in the statistical analysis. Instantaneous control over multiple comparisons was obtained by computing the cumulative distribution function *F* of the maximum absolute value in space of the source activity at each time *t*:

$$\mathbf{M}_p(t) = \max_{\mathbf{x}_j} \left| \mathbf{j}_p^n(\mathbf{x}_j,t) \right| \text{ with } p = \mathbf{1},...,N_p . \tag{2.11}$$

The null hypothesis at time *t* for the source located at coordinates $\mathbf{x}_j$ was rejected with probability of false positives α when

$$\left| \mathbf{j}_\mathbf{0}^n(\mathbf{x}_j,t) \right| > (\mathbf{1}-\alpha) \ 100^{\text{th}} \text{ percentile of } F . \tag{2.12}$$

This method allowed identifying the binary spatiotemporal distribution of **statistically significant sources** ($\mathbf{SS}(\mathbf{x}_j,t)$): $\mathbf{SS}(\mathbf{x}_j,t) = 0$ for all sources ($\mathbf{x}_j$) and time samples (*t*) in which the null hypothesis could not be rejected, $\mathbf{SS}(\mathbf{x}_j,t) = 1$ otherwise. Results were constructed with a significance level $\alpha = 0.01$, estimated from $N_p = 1000$.

The reliability of the inverse solution and of the statistical analysis was tested by applying the analysis to somatosensory (SEP) (right median nerve) and auditory (AEP) (binaural tones) evoked potentials. Significant vertexes were detected in the contralateral hand area of the somatomotor cortex in the case of SEP and, bilaterally, in the supra-temporal plane in the case of AEP. The results of this analysis are reported in Figure (2.5).

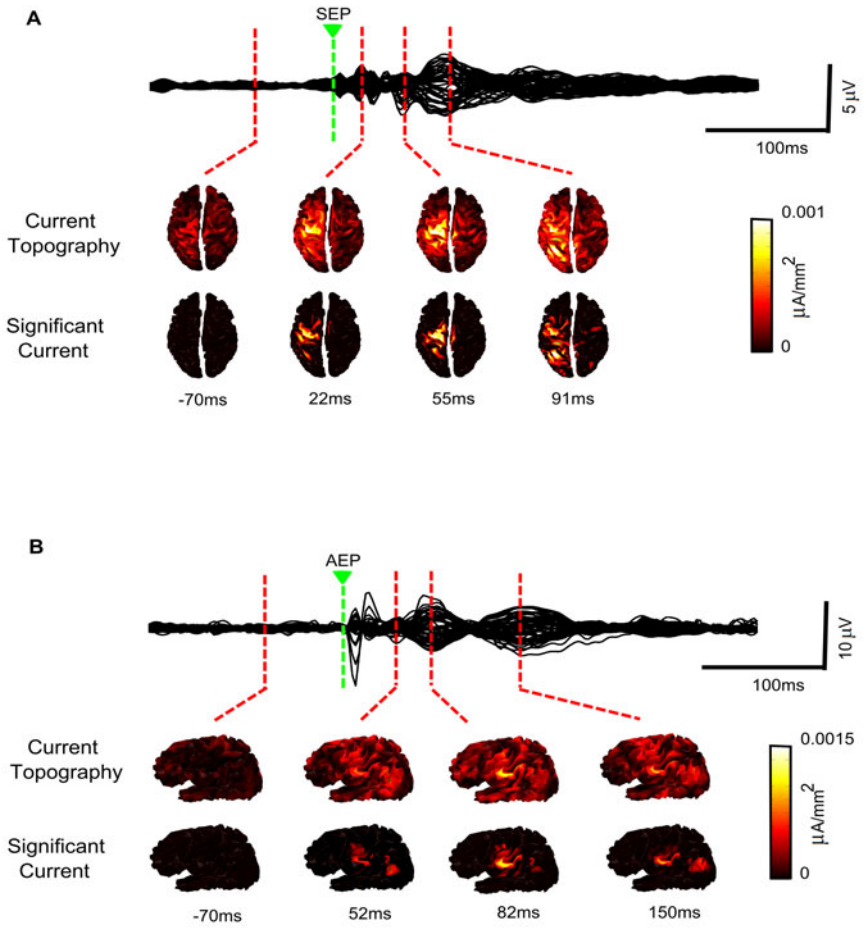

Figure 2.5: Accuracy of source reconstruction after non-parametric statistics. The butterfly plot (black traces) of (**A**) somatosensory-evoked potentials (SEPs) and (**B**) auditory-evoked potentials (AEPs) are displayed. The spatial distributions (cortical maps) of the current sources are shown for some relevant components.

2.5 – Can we quantitatively describe the pattern of cortical activation evoked by TMS?

In the previous two sections, we described how it is possible to access the cortical activity produced, in the region under the stimulator, **directly by the magnetic stimulation**, and, in distant cortical areas, by the **causal effect of the activation of the stimulated cortical neurons on other neuronal groups**. In neurophysiology, the term "**excitability**" usually

refers to the amplitude of the immediate neural response to a perturbation, while the effect of the activation of a neuronal group on another set of neurons is termed "**effective connectivity**" (Friston, 2002). Therefore, TMS/hd-EEG opens up the possibility to measure directly the cortical excitability of virtually any cortical area and to detect specific patterns of effective connectivity in the thalamocortical system.

In Casali et al. (2010) we implemented a data-driven procedure to characterize TMS-evoked cortical potentials based on a limited set of indices of cortical excitability and connectivity. These general indices can be used to describe synthetically the large-scale effects of TMS on cortical circuits, even when very little *a priori* knowledge is available.

2.5.1 – *Indices of Cortical Responsiveness*

Starting from the spatiotemporal distribution of statistically significant sources (**SS**($\mathbf{x}_j$,t)), we derived three general indices of global cortical responsiveness. These indices are designed to capture different aspects of the electrical reaction of the cerebral cortex to TMS.

Significant current density (SCD). SCD is designed to capture the overall strength of the currents evoked by TMS in the cerebral cortex. It is calculated by cumulating the absolute amplitude of all significant TMS-evoked currents over a time interval σ and/or a cortical region s:

$$\begin{aligned} \mathrm{SCD}_s(t) &= \sum_{\mathbf{x}\in s} \mathrm{SS}(\mathbf{x}_j,t)\cdot\left|\mathbf{j_0}(\mathbf{x}_j,t)\right| \\ \mathrm{SCD}_\sigma(\mathbf{x}_j) &= \sum_{t\in\sigma} \mathrm{SS}(\mathbf{x}_j,t)\cdot\left|\mathbf{j_0}(\mathbf{x}_j,t)\right| \\ \mathrm{SCD}_{s\sigma} &= \sum_{\mathbf{x}\in s}\sum_{t\in\sigma} \mathrm{SS}(\mathbf{x}_j,t)\cdot\left|\mathbf{j_0}(\mathbf{x}_j,t)\right| \end{aligned} \qquad (2.13)$$

$\mathrm{SCD}_{s\sigma}$ is a single number representing the absolute total current of the sources significantly activated by TMS pulses in a cortical volume s and time interval σ. Similarly, SCD_s and SCD_σ describe, respectively, the temporal modulation of all TMS-evoked currents in a volume s and the spatial distribution of TMS-evoked currents during a finite time interval σ.

Phase-locking (bPL). This index reflects the broad-band ability of TMS to affect the phase of ongoing oscillations at the cortical source level.

Given a set of complex numbers on the unit circle, broad-band phase-locking (bPL) can be defined as the absolute value of their average. Single-trial time series of TMS-evoked currents from each cortical source $\mathbf{j}(\mathbf{x}_j,t,k)$ can be transformed into complex analytic signals by means of the Hilbert transform (H(•)),

$$\mathbf{z}(\mathbf{x}_j,t,k)=\mathbf{j}(\mathbf{x}_j,t,k)+i\mathrm{H}(\mathbf{j}(\mathbf{x}_j,t,k)), \tag{2.14}$$

and then normalized to obtain unitary absolute values $\mathbf{z}_n(\mathbf{x}_j,t,k)$. Therefore, instantaneous bPL of a single source can be obtained as (Fründ et al., 2007; Palva et al., 2005; Sinkkonen et al., 1995; Tallon-Baundry, et al., 1996)

$$\mathrm{bPL}(\mathbf{x}_j,t)=\left|\bar{\mathbf{z}}_n(\mathbf{x}_j,t)\right|=\left|\frac{\sum_k \mathbf{z}_n(\mathbf{x}_j,t,k)}{\mathrm{N}_k}\right|. \tag{2.15}$$

bPL ranges from 0 (random phases) to 1 (perfect phase locking). Under the null hypothesis that the complex numbers $\{\mathbf{z}_n(\mathbf{x}_j,t,k)\}$ at each time sample t are randomly uncorrelated and uniformly distributed, bPL values follow a Rayleigh distribution

$$P(\mathrm{bPL}(\mathbf{x}_j,t))=\frac{\pi\cdot\mathrm{bPL}(\mathbf{x}_j,t)}{2m^2}e^{-\frac{\pi\cdot\mathrm{bPL}^2(\mathbf{x}_j,t)}{4m^2}}, \tag{2.16}$$

with m being the mean distribution value, that can be estimated from baseline by

$$m(\mathbf{x}_j)=\frac{\sum_{t=-T}^{-1}\mathrm{bPL}(\mathbf{x}_j,t)}{T}. \tag{2.17}$$

The null hypothesis of random phase distribution of TMS-evoked potentials can be rejected with a level of significance α when $P(\mathrm{bPL}(\mathbf{x}_j,t)) > 1-\alpha$. Correction for multiple comparisons was performed by applying to the bPL obtained from each source the Rayleigh distribution with $m = m^*$, where

$$m^*=\max_{\mathbf{x}_j}[m(\mathbf{x}_j)]. \tag{2.18}$$

bPL is amplitude independent, robust to large artefacts, is not based on the spectral content of the data and has excellent temporal resolution. Similarly to SCD, bPL can also be cumulated over a cortical volume *s* and/or a time interval σ to obtain synthetic indices of broad-band phase-locking (bPL_s, bPL_σ, and $bPL_{s\sigma}$,).

Significant current scattering (SCS). This index captures the spatial spreading of the significant activations evoked by TMS. SCS was calculated by cumulating the geodesic distances ($d(\mathbf{x}_j)$) between any significant current source and the TMS target over a time interval σ and/or a cortical volume *s*:

$$\begin{aligned} \mathrm{SCS}_s(t) &= \sum_{\mathbf{x}\in s} \mathrm{SS}(\mathbf{x}_j,t)\cdot d(\mathbf{x}_j,t) \\ \mathrm{SCS}_\sigma(\mathbf{x}_j) &= \sum_{t\in\sigma} \mathrm{SS}(\mathbf{x}_j,t)\cdot d(\mathbf{x}_j,t) \\ \mathrm{SCS}_{s\sigma} &= \sum_{\mathbf{x}\in s}\sum_{t\in\sigma} \mathrm{SS}(\mathbf{x}_j,t)\cdot d(\mathbf{x}_j,t). \end{aligned} \tag{2.20}$$

This index is maximally sensitive to the spatial extent of the significant activations evoked by TMS and depends minimally on the amplitude of the TMS-evoked responses. Thus, SCS increases proportionally to the spreading of TMS-evoked activity from the site of stimulation.

2.5.2 – *Interpretation of Indices of Cortical Responsiveness*

In order to test the sensitivity of the above synthetic measures with respect to specific properties of the stimulated circuits, TMS-evoked potentials were recorded from five healthy adults (two males and three females, age range 23–37years) at different stimulation intensities, during randomly ordered sub-sessions, when the medial third of the left superior occipital gyrus in BA19 was stimulated. The intensity of TMS was referred to the nearest-sensor EEG threshold (NSET), individually defined as the maximal TMS intensity (expressed as a percentage of the stimulator's maximal output) that failed to evoke visible potentials in the electrode nearest to the coil. NSET was assessed at the beginning of each experiment by visually inspecting online the EEG potentials obtained by averaging 20–30 single-trial responses to TMS. Full sessions of stimulation were then delivered to each subject at 65%, 100%, 135%, 145%, 170%, 205%, 240% and 275% of individual NSET, with a tolerance of 5%. For each intensity, 150–200 stimuli were delivered at a frequency randomly distributed

between 0.43 and 0.50 Hz. During the experiment, subjects were lying on an ergonomic chair, relaxed and with eyes open looking at a fixation point on a screen.

Figure 2.6 depicts the time course over the whole brain (panel A) and the spatial distribution over the full post-stimulus period (panel B) of SCD, bPL and SCS obtained when BA19 was stimulated at increasing intensities in one subject. The time courses and the spatial maps of the three indices show that the duration of the overall cortical response progressively increased in time and that the set of cortical regions engaged by TMS progressively enlarged when stimulation intensity was increased above 100% NSET. Albeit these general similarities, SCD, bPL and SCS seemed to capture different features of the brain response to TMS. In all subjects SCD peaked early and decayed rapidly, bPL was more sustained in time, while SCS was the last one to attain its peak values and the last one to decay (Figure 2.7). Also, considering the spatial distribution maps exhibited in Figure 2.8 for all subjects SCD had the highest values in the stimulated area (left Superior occipital gyrus), bPL showed a rather balanced activation between cortical regions located near and far from the TMS target, while SCS values were maximal in distant brain areas (frontal lobe). Based on the individual time course of these indices, it was possible to map, in each subject, the instantaneous spatial distribution of SCD, bPL and SCS at the peak of their activity (Figure 2.9). This representation revealed a consistent pattern across individuals, showing that the strongest cortical activation (SCD peak) always occurred in the stimulated area between 28 and 36 ms, while the maximum spread of activity (SCS peak) occurred in left frontal regions (ipsilateral to the stimulation target) between 77 and 101 ms.

Thus, SCD, being highly sensitive to the response's amplitude, was higher shortly after the TMS, when the target area reacts to the direct stimulation, and was more sensitive to the immediate brain's reaction to TMS. SCS was minimally dependent on the response's amplitude but maximally sensitive to its spatial extension, being capable of detecting the subsequent long-range spreading of activation. bPL being an amplitude-independent measure that is not weighted towards distant activations lied in between SCD and SCS and was equally sensitive to the local (excitability) and the distant (connectivity) effects of TMS.

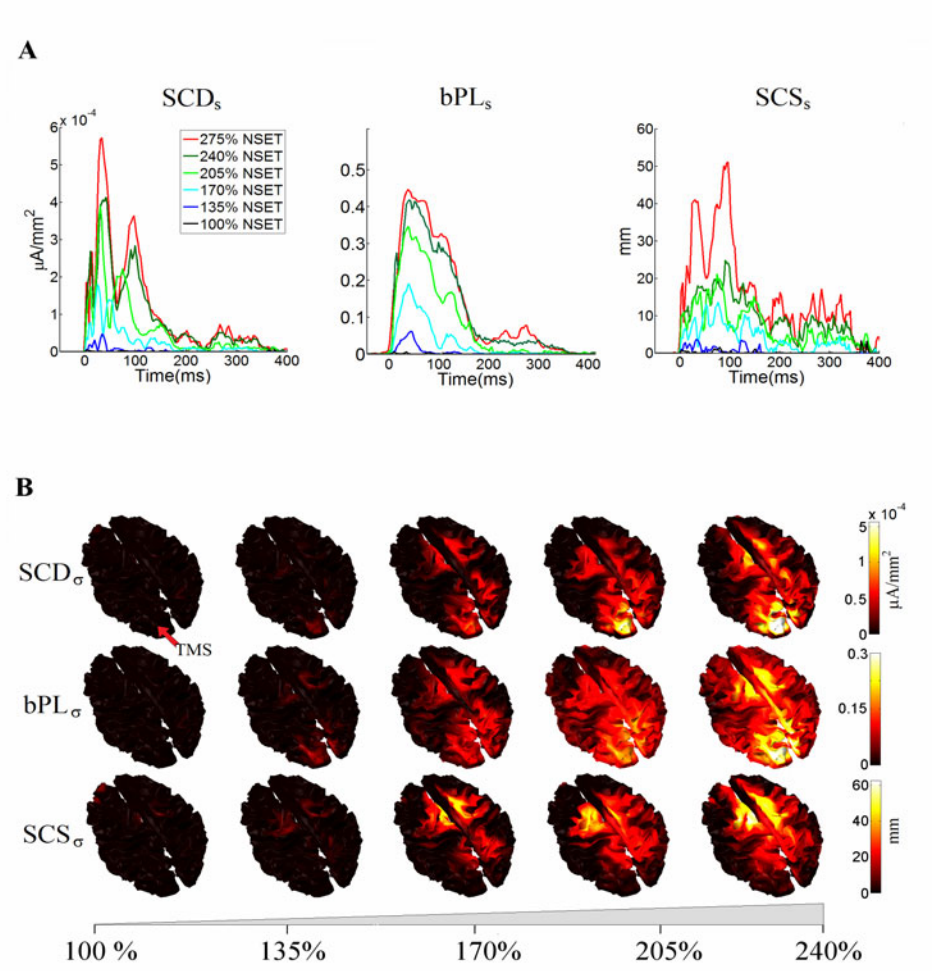

Figure 2.6: Synthetic indices of cortical responsiveness to TMS in a single subject at different stimulation intensities. Time course (**A**) and spatial distribution (**B**) of significant current density (SCD), broad-band phase-locking (bPL) and significant current scattering (SCS), averaged over the whole brain(s) and the full poststimulus period(σ) respectively (adapted from Casali et al., 2010).

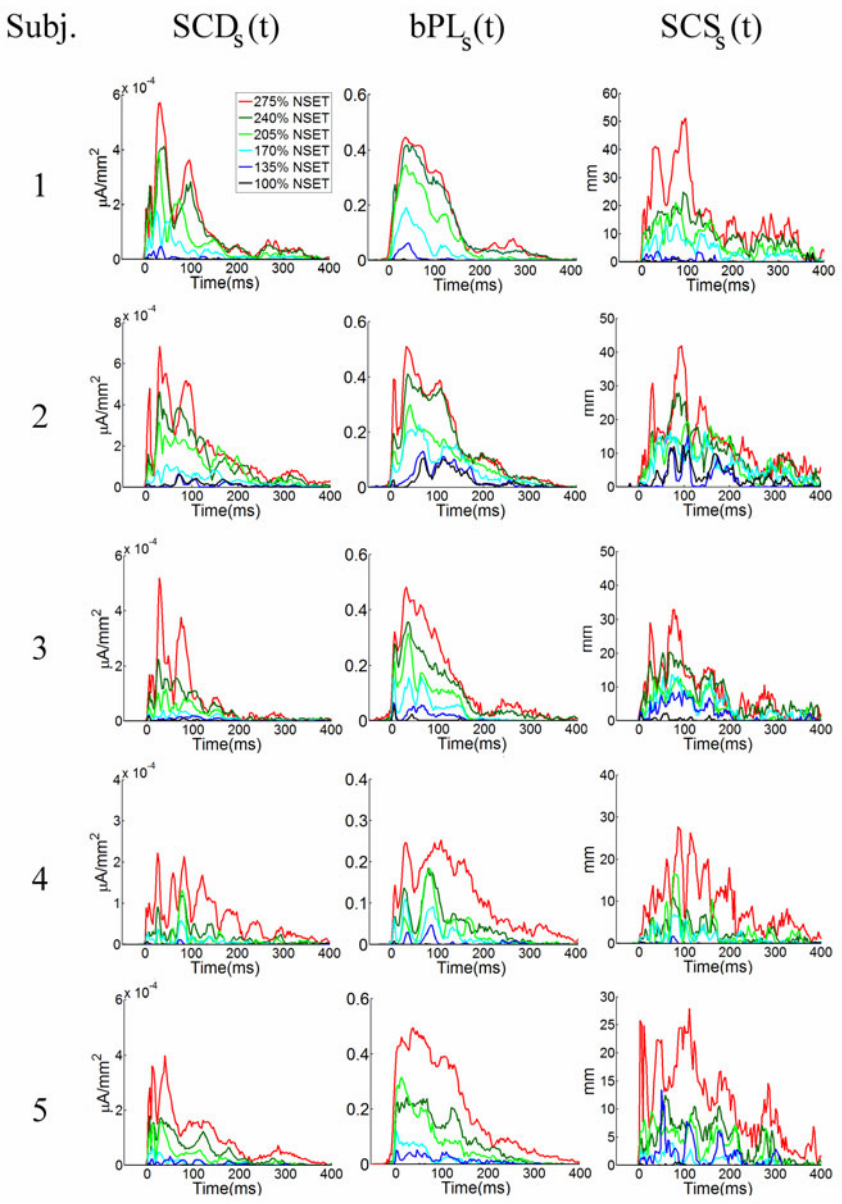

Figure 2.7: Time course of SCD, bPL and SCS averaged over the whole brain (s), for all subjects at different stimulation intensities (adapted from Casali et al., 2010).

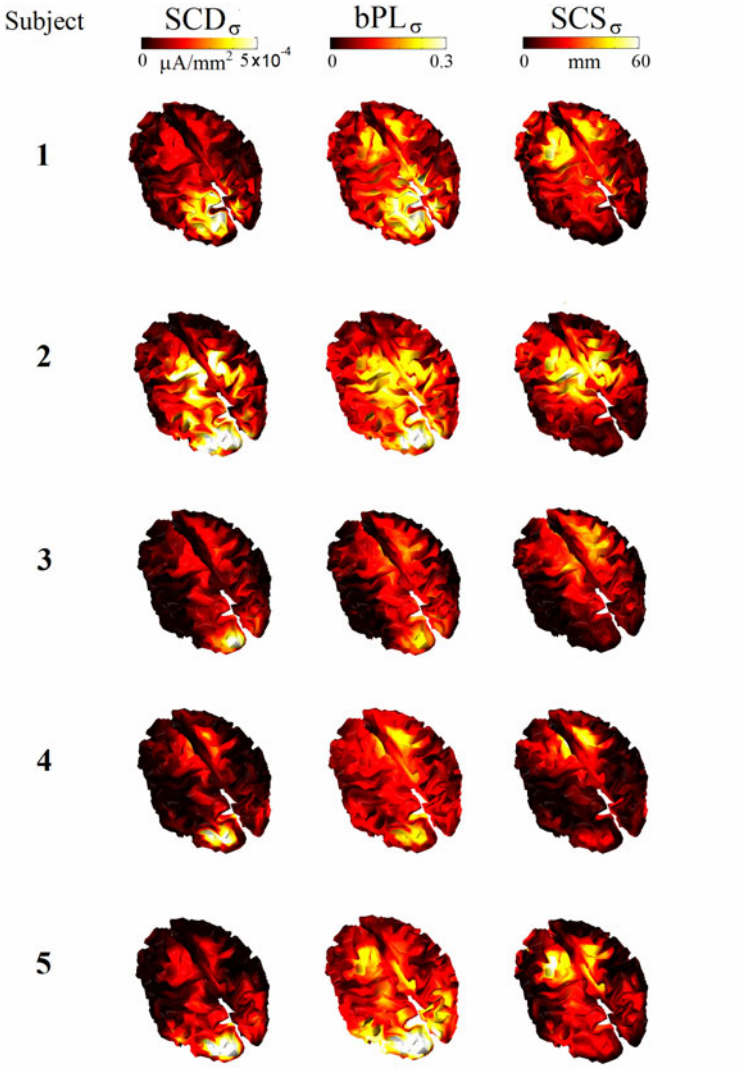

Figure 2.8: Spatial distribution of SCD, bPL and SCS averaged over the entire post-stimulus period, for all subjects at 275% NSET (adapted from Casali et al., 2010).

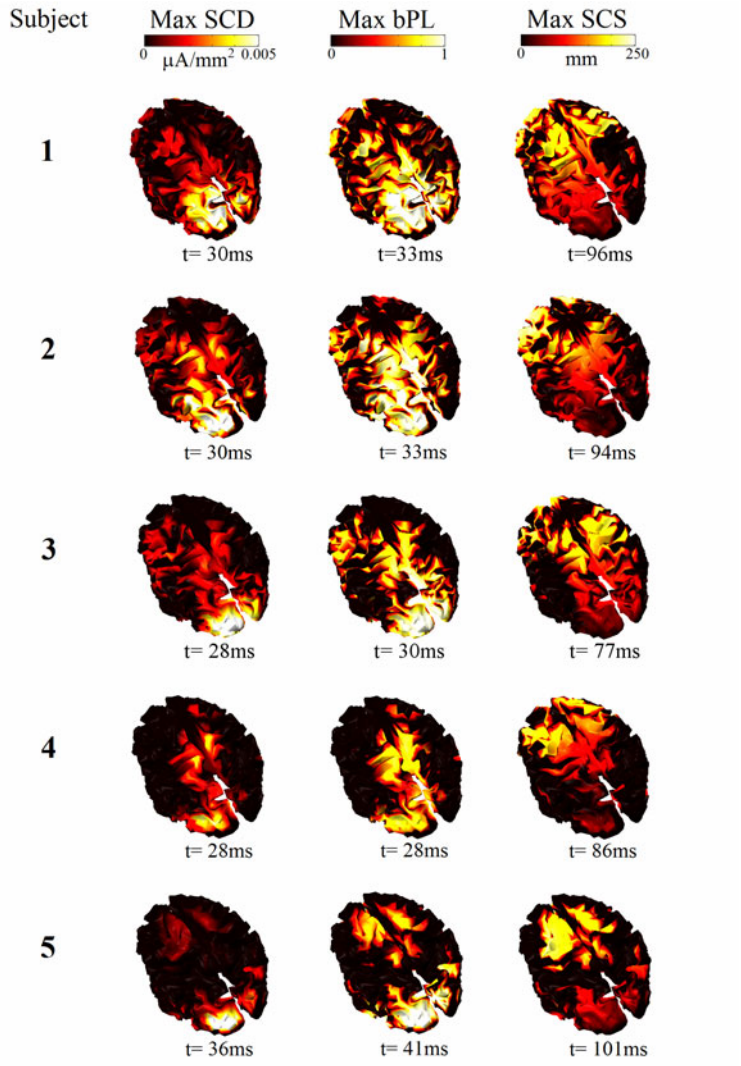

Figure 2.9: Instantaneous spatial distribution of SCD, bPL and SCS, in correspondence to their respective peak of activity in time, at 275% NSET stimulation intensity for all subjects (adapted from Casali et al., 2010).

2.5.3 – *Measuring Local Excitability*

SCD, bPL and SCS reflected different properties of the stimulated circuits. In particular, SCD suggested that a strong reaction to TMS occurs within the first 50 ms in a region located immediately under the coil. Thus, based on this information, we restrict our analysis to detect the local excitability (threshold) of the stimulated area (BA19).

Typically, cortical excitability, or the amplitude of the immediate neural response to a perturbation, is defined operationally as the local activation threshold, i.e., the minimum stimulation intensity eliciting a reliable motor response of the target muscles (Rossini et al., 1994). This approach, although practical and reliable, inevitably confines the assessment of excitability to cortico-spinal circuits. TMS/hd-EEG may extend the measurement of cortical excitability to a larger set of cortical areas. Using this technique, one could, for instance, implement a method to detect the minimum stimulation intensity required to trigger significant neuronal currents in the stimulated area. Observing and comparing the time courses and spatial distributions of SCD and SCS (Figs. 2.7 and 2.8) reveals that TMS always evoked, during the first 50 ms, strong currents localized in the target region, possibly reflecting the initial trans-synaptic activation of the neuronal population located under the coil. Based on this observation, we defined the TMS/hd-EEG threshold as the minimum stimulation intensity (TMS-induced electric field on the cortical surface) required for significantly activating more than 1% of the cortical sources within the stimulated area during the first 50 ms after TMS pulse. The temporal and spatial boundaries of this definition were directly derived from the data, while the choice of the minimum percentage of active sources was motivated by the fact that a non-parametric statistics with $\alpha = 0.01$ admits a 1% rate of false positives.

Application of this procedure to real data showed a dramatic drop of significance (~<1% of significantly active sources) in all subjects for stimulation intensities below 50 V/m and no significant activations below 40 V/m (Figure 2.10). These values are in good agreement with the activation threshold of the cat's phrenic nerve measured *in vitro* (Maccabee et al., 1993) and with the threshold of the human median nerve measured *in vivo* (McRobbie and Foster, 1984). These results are also similar to the ones obtained by Komssi et al. (2007) using TMS/hd-EEG on motor cortex. While in Komssi et al. (2007) the threshold was defined as the TMS intensity yielding a 50% probability of evoking some pre-selected

EEG components in a sample of healthy subjects, the procedure proposed here allows detecting the cortical activation threshold automatically at the single-subject level.

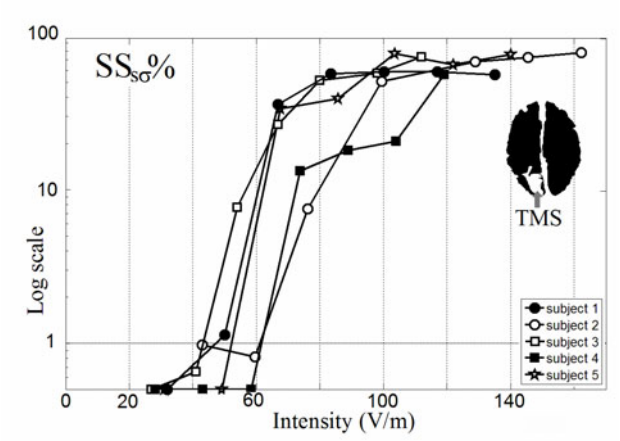

Figure 2.10: Caracterizing Cortical Excitability: log-transformed percentage of significant current sources cumulated over the TMS target region (s = left superior occipital lobule, white area in the cortical map on the right) and the temporal window 0–50 ms (σ) as a function of the TMS-induced cortical electric field (automatically estimated in V/m by the navigation system) for all subjects. Activation threshold (dotted line) is set at 1%, based on the α-value of significance of non-parametric statistics (adapted from Casali et al., 2010).

2.5.4 – *Measuring Cortical Connectivity*

In this work, we are particularly interested in employing the combination of TMS and hd-EEG to access the capability of different elements of the thalamocortical system to causally interact with each other. Figure 2.11 represents an example of how the above standardized analysis of TMS/hd-EEG data can help identifying subsets of cortical regions that are specifically linked by causal interactions.

In this example, TMS was applied to the left Superior occipital lobule at high intensity (275% of NSET) during two different days (day 1, red color; day 2, blue color) in order to evaluate test-retest reproducibility. In the same subject, the left motor cortex was also targeted (green color) at the same TMS intensity, in order to appreciate the specificity of TMS-evoked cortical activation. Figure 2.11A depicts the total value (bars), the spatial distribution (maps) and the time course (lines) of SCD, bPL and SCD calculated in the three different conditions and respectively integrated over the whole brain and post-stimulus period. While repeating the same occipital stimulation on two different days evoked a similar spatiotemporal pattern

of cortical activation, TMS on motor cortex resulted in clearly different spatial distributions and time courses of SCD, bPL and SCD. The same observations can be derived from Figure 2.11B where the spatial distribution of the indices is shown at their respective peak values; following a strong initial response in the target area, activity propagated to the ipsilateral frontal cortex within the first 100–110 ms in the case of the two occipital stimulations, while a dominant contralateral activation was detected within the first 50 ms after motor cortex stimulation.

Cumulating SCD and SCS values over discrete regions of interest (see section 2.3.3) allowed reducing the data space and identifying a set of distant cortical regions specifically engaged by stimulation of the left Superior occipital lobule. The histograms depicted in Figure 2.11C represent (for the three conditions) the values of SCD and SCS calculated in all cortical lobules, sorted according to their geodesic distance from the left Superior occipital lobule. Stimulation of the occipital cortex on day 1 and day 2 resulted in histograms characterized by the same peaks of regional activation (the regions with higher SCS values are color-coded and plotted on the cortical surface below), involving the left superior parietal lobule, the left precuneus, the left supplementary motor area, the left precentral region and left superior frontal lobule. Conversely, the same analysis applied to motor cortex stimulation detected a strong bilateral activation of motor cortex, somatosensory cortex, precuneus and of ipsilateral parietal lobule. Extracting the time course of the significant currents from a subset of cortical regions (Figure 2.11D) also highlighted the reproducibility and the specificity of TMS-evoked cortical activations.

This pattern of effective connectivity evoked by stimulation of the Superior occipital lobule is consistent with cortico–cortical interactions occurring through the dorsal visual stream (Colby et al., 1988; Rizzolatti and Matelli, 2003; Tanné et al., 1995) or through the occipito-frontal fascicle. Intracranial recordings in monkeys (Lamme and Roelfsema, 2000; Schroeder et al., 1998) and humans (Gaillard et al., 2009) have demonstrated that peripheral visual stimulation results in a progressive posterior–anterior propagation of neural activity that reaches area V3 (BA19) at around 50 ms and the ipsilateral frontal lobe at latencies between 120 and 150 ms. Consistent with the results of these invasive recordings, the present experiment showed a maximum spread of activation (Max SCS) in ipsilateral frontal cortex occurring 77–101 ms after direct stimulation of left Superior occipital—BA19 (Figure 2.9).

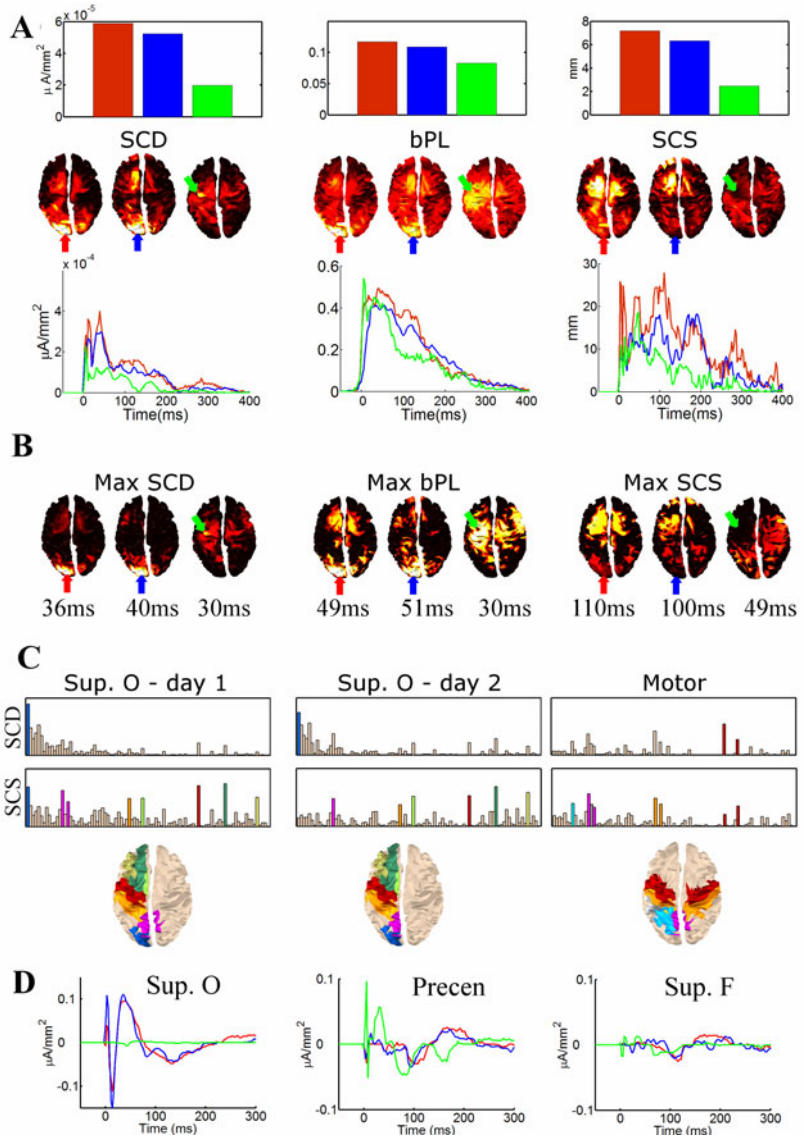

Figure 2.11: Caracterizing Cortical Connectivity. Synthetic indices of cortical responsiveness to TMS computed in a single subject at a stimulation intensity of 275% NSET delivered to the left Superior occipital cortex in two different days (red and blue colors, respectively) and to the left hand motor area (green color). **A)** SCD, bPL and SCS values cumulated either over the whole brain (time courses) or over the full post-stimulus period (spatial maps) or over both the whole brain and the post-stimulus period (bars). **B)** Instantaneous spatial distribution of SCD, bPL and SCS indices in correspondence to their respective peak of activity in time. **C)** SCD and SCS cumulated over the post-stimulus and over single regions of interest, sorted according to the distance from the left Superior occipital cortex. Regions with higher SCS activity and the region with highest SCD activity are color-coded and highlighted in the cortical outline below. **D)** Time course of dipolar currents cumulated over three regions of interest ipsilateral to the TMS stimulation— superior occipital (Sup. O) lobule, precentral area (Precen) and superior frontal (Sup. F) cortex—when TMS is delivered to the left Sup. O (red and blue lines) and to the left hand motor (Motor) area (green line) (adapted from Casali et al., 2010).

Chapter 3.

Causal Responses of the Thalamocortical System to Transcranial Magnetic Stimulation

The previous chapter described how to employ the combination of TMS and hd-EEG to perturb directly, in a controlled and reproducible way, specific cortical areas and measure the corresponding neural responses that reflect causal patterns of interaction between different groups of neurons in the thalamocortical system. What then can be said about such responses evoked by TMS in conditions such as alert wakefulness, in which consciousness is unambiguously present, or when it is clearly reduced, such as during deep sleep, anesthesia and vegetative state? In this chapter we will show that the causal interaction of different connected neuronal modules with specific oscillatory frequencies produces qualitatively complex patterns of cortical response to TMS during alert wakefulness. We will also show that this ability of the thalamocortical system to sustain complexity is consistently impaired among unconscious healthy subjects and brain-injured patients.

3.1 – Alert Wakefulness

3.1.1 – *The Natural Frequencies of Corticothalamic Circuits*

A system of causally interacting neurons can be generally thought as a set of coupled electrical oscillators. It is well known that the perturbation of coupled oscillators may reveal

the tuning frequency, or "natural frequency" of the system which, in turn, indicates structural properties of the state of the system under study. In Rosanova et al. (2009) we employed TMS/hd-EEG to measure the natural frequencies of different corticothalamic modules in awake healthy subjects. We targeted TMS to three different cortical areas (BAs 19, 7 and 6) of 6 subjects, stimulating each cortical area at 8 different intensities (range: 20-160 V/m). During the experiment, subjects were lying on an ergonomic chair, relaxed and with eyes open looking at a fixation point on a screen. For each intensity we delivered between 100 and 200 stimuli at a frequency jittering between 0.4-0.5 Hz (period: 2000 plus a jitter $\leq$ 300 ms).

Figure 3.1 displays the EEG responses recorded in one subject from all sensors after stimulation of area 19, area 7 and area 6 at 120V/m on the cortical surface. The butterfly plots reveal that the brain electrical response to TMS varied markedly depending on the site of stimulation: following an early (0-20 ms) stereotypical sharp component, TMS of area 19 resulted, during the first 200 ms, in a low frequency, large response; stimulation of area 7 elicited faster and smaller components; and stimulation of area 6 evoked the fastest EEG oscillations.

In order to detect the frequency of the oscillations induced by TMS at different cortical sites, we calculated the Event Related Spectral Perturbation (ERSP) based on Morlet wavelets. The procedure was implemented using the public license toolbox EEGLAB (Delorme and Makeig, 2004). ERSP decomposes the EEG response recorded from each sensor in the time-frequency space and allows tracking the significant spectral modulations induced by TMS during the post-stimulus time. We averaged, for each condition, the ERSPs matrices across all channels in order to obtain a global-ERSP (gERSP). The bottom panels of Figure 3.1 display the resulting plots for the three stimulation sites. In all cases, TMS resulted in a significant early (8-20 ms) activation in the β_2/γ bands (21-50 Hz). After this short-lasting stereotypical event, the frequency content of the global brain response to TMS varied markedly depending on the stimulated area: dominant oscillations in the α range (8-12 Hz) were detected when area 19 was stimulated, in the β_1 range (13-20 Hz) when area 7 was stimulated and in the β_2/γ range (21-50 Hz) when area 6 was stimulated.

In order to extract the natural frequency associated to the TMS-evoked potentials in each condition, we averaged the gERSP in a time window between 20 and 200 ms post-stimulus, minimizing the effect of possible artifacts occurring at the time of stimulation. We

detected, on the resulting spectral profiles (black trace on the right of each panel), the frequency with maximum power (indicated by a dotted line). Across all subject, at the intensity of 120 V/m, the natural frequency of the global scalp response to stimulation of area 19, area 7 and area 6 was 11 ± 1.5 Hz, 18.3 ± 2.0 Hz and 29 ± 2.0, respectively.

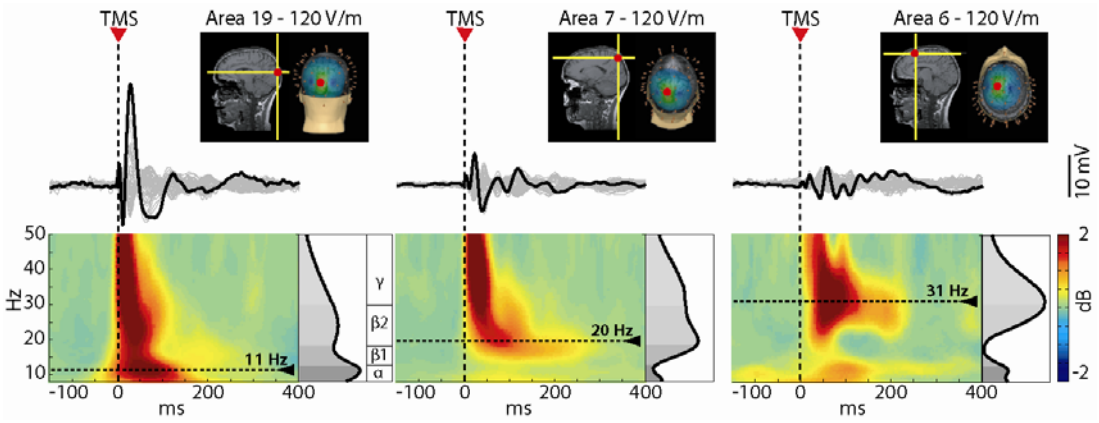

Figure 3.1: TMS induced global EEG oscillations that are specific for the stimulated site. The figurine illustrates the 3 cortical sites targeted by TMS (hot spot on the individual MRI). The traces below represent butterfly plots; the black trace highlights the electrode directly underlying the stimulator. The bottom panels show the ERSP patterns calculated globally on the scalp (average of all electrodes). The rightmost graphs depict the power spectrum profile induced during the first 200ms after TMS. The dotted lines highlight the frequency with maximum power. TMS elicited early gamma components immediately followed by prominent alpha band oscillations after occipital stimulation, beta band oscillations after parietal stimulation and fast beta/gamma oscillations after perturbation of frontal cortex (adapted from Rosanova et al., 2009).

An important parameter that may potentially bias the frequency of the EEG response to TMS is the strength of the applied stimulus. We employed an MRI-guided navigation system to ensure that the strength of the electric field induced by TMS in different areas was always comparable. The maximum electric field was always kept on the convexity of the gyrus with the induced current perpendicular to its main axis. Moreover, in order to exclude an effect of local activation thresholds in determining the observed site-specificity of the response's frequency, we probed each site at several different intensities, from sub-threshold levels to near-saturation levels. As shown in Figure 3.2, TMS intensities below 40 V/m did not result in any significant power modulation in any of the four EEG bands. Intensities of stimulation between 60 and 120 V/m evoked progressively larger responses in all four

frequency bands. At this level the responses of the three corticothalamic modules clearly differed, displaying higher α band power after stimulation of area 19, higher β_1 band power after stimulation of area 7 and higher β_2/γ band power after stimulation of area 6. These differences further increased at stimulation intensities (140-160 V/m) that produced saturation of the TMS-evoked potentials. Thus, the specific frequency of the response did not depend on stimulation intensity, or activation threshold, but most likely depended on a number of endogenous properties of the activated circuit.

These results were reproducible in all subjects: provided that TMS intensity was above the threshold for triggering a significant EEG response, the average waveform differed according to the site of stimulation (Figure 3.3). Occipital stimulation triggered high-amplitude, lower-frequency components, while stimulating more rostral sites elicited smaller waves separated by shorter intervals. Panel B of Figure 3.3 shows the natural frequency calculated at the global sensor level in all subjects after stimulation of area 19, area 7 and area 6 at 160 V/m. In all cases, the natural frequency progressively increases from α, to β and γ when the three sites are plotted in the caudal-rostral order.

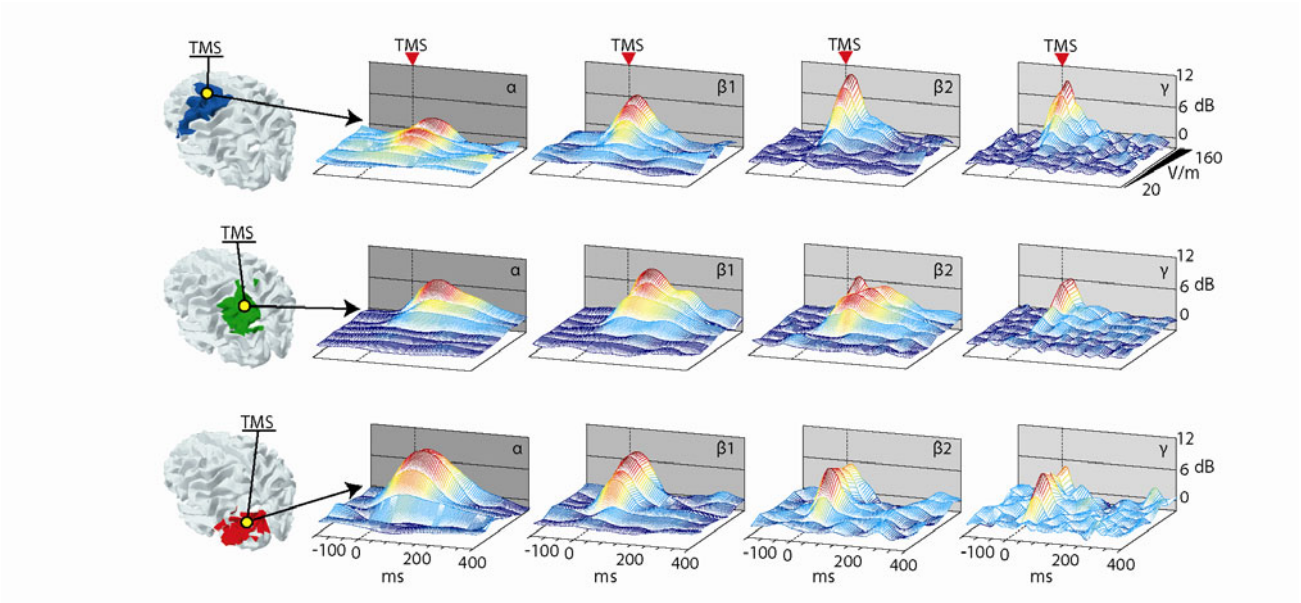

Figure 3.2: The specificity of natural frequencies was preserved with different stimulation intensities. Time series (ERSP) of standard EEG frequency bands (alpha: 8-12; beta1: 13-19; beta2: 20-29; gamma: 30-50) for the three sites of stimulation (first row: area 6; second row: area 7; third row: area 19) are plotted as a function of stimulation intensity, ranging from 20 to 160 V/m (adapted from Rosanova et al., 2009).

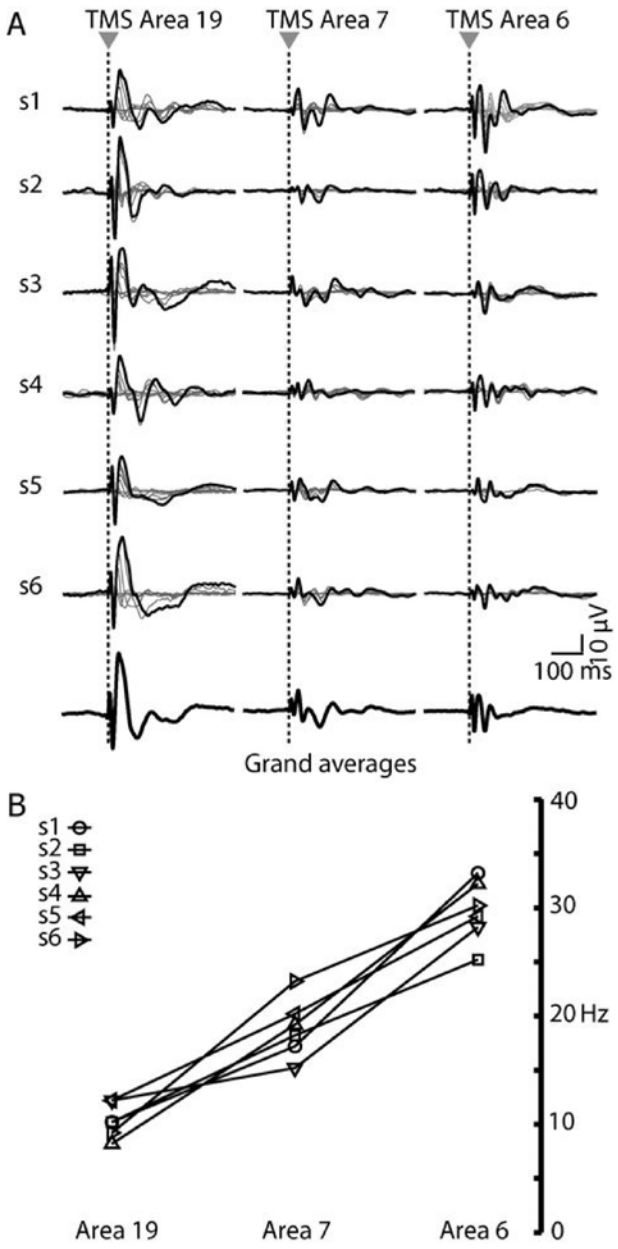

Figure 3.3: The specificity of the natural frequency was reproducible across subjects. A) The traces represent TMS responses for each subject (n=6) evoked by 6-8 intensities of stimulation, recorded from one electrode underlying the stimulator. The black trace highlights the response obtained at maximal stimulation intensity (160 V/m). The thicker black traces at the bottom represent the grand averages calculated from all 6 subjects. **B)** The frequencies with maximum power, obtained by stimulating each cortical area at the maximal intensity, are plotted for each subject. A clear posterior-anterior gradient of increasing frequencies is observable (adapted from Rosanova et al., 2009).

Did these dominant frequencies, recorded globally over the scalp, reflect the frequencies of neural oscillations generated by TMS? Time series of local currents evoked by TMS activation were obtained, after source modeling and automatic classification of the individual's cortical surface, by cumulating the dipole activities recorded in each one of the 47 Brodmann's areas (see section 2.3.3). At this point, the same time-frequency decomposition analysis performed at the global sensors level was carried out at the local source level. Figure 3.4 shows the time series of the local current and the ERSP plots recorded from all three cortical areas of interest (area 19, area 7 and area 6) when TMS was directly applied to each of them. These sources, directly activated by TMS, generated strong currents matching the dominant frequency recorded globally at the scalp level. Moreover, even when not directly stimulated, each cortical area still tended to oscillate at a rate closer to its own natural frequency. Across all subjects, when area 19 was stimulated, area 19 responded at 10.8 Hz, area 7 at 20 Hz and area 6 at 31.3 Hz; when area 7 was stimulated, area 19 responded at 13.5 Hz, area 7 at 18.6 Hz and area 6 at 27.3 Hz; when area 6 was stimulated, area 19 responded at 10.6 Hz, area 7 at 19 Hz and area 6 at 29 Hz. Hence, each area, whether directly activated by TMS or engaged through long-range connections, expressed local oscillations at a rate closer to its own natural frequency.

The electrical rhythms triggered by TMS are more likely to reflect overall circuit properties at the level of whole cortical areas and connected thalamic/subcortical nuclei. For instance, a recent study has linked α oscillations to the presence of a subpopulation of electrically coupled neurons localized in the lateral geniculate nucleus which fire bursts of action potentials in the α range when activated by a cortical glutamatergic input (Hughes et al., 2004). This mechanism, involving a whole corticothalamic module, may explain why TMS of visual cortex readily triggers α oscillations, while TMS of frontal cortex fails to do so. Similarly a role of the thalamus can be postulated in the genesis of faster oscillations (Llinas et al., 2007). Interestingly, TMS/EEG experiments performed in patients with lesions in the ventrolateral thalamus have demonstrated a significant decrease of β band oscillations after TMS of the ipsi-lesional motor cortex (Van Der Werf et al., 2006). In this perspective, our results suggest that TMS is capable to engage the thalamocortical system of alert healthy subjects, setting in motion different connected neuronal oscillators and generating complex EEG responses composed of strong fluctuations at the natural frequency of the stimulated area and by weaker fluctuations at around the natural frequency of distant regions.

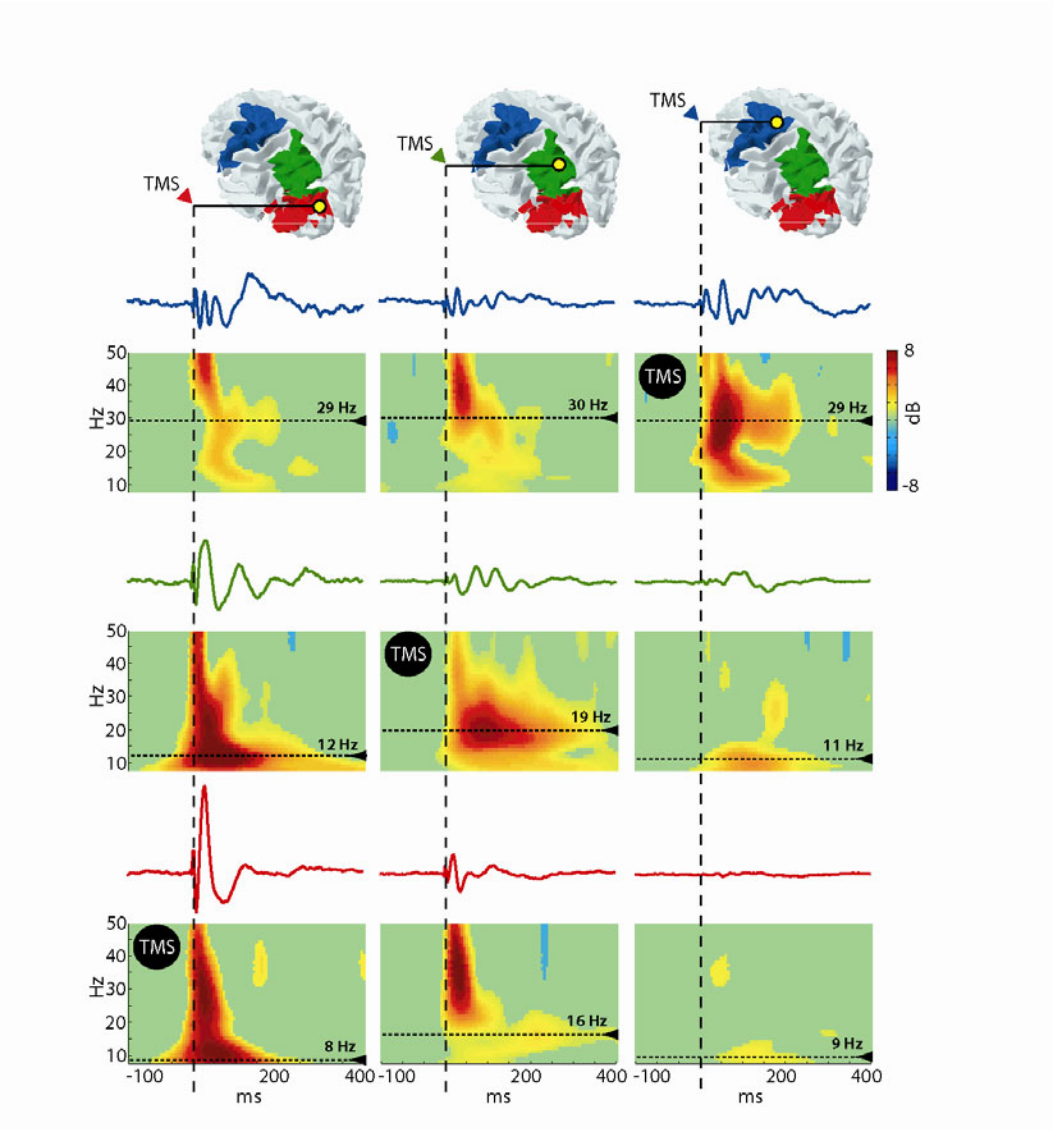

Figure 3.4: The natural frequency is a local property of individual thalamocortical modules. The coloured patches on the cortical surface mark the areas from which cortical currents are recorded after source modeling. Below, time series and ERSP plots of local cortical currents are displayed for area 6 (first row, blue traces), area 7 (second row, green traces) and area 19 (third row, red traces) when area 19 was stimulated (first column), area 7 was stimulated (second column) and area 6 was stimulated (third column). The dotted lines highlight the peak frequency for each plot. Comparing the plots on the diagonal line marked by TMS icon, reveals that each cortical area responded with a clear-cut natural frequency when directly stimulated. Comparing the plots on the horizontal and on the vertical lines, reveals that the natural frequency is a local property that was partially preserved also when its cortical generator was not directly stimulated (adapted from Rosanova et al., 2009).

3.1.2 – *The Spatiotemporal Complexity of Cortical Activation Evoked by TMS during Wakefulness*

As we saw in the first chapter, if the brain is the vehicle for the objects of consciousness, during alert wakefulness the thalamocortical system is likely to be in a state of dynamical equilibrium, being capable of integrating activity across several neuronal modules through feedforward and feedback causal connections. Thus, if these widespread and specialized cortical areas have distinct intrinsic dynamical properties and are engaged at different latencies, such complex reentrant system will generate complex spatiotemporal patterns of causal activation. In fact, the results described in the previous section indicates that TMS/hd-EEG is capable of probing this capability of integration across neuronal modules, each one oscillating with its particular natural frequency, when TMS-evoked neural currents activated different areas of the cortex at different latencies. Figure 3.5 displays the significant neuronal currents evoked by TMS in a representative awake healthy subject. The black traces in the figure are the time-series for the Global Mean Field Power (GMFP) (Lehmann and Skrandies, 1980) calculated from the multichannel average signals as:

$$\mathrm{GMFP}(t) = \sqrt{\frac{\sum_{i=1}^{k} (V_i(t) - \bar{V}(t))^2}{k}}, \qquad (3.1)$$

where k is the number of channels, V_i is the voltage measured with channel i, and $\bar{V}$ is the mean of the measured voltages across channels (average reference). The significant current distributions (SCD) are exhibited at latencies corresponding to high values of GMFP. TMS was targeted to BAs 6 (A), 7 (B) and 19 (C) at 100V/m. In all cases, TMS evoked a qualitatively complex spatiotemporal pattern of cortical currents: a series of low-amplitude waves of activity associated with cortical activations that propagated along long-range ipsilateral and transcallosal connections, lasting for around 300ms, and in which activity shifted through cortical areas in different times.

What does happen to these complex patterns of TMS-evoked activity when consciousness fades? Alterations in the membrane properties of subsets of cortical and subcortical (especially thalamic) neurons, as well as alterations in their patterns of connectivity may result in distinctive and detectable changes in the oscillatory properties of neuronal groups, possibly resulting in altered patterns of response to TMS. Hence, we ask in

the following sections whether this complexity of neuronal response to TMS due to the integration of oscillators with different site-specific natural frequencies is preserved in states where consciousness is lost, such as during sleep, anesthesia and in patients suffering from disorders of consciousness.

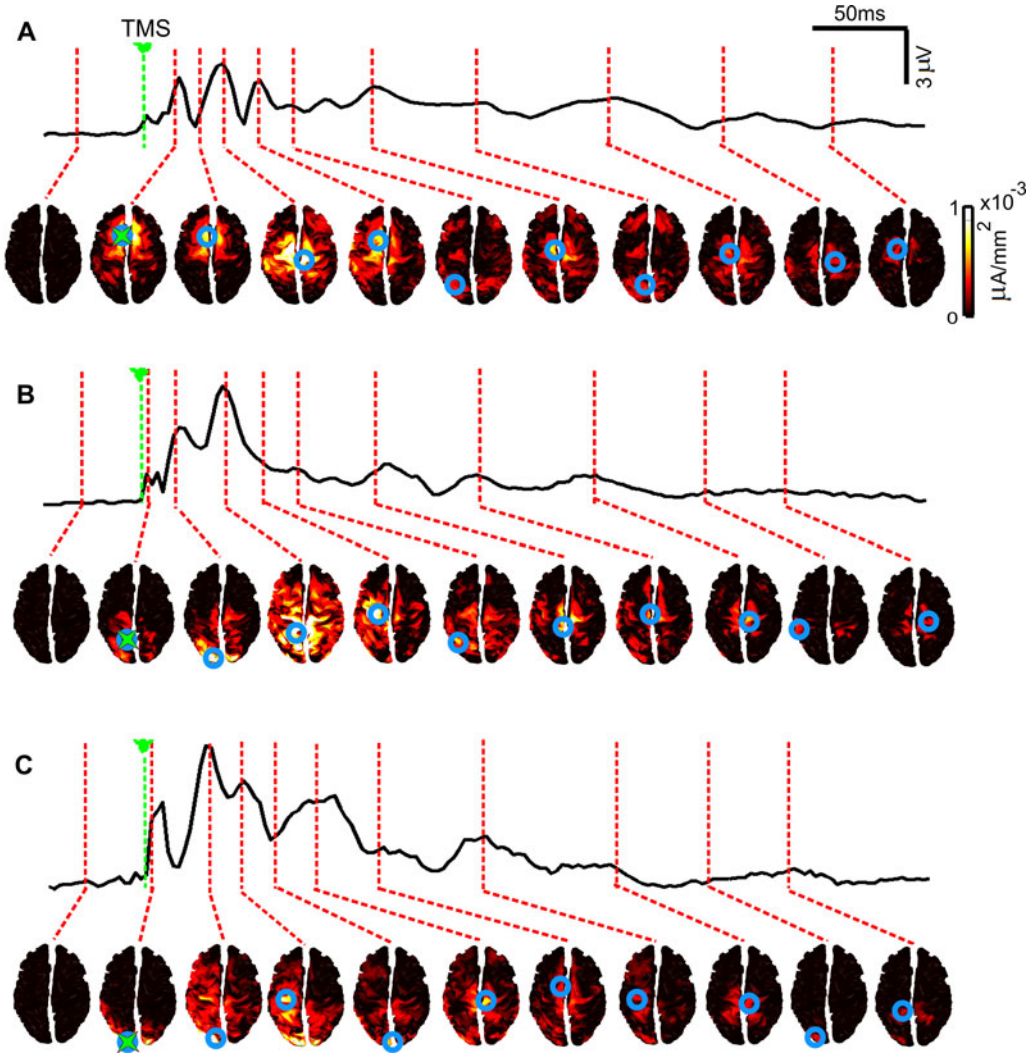

Figure 3.5: TMS evoked a balanced, long-range differentiated pattern of activation during wakefulness. hd-EEG GMFP (black traces) and SCD (cortical maps) are shown for a representative healthy subject during alert wakefulness with TMS (green star) target to BAs 6 (**A**), 7 (**B**) and 19 (**C**) at 100V/m. In all cases, TMS evoked a widespread activation pattern which lasted for more than 300ms and in which the maximum significant current (blue circle) shifted through different cortical areas at different latencies.

3.2 – Sleep

The most common situation in which the level of consciousness changes is early NREM sleep, when subjects, if awakened, report no or little conscious experience (Hobson and Pace-Schott, 2002), despite the fact that their brain remains highly active (Steriade et al., 2001). In a series of studies Massimini et al. (2005, 2007, 2010) described the TMS-evoked responses during the transition from wakefulness into different stages of sleep. While during wakefulness, TMS evoked an initial local cortical activation which invariably engaged distant cortical areas in a complex and differentiated way, the exactly same stimulation, applied 15 min later, during sleep stages 2 and 3, triggered a larger, low-frequency wave, associated with a strong initial cortical activation that did not propagate to connected brain regions, dissipating rapidly (Massimini et al., 2005). Increasing intensities of stimulation during NREM sleep may result in long-range bursts of cortical activity, but always associated to simple stereotypical and nonspecific responses (Massimini et al., 2007). Interestingly, during REM sleep, TMS triggered more complex patterns of cortical activation, resembling those observed in wakefulness (Massimini et al., 2010).

3.2.1 – *Slow-wave Sleep*

Figure 3.6 shows our results reproducing the experiment of Massimini et al. (2005, 2007) for a subject submitted to TMS targeted to the sensory-motor cortex at 90V/m. A first TMS-EEG session (180 stimuli) was collected while the subject was awake. Subject was then allowed to fall asleep and after entering a consolidated period (>5 min) of NREM sleep stage 2, a second TMS-EEG session was collected using the same stimulation parameters. As expected, during wakefulness, averaging the TMS-locked responses revealed low-amplitude, high-frequency waves of activity associated with cortical activations that propagate along long-range ipsilateral and transcallosal connections (Figure 3.6 A). The same stimulation however produced a slow, brief and local response during NREM sleep (Figure 3.6 B). This finding is general and can be reproduced after the stimulation of different cortical areas, as long as the subjects are in slow-wave sleep stages.

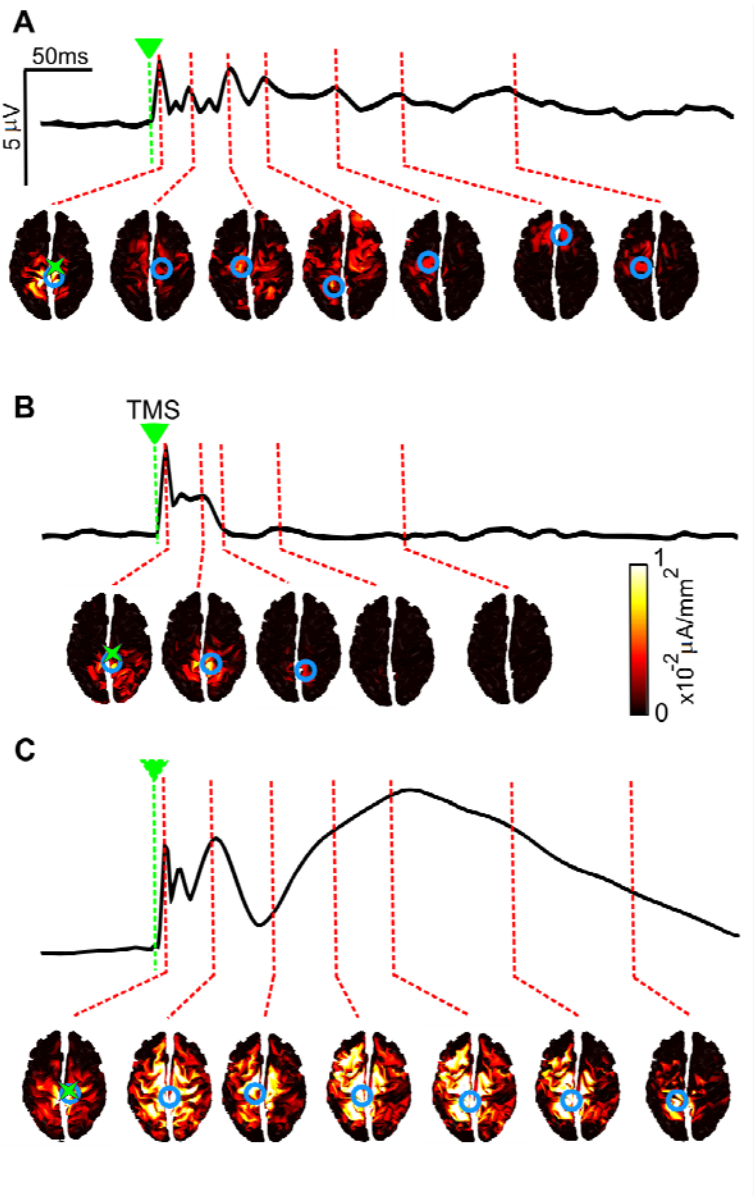

Figure 3.6: The balanced, long-range, differentiated pattern of activation observed in wakefulness was impaired during slow-wave sleep. hd-EEG GMFP (black traces) and SCD (cortical maps) are shown for a representative subject in whom the sensory-motor cortex was stimulated with TMS (green star). **A)** During waking, the stimulation intensity of 90V/m evoked an activation pattern lasting for more than 300ms and spreading from near the stimulation site to other cortical locations, recruiting different cortical areas at different latencies. **B)** During slow-wave sleep, the same stimulation evoked a response that remained local, fading shortly (<100ms). **C)** Higher stimulation intensity (160V/m) during slow-wave sleep evoked strong and global but yet stereotypical responses, in which the maximum significant current (blue circles) remained fixed near the TMS target.

Thus, TMS/hd-EEG revealed a clear-cut reduction of cortico-cortical integration occurring during sleep early in the night: while the cortical area that is directly engaged by TMS reacted to the stimulation, it generally behaved as an isolated module. Moreover,

TMS/hd-EEG measurements not only indicated that during slow-wave sleep the thalamocortical system tends to break down into isolated modules, but also showed that the ability of thalamocortical circuits to produce differentiated responses is impaired. We targeted TMS to premotor cortex (BA 6) and to the visual cortex (BA 19) of a healthy subject during wakefulness and slow-wave sleep. Figure 3.7 displays, for each condition, the significant current evoked by TMS as cumulated over the entire post-stimulus interval and plotted on the cortical surface; on the right side of each cortical surface, the time course of the currents recorded from three selected areas are depicted. This example, as the one reported in the previous figure, confirms a clear-cut loss of integration during slow-wave sleep by showing that distant cortical areas ceased to be causally affected by the initial perturbation. On the other hand, it also reveals a clear loss of response specificity. While during wakefulness cortical areas reacted to the stimulus with a specific activation pattern which had a characteristic shape and frequency content (section 3.1), this distinction was clearly obliterated during sleep; the local response to TMS became, in both cases, a stereotypical response: a positive wave followed by a negative rebound.

The stereotypical pattern of the response to TMS during slow-wave sleep was not caused by an effect of local activation thresholds. As described by Massimini et al. (2007), increasing the intensity of TMS up to saturation levels (160V/m) during slow-wave sleep did not produce a recovering of the balanced widespread activity observed in wakefulness. Instead, the slow positive–negative component may be amplified, developing towards a full-fledged sleep slow wave (figure 3.6 C).

Altogether, these TMS/hd-EEG measurements suggest that, during slow-wave sleep, the thalamocortical system, despite being active and reactive, either breaks down in causally independent modules or bursts into an explosive and non-specific response. In no case, during slow-wave sleep, does TMS result in a balanced, long-range, differentiated pattern of activation.

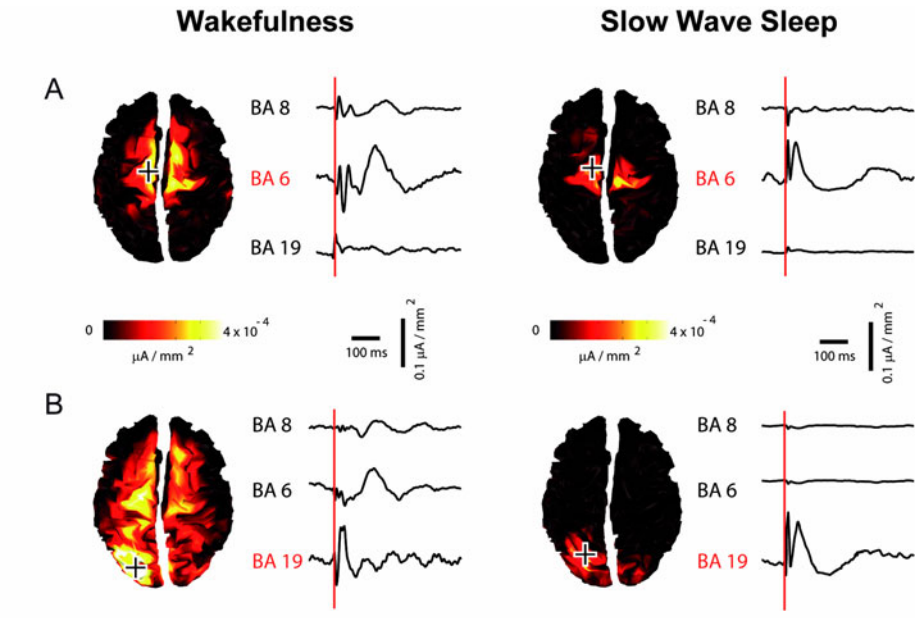

Figure 3.7: The ability of thalamocortical circuits to produce differentiated responses was impaired during slow-wave sleep. TMS was applied to premotor cortex (**A**) and to visual cortex (**B**) during wakefulness (the left panels) and during slow-wave sleep (the right panels). For each condition, the significant current distribution (SCD) recorded during the entire post-stimulus interval are plotted on the cortical surface. On the right side of each cortical surface, the time series of the currents recorded from three selected areas (BAs 8, 6 and 19) are depicted (the time of stimulation is marked by a red line). With the transition from wakefulness to slow-wave sleep, distant cortical areas ceased to be causally affected by the initial perturbation and cortical responses to TMS became stereotypical (adapted from Massimini et al., 2009).

3.2.2 – *REM Sleep*

Following Massimini et al. (2010), we were able to observe the gradual change of activity evoked by TMS during the transition from wakefulness through stage 1 to NREM (stages 2 and 3) and REM sleep. TMS was targeted to the rostral portion of the right premotor cortex with a maximum electrical field of 90V/m at the cortex surface while the subject, lying with eyes closed on a reclining chair, was allowed to sleep. After the experiment, trials were assorted by sleep stage, and four TMS sessions were constructed with a total of 258 trials during wakefulness, 274 at stage 1, 247 at stage 2-3 and 322 during REM sleep. Figure 3.8 displays how the widespread differentiated pattern observed during wakefulness (A) gradually shifted through sleep stage 1 (B) to a short-lived local and stereotypical response during stages 2 and 3 (C). Interestingly, during REM sleep (D), late in the night, when dreams become long and vivid and the level of consciousness is likely to return to levels close to

wakefulness (Kahan and Laberge, 2011), thalamocortical integration partially recovered and TMS triggered a more widespread and differentiated pattern of activation (Massimini et al., 2010).

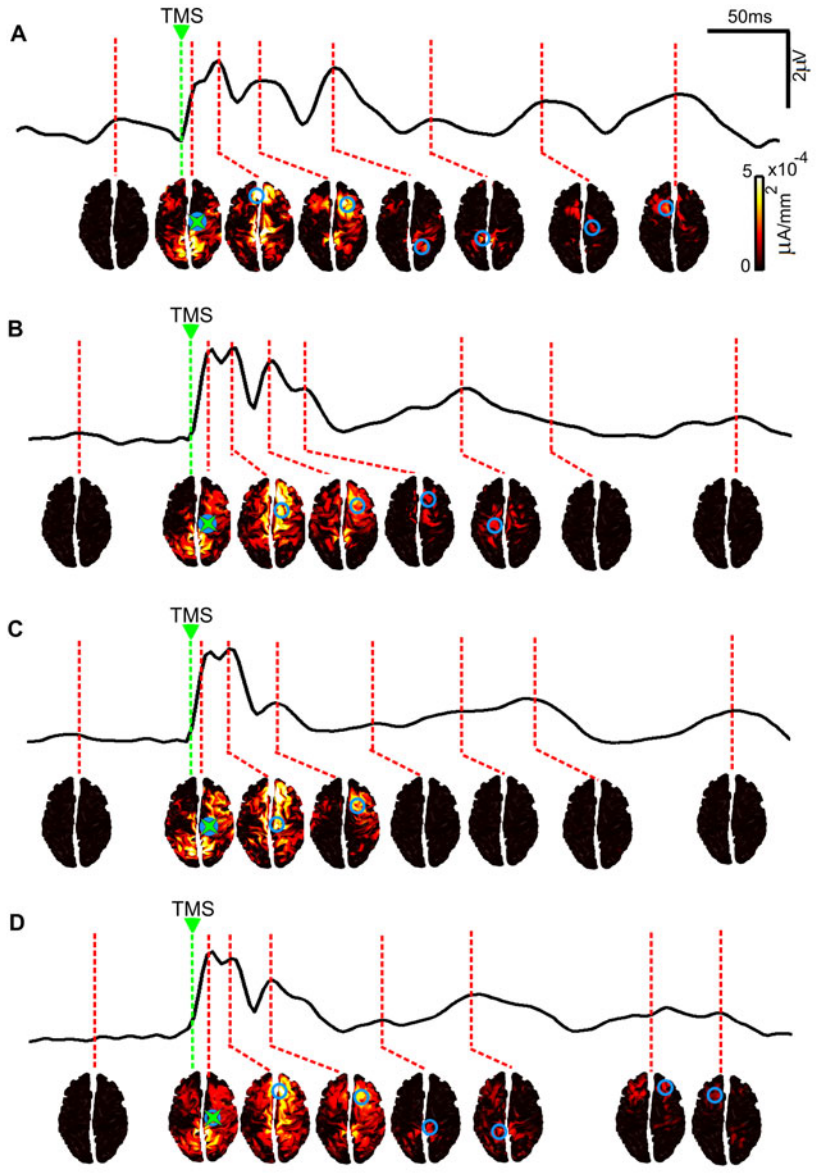

Figure 3.8: The widespread differentiated pattern of TMS-evoked activation observed in wakefulness was gradually impaired upon falling into NREM sleep and partially recovered during REM sleep. hd-EEG GMFP (black traces) and SCD (cortical maps) are shown for a subject in whom the premotor cortex was stimulated (green star) while transiting from wakefulness (**A**) through sleep stage 1 (**B**), NREM sleep stages 2-3 (**C**) to REM sleep (**D**). Duration and spreading of TMS-evoked currents observed during wakefulness were lost in sleep stages 2-3 but recovered during REM sleep.

3.3 – Anesthesia

In addition to sleep, the most common condition in which consciousness can be lost is general anesthesia. Although several anesthetics can induce states with behavioral and electrophysiological features not unlike those of deep NREM sleep, pharmacological anesthesia and sleep are not identical and differ in terms of both neurophysiology and neurochemistry (Van Dort et al. 2008). Moreover, general anesthesia offers several advantages for investigating the neural correlates of LOC (Franks, 2008). Specifically, in sleep studies it is not feasible to evaluate an individual's level of alertness repeatedly and reliably, because the depth of sleep varies unpredictably and subjects awakened to assess consciousness cannot rapidly return to sleep. By contrast, during general anesthesia, a subject's level of alertness may be assessed repeatedly without reversing the pharmacologically induced LOC.

In Ferrarelli et al. (2010) we described the patterns of TMS-evoked responses during loss of consciousness induced by anesthesia. In this first study of TMS/hd-EEG during anesthesia-induced LOC we chose to use the benzodiazepine midazolam, an agent that has a marked anticonvulsant effect, since TMS can, although rarely, induce seizures in epileptic patients. Also, the clinical tool used to evaluate the subjects' alertness throughout the experimental procedure, the Observer's Assessment of Alertness/Sedation (OAA/S) scale, was initially tested in subjects who received titrated doses of midazolam (Chernik et al., 1990). Finally, unlike most anesthetic agents midazolam targets $GABA_A$ receptors exclusively, leading to increased inhibitory postsynaptic currents that presumably underlie its behavioral/cognitive effects (Tanelian et al., 1993). Other general anesthetics, such as volatile and i.v. drugs, have effects that are more difficult to interpret because of multiple interactions with several proteins, such as voltage-gated and leak channels (Verbny et al., 2005). One potential drawback of midazolam is its pharmacokinetic profile, which leads to slower recovery compared with shorter-acting induction agents (Gan, 2006). Because of this slow recovery, we were able to measure TMS-evoked EEG responses after full recovery of vigilance (Level 5 OAA/S) in just one subject within the limited time frame of the TMS/hd-EEG recordings.

3.3.1 – *Experimental Protocol*

Six male subjects received i.v. midazolam at doses up to 0.2mg/kg, with OAA/S reaching scores of "1" (unresponsive to verbal and mild physical stimulus) for a sufficient period that hd-EEG responses to TMS could be measured. A 20-gauge i.v. catheter was placed for anesthetic drug delivery, and participants were given supplemental oxygen at 3 L/min via nasal cannula and anantacid (Bicitra) to minimize possible complications in the event of nausea and vomiting caused by the anesthetic drug. During the TMS procedure, the participant's electrocardiogram, noninvasive blood pressure, SaO_2, exhaled CO_2, and axillary skin temperature were continuously monitored by an anesthesiologist. Additionally, the subject's level of consciousness was evaluated before and after each TMS session with the OAA/S.

TMS was targeted to the right premotor cortex and a first 8- to 10-min TMS/hd-EEG session (~250 stimuli, with a 2,000-ms period and a ±250-ms jitter) was collected in each subject before midazolam injection (level5 responsiveness of the OAA/S). Midazolam was then given at an initial dose of 0.1mg/kg, followed by additional doses of 0.02mg/kg each 2–3 min until the subject was unresponsive (level 1 of the OAA/S), up to a maximal dose of 0.2 mg/kg. During midazolam administration, 3-min TMS blocks at 0.2 Hz interleaved by alertness assessments with the OAA/S were performed and a longer TMS session mirroring the preinjection TMS session was collected.

3.3.2 – *Results*

Compared with wakefulness, we found a marked change in TMS-evoked brain responses during midazolam-induced LOC (Figure 3.9). Before the injection of the anesthetic (level 5 alertness, OAA/S), TMS pulses to premotor cortex evoked a complex spatiotemporal pattern of low-amplitude, fast-frequency scalp waves, associated to cortical currents which lasted for at least 300ms following the stimulation and shifted among cortical areas distant from the TMS-targeted brain area (panels A and B). Conversely, following midazolam-induced LOC (level 1 OAA/S), TMS pulses gave rise to high-amplitude, low-frequency EEG voltages, produced by cortical currents which faded within 150ms and remained more localized to the stimulated site (panels A' and B').

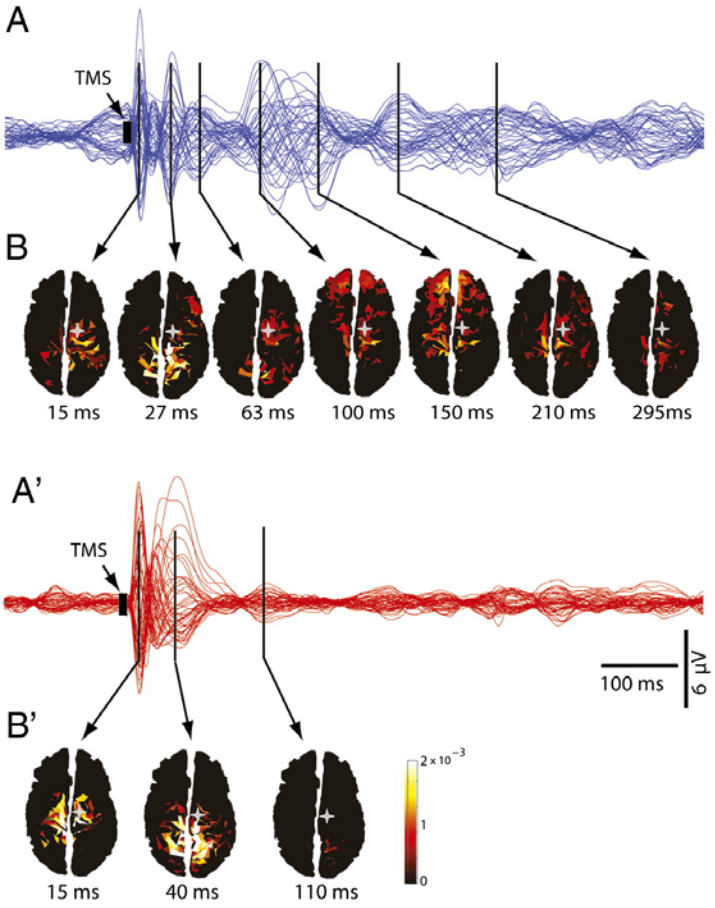

Figure 3.9: The balanced, long-range, differentiated pattern of activation observed in wakefulness was impaired during anesthesia-induced LOC. A and **A′**: Averaged TMS-evoked potentials at all electrodes, superimposed in butterfly plots (the blue traces for waking, red traces for anesthesia). **B** and **B′**: Cortical currents calculated on individual cortical meshes are shown from minimal (dark red) to maximal (white) values. During wakefulness, TMS of premotor cortex determined low-amplitude, complex scalp waves corresponding to cortical currents that lasted >300 ms and shifted among distant cortical areas. Conversely, during anesthesia, TMS gave rise to high-amplitude, short-lasting scalp voltages reflecting cortical currents that remained local, and faded within 150 ms. The gray stars represent the TMS target (premotor cortex); the black arrows represent the local maxima in periods of significant TMS-evoked activation (adapted from Ferrareli et al., 2010).

Although the activity evoked by TMS at the stimulation site, the premotor cortex (BA 6), was similar across conditions, its time course was markedly different. While during wakefulness, the immediate response to TMS consisted of oscillations on the β range, during anesthesia, the oscillatory frequency of the premotor cortex was altered and TMS-evoked slower responses, consisting of a large positive wave followed by a negative deflection. This local slow response was found in all subjects during LOC. Moreover, the initial response

remained largely restricted to the premotor cortices and affected only marginally the activity of other cortical areas. This breakdown of cortico-cortical effective connectivity was also evident when inspecting the time courses of the TMS-evoked cortical currents on other areas. For instance, the time course of cortical currents in BA 8 (prefrontal cortex), which is anatomically connected to BA 6 (premotor cortex), was not affected by the strong activation evoked by TMS in the premotor cortex (Figure 3.10).

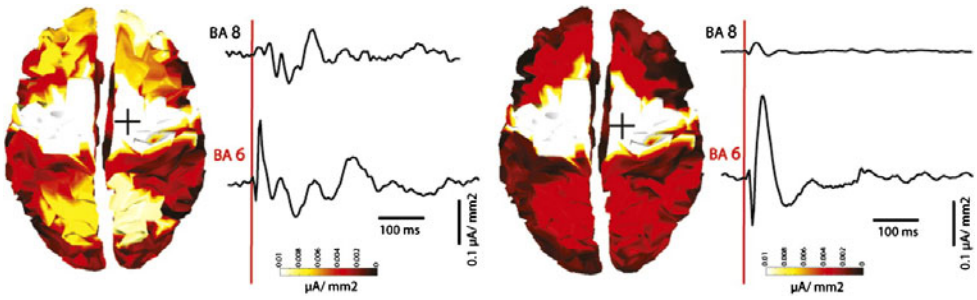

Figure 3.10. TMS during anesthesia evoked a large positive–negative wave in the stimulation site but little activation in distant areas. SCD cumulated in a 0–500 ms post-TMS interval and displayed on the corresponding BA in wakefulness (Left) and anesthesia (Right). To the right of each topographic plot are time courses of currents recorded from the stimulated area, premotor cortex (BA 6), and from a more anterior cortical area (BA 8). During anesthesia, TMS-evoked SCD in BA6 was similar to the SCD recorded in wakefulness, as reflective of an initial stronger but shorter-lived response during anesthesia compared to wakefulness. Conversely, SCD from BA 8, which is anatomically connected to BA 6, were markedly reduced in anesthesia compared to wakefulness, suggesting a marked decrease in cortical effective connectivity (adapted from Ferrareli et al., 2010).

These results bear a striking resemblance with those obtained during early NREM sleep discussed in the precedent section. The depth of sleep, however, can vary unpredictably, and if awakened to assess consciousness, subjects cannot rapidly return to sleep. In this study, subjects could be repeatedly assessed for LOC. In one subject, we could also evaluate the effects of progressively reduced arousal, from level 3 (sedation) to level 1 (LOC). The results show that at sedation level, the TMS-evoked initial activity became stronger than in wakefulness (Figure 3.11) but was followed by smaller and shorter-lived oscillations. This initial response to TMS was even larger during LOC, with a positive–negative wave similar to the spontaneous sleep slow oscillation, while subsequent activity was obliterated, demonstrating that brain responses to TMS became progressively shorter and sleep-like while transitioning into pharmacologic LOC.

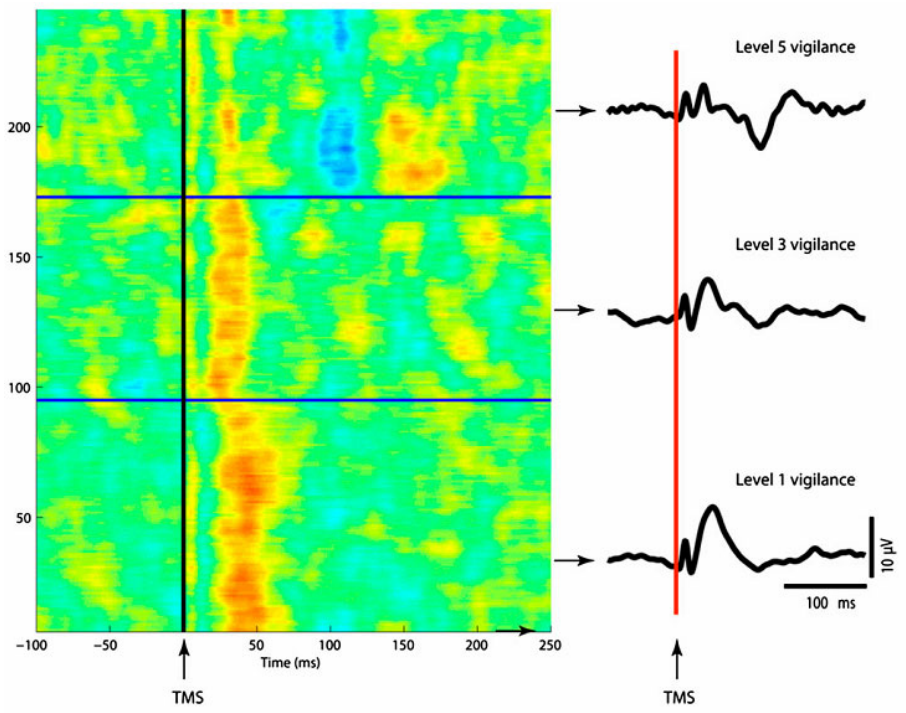

Figure 3.11. TMS-evoked brain responses gradually changed while transitioning from high (OAA/S = 5) to low (OAA/S = 1) levels of vigilance. Left: single trials recorded from one channel located under the TMS coil, color-coded for voltage. Right: averaged TMS-evoked responses obtained during the three levels of vigilance. Both single and average EEG responses showed a progressive increase in the amplitude and latency of an early evoked component (positive peak), followed by the obliteration of succeeding oscillations when reaching low levels of vigilance (adapted from Ferrareli et al., 2010).

In order to further quantify changes in strength (activity) and propagation (connectivity) of TMS-evoked cortical responses during LOC, SCD and SCS were calculated for each subject in wakefulness and anesthesia. The time course of SCD revealed that, in each subject, the initial TMS-evoked cortical activity, related to the large positive–negative wave, was higher in the anesthesia condition, whereas subsequent cortical activity was stronger during wakefulness. When cumulating SCD in two post-TMS time ranges, respectively, 0–50 and 50–500 ms, we found that in the 0–50 ms interval, SCD was significantly higher during anesthesia (p = 0.016, Mann–Whitney), whereas in the 50–500 ms range, SCD was significantly higher during wakefulness (p = 0.016, Mann–Whitney). The average SCD in the entire poststimulus interval was reduced during anesthesia, indicating a diminished response to the TMS, although the two conditions differed only at trend level (p = 0.1, Mann–Whitney). By contrast, SCS discriminated effectively between the two conditions (Figure 3.12). Specifically, in each subject, mean SCS was significantly higher in wakefulness compared to anesthesia (p = 0.009, Mann–Whitney). Furthermore, the time course of SCS

showed that, in each subject, SCS during wakefulness was significantly increased compared to baseline for >200ms, whereas during anesthesia, SCS returned to baseline within 100 ms of TMS.

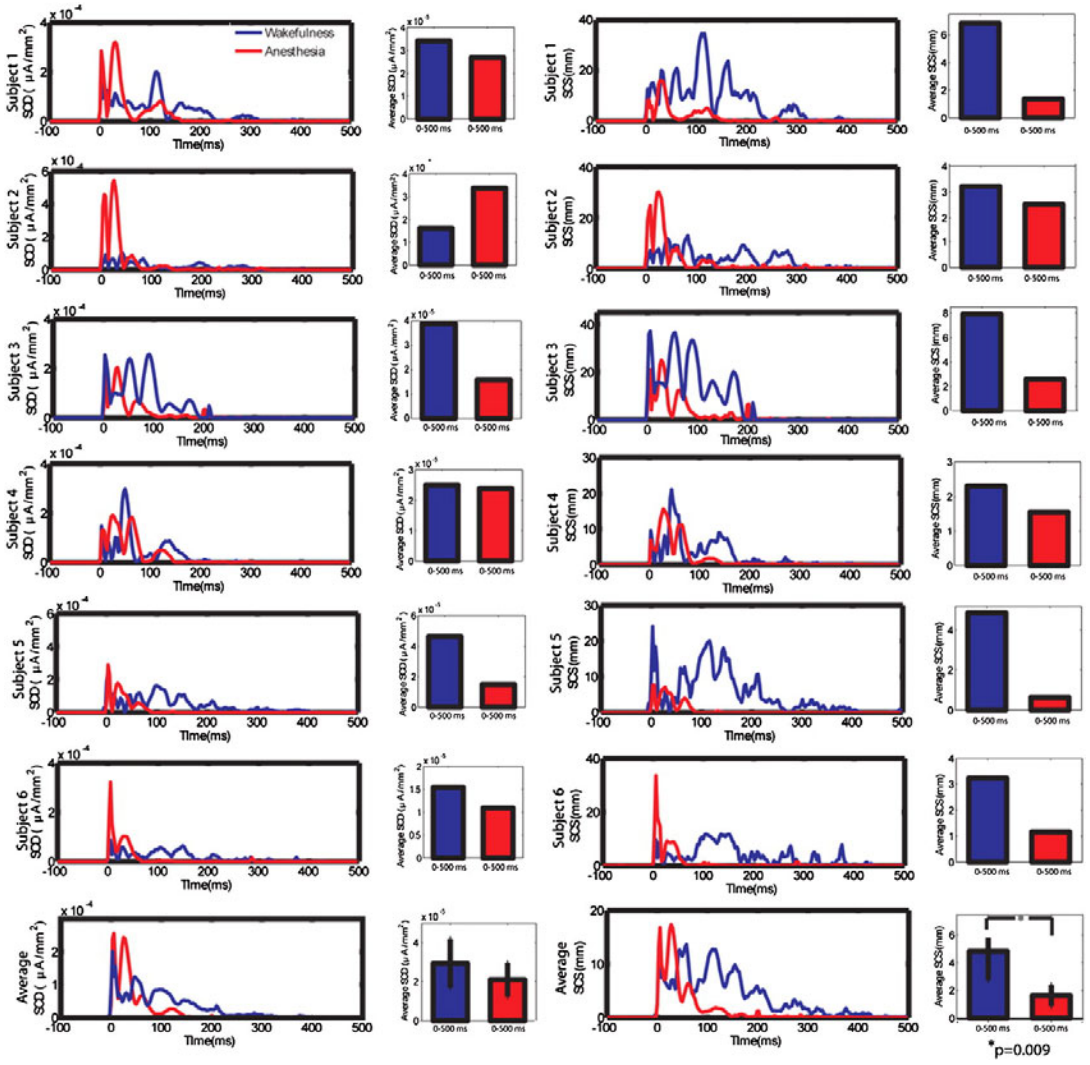

Figure 3.12: Cortical connectivity (SCS), but not reactivity (SCD), captured cortical changes during LOC. SCD and SCS were computed for each subject in wakefulness (blue line) and anesthesia (red line) following TMS of the premotor cortex. Left: Time course of SCD for individual and average data, and mean SCD over the entire post-TMS time interval (0–500ms). In each subject, SCD values were initially higher (first 50ms after TMS) during anesthesia but tended to dissipate shortly thereafter, consistent with a TMS-evoked larger initial response during anesthesia that was, however, short-lived. Mean SCD over the entire post-TMS period were not significantly different between wakefulness and anesthesia. Right: Time course of SCS for individual and average data, and mean SCS over the entire post-TMS interval. In each subject, during wakefulness, SCS was present for>200ms, whereas during anesthesia it faded after 100ms. Notably, mean SCS values in the 0–500ms post-TMS interval were significantly higher (p = 0.009, Mann–Whitney) in wakefulness relative to anesthesia, (adapted from Ferrareli et al., 2010).

Thus, during loss of consciousness induced by midazolam, the effective connectivity in the corticothalamic system was markedly reduced. Moreover, specific local oscillatory properties of neural circuits, as the natural frequencies of corticothalamic circuits of healthy awake subjects, were lost under LOC. Altogether these results suggest that when consciousness is absent the thalamocortical system responds to TMS with qualitatively simpler, slow and non-specific patterns of activation.

3.4 – Disorders of Consciousness

The results exposed in the previous sections demonstrate that in healthy awake and alert subjects TMS induced a sustained EEG response involving the sequential activation of different brain areas and affecting much of the cortex. By contrast, during NREM sleep and after loss of consciousness induced by general anesthesia, TMS pulses invariably produced a simple stereotyped response that either remained localized to the site of stimulation or bursted into an explosive and non-specific activation. Moreover, during REM sleep, when subjects are unresponsive to sensory stimuli and virtually paralyzed but report vivid dreams upon awakening, the cortical response to TMS recovered its complexity and became similar to that observed during wakefulness. If the complexity of the responses to TMS is supposed to correlate to the level of consciousness, what should we expect from the patterns of TMS-evoked cortical activity collected in patients suffering from disorders of consciousness?

3.4.1 – *TMS/hd-EEG in DOC Patients: Experimental Protocol*

In Rosanova et al. (2012) we described the results of employing TMS/hd-EEG at the bedside of 17 patients who evolved from coma into different clinical states: vegetative state (VS), minimally conscious state (MCS), emergence from MCS (EMCS) and locked-in syndrome (LIS). During each TMS/hd-EEG session, patients were lying on their beds, awake and with their eyes open. If signs of drowsiness appeared, recordings were momentarily interrupted and subjects were stimulated using the CRS-R arousal facilitation protocols (Giacino et al., 2004). Throughout every recording session the stability of stimulation coordinates was continuously monitored. If the virtual aiming device was signaling a

displacement >4 mm, session was interrupted and the coil was repositioned. At the end of the experiment, the stimulation coordinates were recorded and the electrode positions were digitized.

Cortical targets were identified on computerized tomography (CT) scans acquired with a Siemens Senatom Sensation 16 from all patients. Stimulation was delivered with an interstimulus interval jittering randomly between 2000 and 2300 ms (0.4-0.5 Hz), at an intensity ranging from 140 V/m up to 200 V/m on the cortical surface. The CT-guided intracranial electric field estimation was a crucial step during the experimental procedure; due to shifts of intracranial volumes in brain-injured patients, it is difficult to assess whether TMS is on target and effective based on extra-cranial landmarks alone and this may result in false-negatives (absence of EEG response due to missed target, or sub-threshold stimulation).

TMS was targeted to four cortical sites: the left and right medial third of the superior parietal gyrus and the left and right medial third of the superior frontal gyrus. These cortical targets were selected for several reasons: i) they are easily accessible and far from major head or facial muscles whose unwanted activation may affect EEG recordings; ii) the posterior parietal cortex, as well as its interactions with more frontal areas, is thought to be particularly relevant for consciousness (Laureys et al., 2004); iii) our previous TMS/EEG studies have been successfully performed in these areas during wakefulness (Rosanova et al., 2009), sleep (Massimini et al., 2005, 2007) and anesthesia (Ferrarelli et al., 2010). In practice, all four cortical sites were not always accessible in all subjects due to skull breaches, external drain derivations. In all cases, we avoided stimulating over focal cortical lesions that were clearly visible in CT scans, since the EEG response of these areas may be absent or unreliable.

3.4.2 – *Data Analysis and Results*

We performed a first set of TMS/hd-EEG experiments (one single session) in a group of 12 patients (**Group I**: 5 females; mean age±standard deviation: 50.3±26.21, for more details see Appendix - Table A.1). These patients were repeatedly evaluated (4 times, every other day) for a period of one week by means the CRS-R, in order to avoid diagnostic errors due to fluctuations in responsiveness and to obtain a stable clinical diagnosis. Five patients of **Group I** (Patients 1, 2, 3, 4 and 5), showed only reflexive behavior and were diagnosed as VS

during the 4 behavioral evaluations. Five patients (Patients 6, 7, 8, 9 and 10) showed non-reflexive behaviours, such as visual tracking or responding to simple commands, satisfying the CRS-R criteria for MCS in at least 3 evaluations, including the one performed on the day of the TMS/EEG session. The two remaining patients (Patient 11 and Patient 12) could communicate reliably and were diagnosed as LIS. The VS and MCS subgroups did not differ systematically in etiology and time from injury (reported in Appendix - Table A.1). In particular, Group I included three chronic patients, one VS (Patient 5: 172 days from injury), one MCS (Patient 8: 1334 days from injury) and one LIS (Patient 12: 1399 days from injury).

Figure 3.13 displays examples of the cortical response evoked by TMS in VS, MCS and LIS patients. VS patients exhibited local, stereotypical evoked responses, closely resembling the cortical activations previously observed during deep sleep and anesthesia (Figures 3.6 and 3.9). In MCS, TMS invariably triggered a complex response associated with a rapidly changing pattern of cortical activation, contrasting starkly with the local, simple wave recorded in VS patients and being, instead, comparable to the one obtained in locked-in subjects. Thus, a clear-cut difference in the cortical response to TMS seems to discriminate reliably between patients who were considered VS after repeated behavioral evaluations and subjects who showed signs of consciousness.

In order to quantify this difference, we applied a statistical measure to extract latencies with significant GMFP activations and count the number of sources involved by maximal currents at these latencies. A bootstrap method (Lv et al., 2007, McCubbin et al., 2008) was applied by shuffling the time samples of GMFP pre-stimulus activity (from -300 to -50 ms) at the single-trial level and by calculating 500 surrogated pre-stimulus GMFPs. From each random realization, the maximum value across all latencies was selected to obtain a maximum distribution (control for type I error) and significance level was set at $p<0.01$. At each significant latency of the post-stimulus GMFP, the location of maximum neuronal current (10 most active sources) was detected on the cortical surface. Plotting and counting the sources involved by maximum neuronal currents across all significant time points in the first 300 ms post-stimulus resulted in the cortical maps and in the values reported in Figure 3.14.

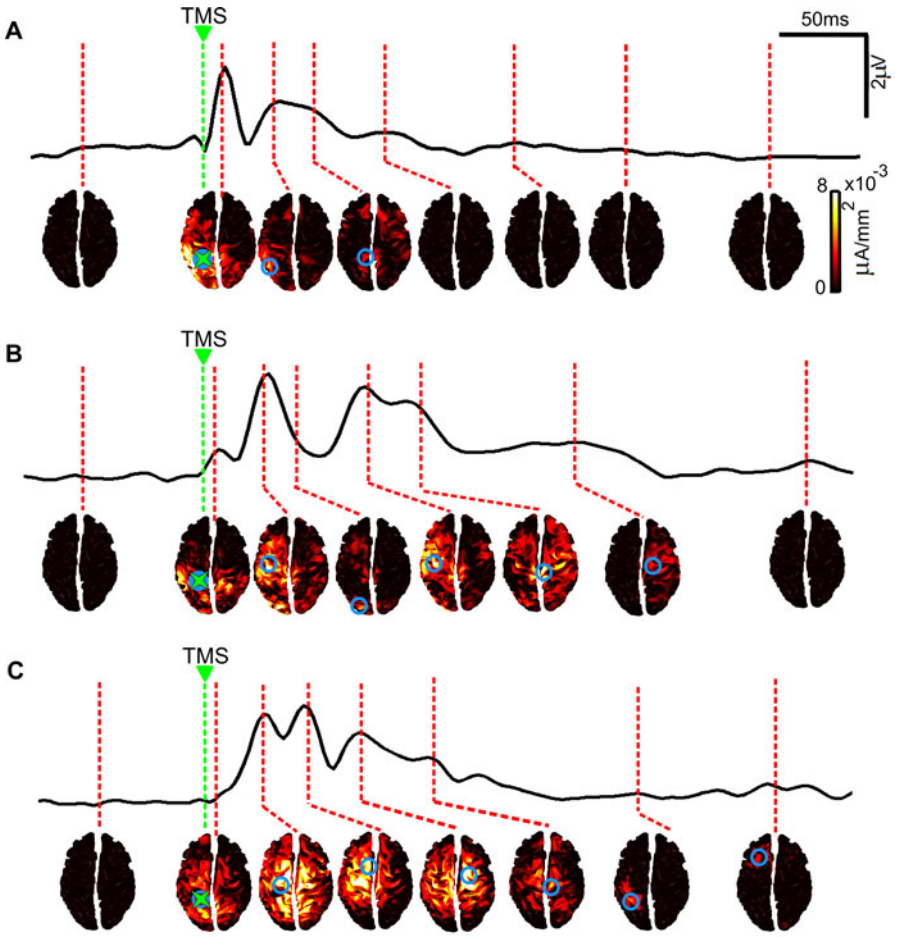

Figure 3.13: Examples of TMS-evoked cortical responses of patients who evolved from coma into different clinical states. hd-EEG GMFPs (black traces) and SCD (cortical maps) are shown for TMS/EEG sessions performed in three patients in different clinical states. **A)** TMS evoked a slow, short-lived and simple response on a patient diagnosed in vegetative state (CRS-R = 4); **B)** A patient who showed fluctuating signs of non-reflexive reactions to external stimuli but were unable to communicate reliably with the examiners, diagnosed in minimally conscious state (CSR-R=14), responded to TMS with a complex, long-lasting changing pattern of cortical activation, where maximum neuronal currents (blue circle) shifted over time across different areas; **C)** A similar complex response was observed in a locked-in subject (LIS) who, though being largely paralyzed at the time of recording, could signal full awareness through vertical eye movements. The green stars mark the site of stimulation (left parietal cortex).

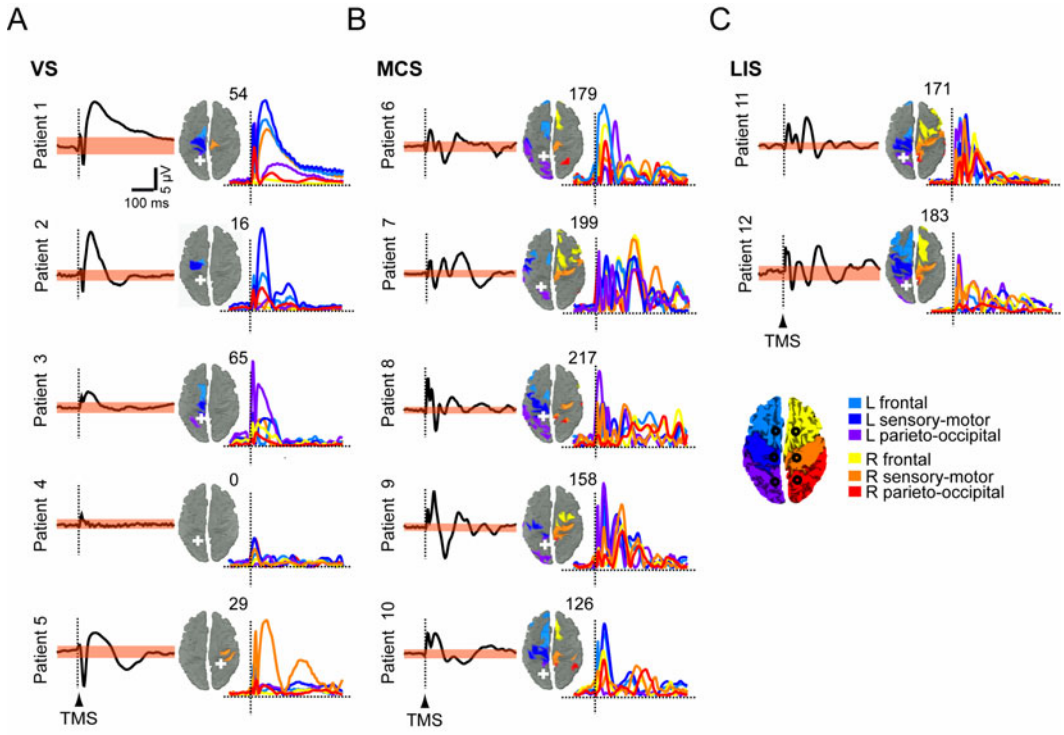

Figure 3.14. TMS-evoked cortical responses in Group I patients. A group of five VS (**A**), five MCS (**B**), and two LIS patients (**C**) underwent one TMS/hd-EEG session after 7 days of repeated evaluations by means of the CRS-R. For each patient, the averaged TMS-evoked potentials recorded at one electrode under the stimulator (black trace) and the respective significance threshold (upper and lower boundaries of the pink bands; bootstrap statistics, $p<0.01$) are shown. The sources involved by maximum cortical currents (10 most active sources) during the significant post-stimulus period of the GMFP are plotted on the cortical surface and color-coded according to their location in six anatomical macro-areas as indicated in the legend at the bottom-right end of the figure; the number of detected sources is indicated at the top-right end of each map. The time-series represent TMS-evoked cortical currents recorded from an array of 6 sources (black circles on the bottom-right cortical map) located about 2 cm lateral to the midline, one for each macro-area. The white crosses mark the sites of stimulation. For all patients the responses to the left parietal cortex stimulation are shown, except for one patient (Patient 5) in whom a significant response could only be detected in the right hemisphere (adapted from Rosanova et al., 2012).

According to this procedure, the number of detected sources is small if TMS triggers stable primary neuronal currents that remain confined to the stimulated area during the entire post-stimulus period. On the contrary, the number of detected sources is large if TMS triggers primary cortical currents that involve different cortical areas at different times. In order to

describe the time-course of TMS-evoked cortical activation in different areas the currents from a grid of six cortical sources (black circles plotted on the bottom-right cortical map in Figure 3.14) were extracted and auto-scaled to the maximum value of each session. Sources and time-series of cortical currents were color-coded according to their anatomical location in 6 arbitrary macro-areas (see colors legend in Figure 3.14).

In all VS patients, except for one anoxic patient (Patient 4) in whom no response could be elicited even when TMS was delivered at high intensity (200 V/m) in both hemispheres, TMS elicited maximum cortical currents that remained localized during the entire significant post-stimulus period, involving a small number of sources around the stimulated area (Figure 3.14A). Conversely, in all MCS patients, maximum neuronal currents shifted over time from the stimulated site to a large number of distant sources (Figure 3.14B). This pattern was comparable to the one obtained in two locked-in (LIS) subjects (Figure 3.14C). Notably, this difference allowed for a single-subject discrimination between VS and MCS patients (see also Figure 3.17).

We also performed longitudinal TMS/hd-EEG measurements (session 1, session 2 and session 3) in a group of five patients (**Group II**: 3 females; mean age±standard deviation: 51.2±23.05, see the Appendix - Table A.2 for more detail) as they awakened from coma and progressed towards different clinical states (Figure 3.15). As assessed by the CRS-R, three of these patients (Patients 13, 14 and 15) regained functional communication evolving from VS, through MCS to EMCS, while two patients (Patients 16 and 17) remained VS. In all cases the first TMS/hd-EEG session (session 1) was performed at least 48 hours after withdrawal of sedation, when patients opened their eyes and were diagnosed VS. At this time, similar to the VS patients of Group I, TMS evoked a simple wave and a local pattern of activation or no response at all. Following session 1, two additional TMS/hd-EEG measurements were performed in the three patients who eventually recovered consciousness: session 2 was recorded as soon as they satisfied the CRS-R criteria for MCS and session 3 when they recovered functional communication and emerged from the minimally conscious state (EMCS). In these patients, TMS triggered a complex pattern of activation that sequentially involved a large set of cortical areas already during session 2; this response was substantially different from the simple, local activation of session 1 and was instead comparable to the one obtained in session 3, when subjects had recovered their ability to communicate (Figure 3.15A). In the two patients who did not show any clinical improvement beyond VS, a second

TMS/hd-EEG measurement (session 2) was performed more than a month after session 1 and showed either a local, simple wave of activation (Patient 16) or no response (Patient 17, anoxic), although subjects were awake and open-eyed when their brains were stimulated (Figure 3.15B).

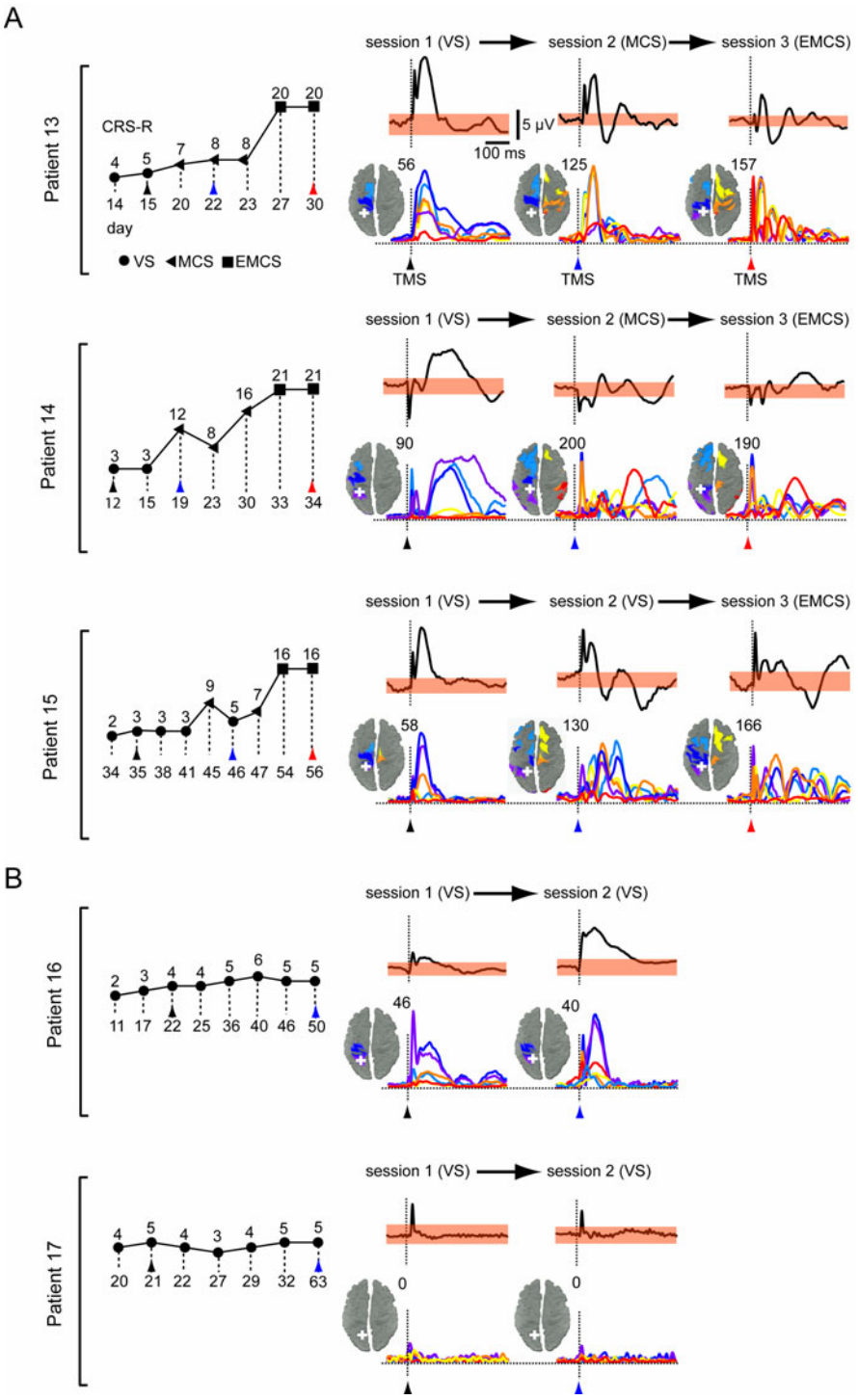

Figure 3.15. Clinical evaluation and TMS-evoked cortical responses in Group II patients. CRS-R total scores are plotted for the patients who were studied longitudinally (Group II) and eventually emerged from MCS (**A**) or remained VS (**B**); the first

assessment (session 1) was carried out 48 hours after withdrawal of sedation, as patients exited from coma. The symbols indicate the associated clinical diagnosis (circles=VS; triangles=MCS; squares=EMCS). Colored arrow tips mark the days when TMS/EEG recordings were performed and the time of TMS delivery (black=session 1; blue=session 2; red=session 3). For every patient and session, averaged potentials triggered by TMS (vertical dashed lines) of parietal cortex and recorded from the electrode under the stimulator are shown. The sources involved by maximum neuronal currents during the significant post-stimulus period are plotted on the cortical surface and color-coded according to their location in six anatomical macro-areas (see Figure 3.14); the number of detected sources is indicated at the top-right end of each map. The time-series represent TMS-evoked cortical currents recorded from an array of 6 sources (see their locations in Figure 3.14) located about 2 cm lateral to the midline, one for each macro-area. The white crosses mark the sites of stimulation; in each patient, the left parietal cortex was stimulated when patients entered VS from coma (session 1), soon after transition to MCS or at least 30 days of permanence in VS (session 2) and after emergence from MCS (EMCS; session 3), when subjects recovered functional communication (adapted from Rosanova et al., 2012).

These results with Group II indicate that the breakdown of the widespread balanced pattern of cortical activation observed in VS patients can be reversible and that a substantial improvement in the brain's ability to sustain internal communication occurred at an early stage during recovery of consciousness, before reliable communication could be established with the patient.

Are these changes in the activation pattern evoked by TMS associated to changes in electrophysiological arousal (EEG activation)? Spectral analysis of spontaneous EEG showed a consistent increase of high-frequency (> 7Hz) oscillations in LIS compared to MCS (Group I) and in EMCS compared to MCS (Group II). By contrast, in spite of a clear-cut change in the electrical response to TMS, no systematic changes of the background EEG could be detected between VS and MCS (Figure 3.16) consistent with previous reports (Kotchoubey et al., 2005). These results suggest that the transition from VS to MCS involves a substantial improvement of cortical integration that is not necessarily associated with an obvious change in the level of activation of the ongoing EEG.

Figure 3.17 summarizes the results obtained after applying TMS in all 17 patients, showing that it was possible to discriminate reliably between VS and MCS at the single-subject level. Crucially, this discrimination was achieved in a way that is completely independent on the patient's ability to exchange information with the surrounding environment.

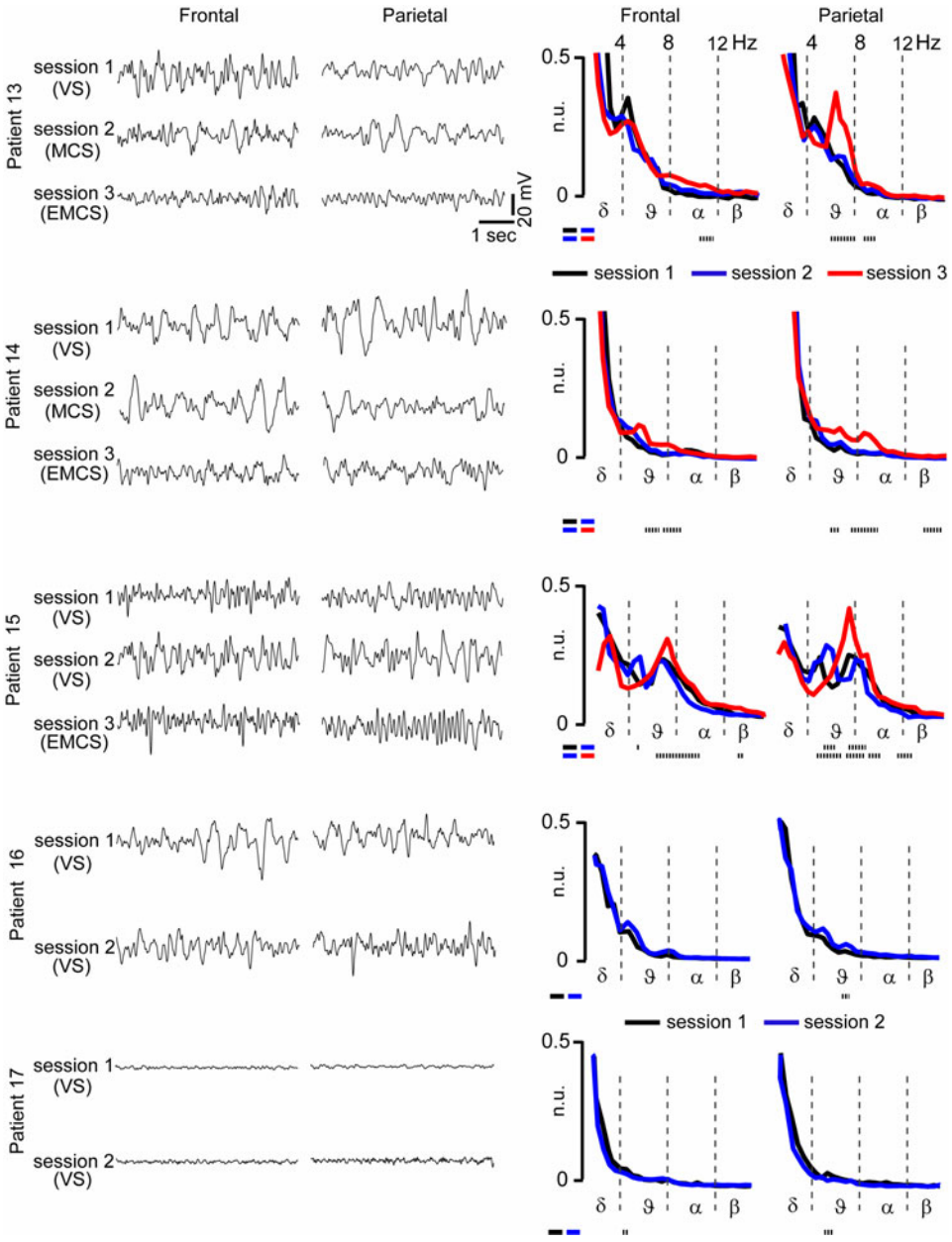

Figure 3.16. EEG spectra showed evident changes from MCS to EMCS but not from VS to MCS. Spontaneous EEG traces (5 seconds) and EEG spectra (calculated on 2 minutes) are shown for the 5 subjects who underwent longitudinal recording sessions (Group II); in these patients, changes in the EEG spectrum were assessed statistically by means of a two tailed paired t-test. The dotted lines at the bottom of each plot indicate the frequency bins that show statistically significant differences of power (t-test, $p<0.01$) (adapted from Rosanova et al., 2012).

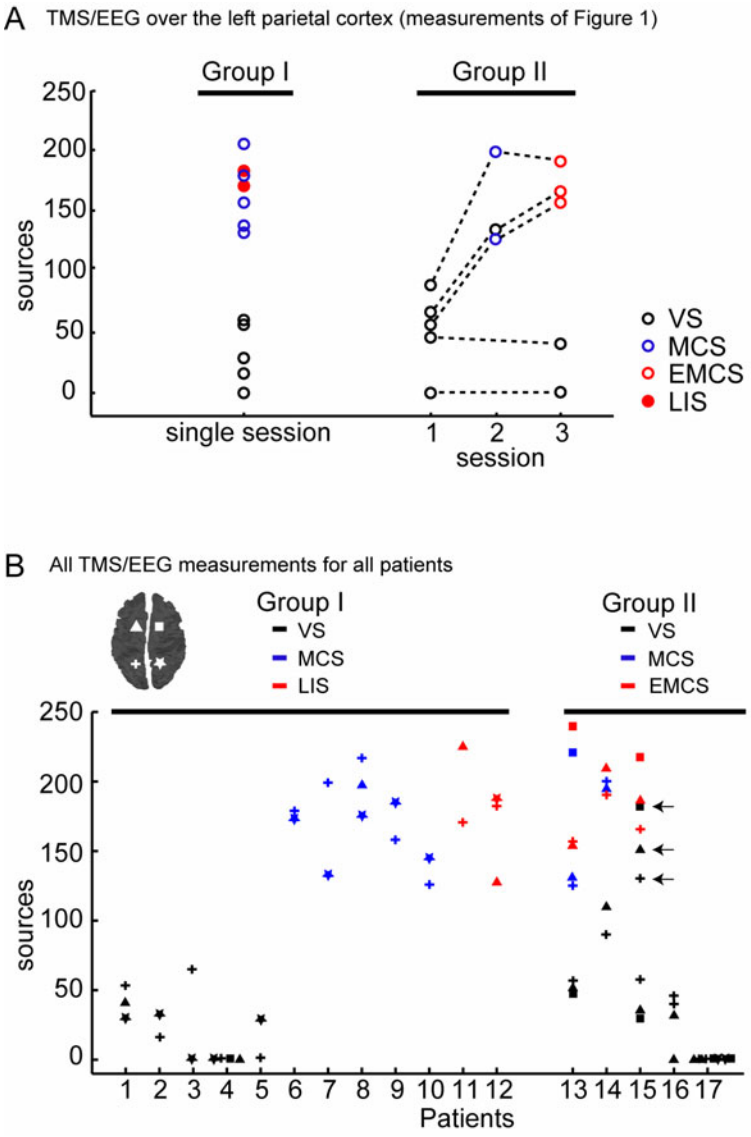

Figure 3.17. Effective connectivity for all patients and TMS/EEG measurements. A) For each patient and TMS/EEG measurement (parietal stimulation, same measurements of Figures 3.14 and 3.15), the number of sources involved by TMS-evoked currents are plotted. The circles indicate the clinical diagnosis at the time of recording (open black circles for VS; open blue circles for MCS; open red circles for EMCS and filled red circles for LIS). **B)** The number of cortical sources involved by maximum cortical currents detected in all TMS/EEG measurements (n=72) is plotted for all patients (Group I on the left and Group II on the right). Each value refers to one cortical target and is marked according to both the site of stimulation (the correspondence between symbols and stimulation sites is graphically reported on the cortical map in the left upper corner of the panel) and the CRS-R diagnosis at the time of recording (black for VS; blue for MCS; red for LIS in Group I and EMCS in Group II). In all cases, effective connectivity was higher in patients who showed some level of consciousness (MCS, EMCS and LIS) compared to VS patients. An exception is represented by the 3 measurements (left parietal, left frontal, right frontal) performed in Patient 15 during session 2 (open black circles indicated by arrows). This patient was diagnosed MCS the day before the measurement, slipped back into a behavioral VS on the day of session 2 and, within days, was reassessed clinically as MCS and then EMCS (during session 3). Effective connectivity was null in the two anoxic subjects (Patients 4 from Group I and Patient 17 from Group II) (adapted from Rosanova et al., 2012).

The quantitative measure employed in this section - the number of different sources involved in the maximum neural current at latencies where the GMFP was statistically significant - was able to detect a clear-cut difference between VS and MCS. This is a particularly relevant distinction if one considers that the most challenging task at the bedside is distinguishing between VS patients and non-communicating MCS patients (Majerus et al., 2005). However, this measure, designed to grasp the increase of complexity in the cortical activation pattern evoked by TMS in MCS patients (Figure 3.13A), was not able to quantify distinctions between lower versus higher levels of consciousness, such as between MCS and patients suffering from locked-in syndrome. We should therefore expect that a more suitable measure of the complexity of the cortical activity evoked by TMS could be devised. In particular, as we saw in the previous two chapters, a measure of the brain's capacity for sustaining causal interactions between specialized cortical areas - a measure of its capacity for internal communication - should rely on statistics constructed from single-trial sources to take into account only primary currents causally related to the perturbation. Volume conduction may cause GMFP to be significant when primary currents are not. For this reason, the measure employed in the previous chapter may overestimate the ability of the thalamocortical system to sustain internal communications in some intermediate conditions, such as MCS. Whereas this bias may favor the sensitivity of the measure for detecting differences from VS, it may also undermine the distinction between MCS and healthy, alert subjects. Moreover, there is no indication that the ability for internal communication, which is supposedly relevant for consciousness, should depend only on those currents that are most strongly activated at a particular time. If the activity of a cortical area is causally linked to the stimulus, it should contribute to a measure that is sensitive to the amount of internal communication in the thalamocortical system, independent of the strength of its corresponding evoked current.

Therefore, although the results presented thus far and the phenomenological considerations discussed in the first chapter qualitatively indicate that the brain's capacity for consciousness depends on its ability to sustain complex patterns of causal interactions between specialized cortical areas, there is still a need for a quantifiable measure of such a peculiar capability. We address this task in the next chapter.

Chapter 4.

The Perturbational Complexity Index

If the brain is supposed to be the vehicle for the content of consciousness, if it is responsible for sustaining our access to the objects of intentionality, then, as we argued in the first chapter, different specialized cortical areas should be allowed to causally interact with each other through complex feedforward-feedback circuitry. Our results in the previous chapter strongly support the notion that when this capacity to sustain causal interactions is present, the thalamocortical system should respond to an effective perturbation with a complex pattern of neural activation, involving different cortical areas at different times. Conversely, the system should react with a simple stereotypical response if the circuitry is no longer integrated or if any specialized module has lost its dynamical specificity. How can we quantify the complexity of the cortical activation pattern evoked by TMS?

4.1 – Calculating the complexity of the cortical activity evoked by TMS

We begin by using NREM sleep as a model of unconsciousness and developing a measure of the complexity of cortical responses to external perturbation. TMS evoked simple stereotypical cortical responses during NREM sleep and a balanced, long-range and differentiated activation pattern during wakefulness, with different cortical areas responding to the stimulation at different times (Figure 3.6). This qualitative notion of the spatiotemporal

complexity of the activation pattern evoked by TMS can be understood from the significant sources binary matrix (**SS**(***x***,*t*), section 2.4), which indicates where and when cortical activity is significantly related to the stimulation. Indeed, observing the **SS** matrix plots (Figure 4.1) calculated for the conditions displayed in Figure 3.6 immediately reveals that the complexity of the **SS** matrices, rather than the strength, duration or extent of the response to TMS, is what distinguishes alert wakefulness (Figure 4.1 A), when consciousness is unambiguously present, from slow-wave sleep, when consciousness fades (Figure 4.1 B and C).

The **SS** matrix displayed in Figure 4.1A is intuitively said to be more complex in that it is more difficult to describe: significant activity occurred at different times and at different places following no simple rule. By contrast, **SS** matrices exhibited in Figure 4.1 B and C are less complex in that they are easier to portray. In general, there is a clear and intuitive relationship between the complexity of a pattern and its information content. This idea motivated Andrei Kolmogorov and others in the 1960s (Chaitin, 1966; Kolmogorov, 1965; Solomonoff, 1964) to define the **algorithmic complexity** of a string as the length of the shortest computer program that could generate the string. The Kolmogorov complexity, however, was quickly deemed impossible to be effectively computed (Li and Vitányi, 2008; Solomonoff, 1964), and approximate measures were developed to reliably quantify algorithmic complexity.

One of the most popular complexity measures of the Kolmogorov class was introduced by Lempel and Ziv (1976). This measure approximates the amount of non-redundant information contained in a string by estimating the minimal size of the "vocabulary" necessary to describe the string. Strings with high Lempel-Ziv complexity require a large number of different patterns ("words") to be reproduced, while strings with low Lempel-Ziv complexity can be largely compressed with a few patterns employed to eliminate redundancy with no loss of information. After being implemented in a universal lossless data compression algorithm (Ziv and Lempel, 1977; Welch, 1984), the Lempel-Ziv routine has become the standard basis for major data compression methods, such as those used to generate TIFF images and ZIP files.

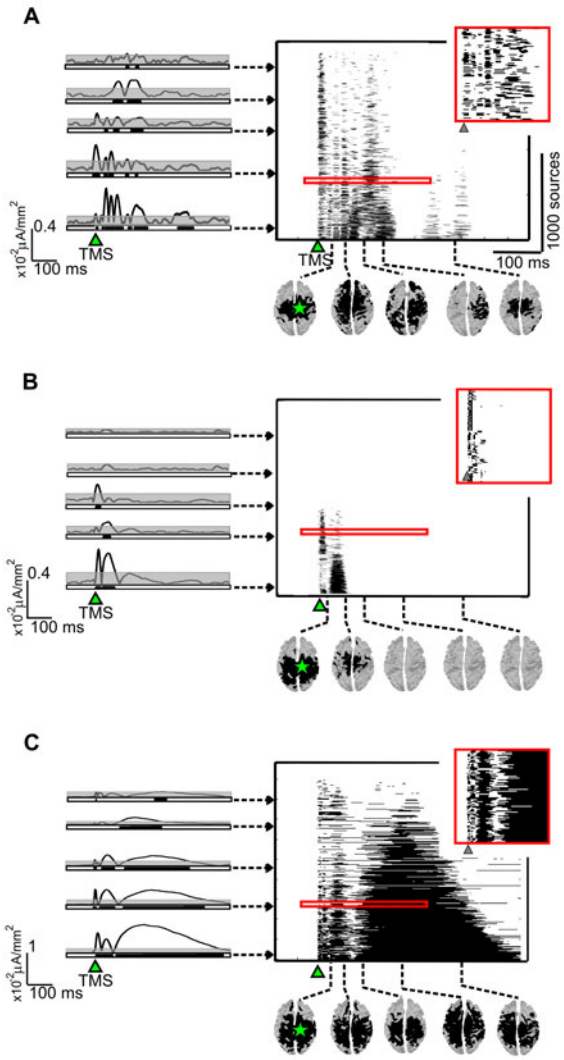

Figure 4.1: Significant sources (SS) binary matrices manifested the reduction of the spatiotemporal complexity of TMS-evoked cortical activation during slow-wave sleep. For each condition described in Figure 3.6 (A, B and C), matrix plots of **SS**(***x***, *t*) are shown (large squares in the center) with sources (rows) sorted by their number of significant samples. TMS-evoked currents for some representative sources are shown on the left (time-series in black). Statistical thresholds (gray areas) applied to each source activity show when sources were significantly activated by the stimulation. The temporal patterns of significant cortical activations comprise the horizontal dimension (lines) of **SS** (the black area in the rectangles below TMS-evoked currents). The vertical dimension (columns) of **SS** contains the spatial distributions of the significant cortical activations (cortical topographical maps on the bottom right). Red rectangular areas of **SS**, with an equal number of vertical and horizontal elements (100 sources × 100 time samples), are displayed in the red squares on the top right after scaling for enlargement. **SS**, when calculated during alert wakefulness (**A**), appears as a complex binary matrix formed by a balanced spatiotemporal pattern of different activated and non-activated sources at different times. The same stimulation during slow-wave sleep generated a simple binary pattern, with some local sources activated earlier (< 100 ms) and no activity later (**B**). Increasing the intensity of stimulation during slow-wave sleep produced a global but stereotypical response, with a simple spatiotemporal pattern of activation, in which a large number of sources were steadily activated for a long period of time (**C**) (adapted from Casali et. al, 2012).

Because the qualitative differences observed in the spatiotemporal patterns of causal activity evoked by TMS between healthy alert wakefulness (Figure 4.1A) and NREM sleep (Figure 4.1B and C) are intuitively related to the notion of complexity in the Kolmogorov sense, the Lempel-Ziv measure is a natural candidate to quantify the patterns of cortical activity appearing to be correlated with the brain's capacity for consciousness. Moreover, the inverse solution and statistics provide the extraction of deterministic patterns of cortical activation, generated by non-linear causal processes, in response to TMS. Therefore, the tools for the analysis of nonlinear dynamics, such as Lempel-Ziv complexity, are best suited to characterize the degree of organization of the activity produced by this class of processes. Other nonlinear tools, such as Lyapunov exponents, depend on the reconstruction of the phase space, the estimation of several parameters and require long time computations, while Lempel-Ziv complexity may be computed with relatively low-cost algorithms.

4.1.1 – *The Lempel-Ziv Algorithm for Binary Matrices*

A simple routine can easily be implemented to calculate the Lempel-Ziv measure for a binary sequence with arbitrary length n (Kaspar and Schuster, 1987). The algorithm scans the sequence searching for subsequences of consecutive characters, or "words". The Lempel-Ziv complexity $c(n)$ accounts for the number of times a new subsequence is encountered: $c(n)$ is a measure related to the number of different "words" necessary to describe the content of the string. We implemented a straightforward generalization of this routine in order to apply it to binary matrices, such as $\mathbf{SS}(\boldsymbol{x},t)$, instead of one-dimensional strings. Our algorithm runs in the vertical dimension of the input matrix (columns), searching for patterns and repeats this search along the horizontal dimension (rows), tracking the patterns encountered in previous columns (see Figure A.1 in the Appendix for a diagram of the algorithm).

The asymptotic behavior of $c(n)$ for random strings generated from a source with entropy H is $nH/\log_2(n)$ (Lempel and Ziv, 1976). Therefore, deviations from maximum complexity may be due to the fact that the "source entropy" H differs from 1 and not to the formation of a pattern in the string (the string is still as random as possible). For this reason, the normalized Lempel-Ziv complexity $\bar{c}(n)$ is usually taken into consideration (Kaspar and

Schuster, 1987) where $\bar{c}(n) = c(n)\log_2(n)/nH$. Asymptotically in n, $\bar{c}(n)$ = 1 for strictly random sequences.

4.1.2 – *The Lempel-Ziv Perturbational Complexity Index* (PCI_{LZ})

The qualitative features of TMS-evoked responses encountered in the previous chapters intuitively suggest that the distinction between alert wakefulness and LOC may be quantified by applying a complexity measure of the Kolmogorov class to the spatiotemporal distributions of significant cortical activity (Figure 4.1). Based on these considerations, we define the Lempel-Ziv perturbational complexity index (PCI_{LZ}) of significant TMS-evoked spatiotemporal activations as the **normalized Lempel-Ziv complexity** ($\bar{c}(n)$) of the binary matrix **SS**.

With **SS** as the input of our algorithm, PCI_{LZ} is a measure that increases with the number of different **spatial patterns** occurring in a given time sample that do not occur in previous samples. In addition, normalizing the Lempel-Ziv measure by the source entropy of **SS**,

$$H = -p_1\log_2(p_1) - (1-p_1)\log_2(1-p_1), \tag{4.1}$$

where p_1 is the fraction of significant spatiotemporal samples contained in **SS**, results in a complexity measure that is minimally dependent on the total amount of significant activity and maximally dependent on the number of different patterns necessary to describe the data. This normalization is justified by the observation that the level of consciousness does not correlate with the fraction of statistically significant sources in the post-stimulus period, which instead depends crucially on external factors such as the stimulation intensity (Figure 4.1 B and C, see also Figure A.2 in the Appendix).

In theory, the complexity of spatial patterns depends on the arbitrary ordination of **SS** rows (sources) and the proper complexity of **SS**(***x***, *t*) should be equated with the minimal complexity across all permutations of all sources. Because searching for all permutations may be computationally unfeasible, we approximate the optimal ordination by calculating PCI_{LZ} from **SS** after sorting sources by their number of significant samples (as depicted in Figure 4.1). Therefore, whereas PCI_{LZ} values will be high for systems that respond with the

activation of different areas at different times, systems that respond with a small number of different spatial patterns will have lower values of PCI_{LZ}, independently of the total amount of significant activity in the post-stimulus period, provided that the stimulation is effective in producing a cortical response. Calculating the PCI_{LZ} for the sessions exhibited in Figure 4.1 confirms this theoretical prediction: during wakefulness, $PCI_{LZ} = 0.51$ for the **SS** exhibited in panel A, and during slow-wave sleep, $PCI_{LZ} = 0.23$ and 0.21 for the matrices exhibited in panels B and C, respectively.

PCI is calculated from the binary spatiotemporal distribution $\mathbf{SS}(\mathbf{x}_j,t)$, which represents where and when the average cortical response to TMS is significantly different from pre-stimulus source activity. Because random noise affecting the EEG signal is considerably attenuated after stimulus-locked averaging and statistics, in contrast to most attempts to quantify physiological data with algorithmic complexity, PCI is not significantly susceptible to confusing complexity with randomness. Still, the correct interpretation of results requires a rigorous control of false positives: **SS** matrices were computed by means of a non-parametric statistics corrected for multiple comparisons, with significance level $\alpha = 0.01$ (see section 2.4). Furthermore, the interpretation of the values of PCI depends on calculation of this measure only when the TMS is able to engage the thalamocortical system effectively. By ensuring that all **SS** matrices in which PCI was calculated had a percentage of spatiotemporal samples (p_1) greater than the rate of false positives of the statistics (1%), corresponding to an entropy $H > 0.08$, we also avoided falsely elevated complexity values due to the division by a number close to zero (H, Eq. 4.1). In general, this control for the signal strength can be achieved at the sensors level by calculating the signal-to-noise ratio (SNR). All TMS sessions in which PCI was calculated had SNR > 1.5 (see Figure A.2 in the Appendix).

4.1.3 – *The Transposed Lempel-Ziv Perturbational Complexity Index* (PCI^{T}_{LZ})

One can also define perturbational complexity in terms of the transpose of $\mathbf{SS}(\boldsymbol{x},t)$, $\mathbf{SS}^{T}(t,\boldsymbol{x})$. This measure will instead increase with the number of **temporal patterns** occurring in a given source that do not occur in other sources. Theoretically, PCI^{T}_{LZ} and PCI_{LZ} are distinct measures: a single source with completely random binary activity may produce high PCI^{T}_{LZ} values due to the high number of temporal patterns contained in the random series but

very low values of PCI_{LZ} (the activation of a single random source has only two spatial patterns). However, because the Lempel-Ziv complexity is combined with a perturbational approach, PCI^{T}_{LZ} and PCI_{LZ} may be predicted to be strongly correlated. As showed by Kaspar and Schuster (1987), when computed in spatially correlated deterministic systems the "**spatial**" Lempel-Ziv complexity, which amounts to the number of different spatial "words" across different time-samples, is also sensitive to the number of **temporal** patterns: if activity in one part of the system is caused by activity in other parts of the system, the Lempel-Ziv complexity $c(n)$ will naturally allow for an analysis of both spatial and temporal patterns. As described in the previous chapters, **SS** contains only those components of cortical activity that are time-locked to the stimulus onset. Therefore, the system responding to TMS is likely to behave as a spatially correlated, deterministic causal system and we expect that when the Lempel-Ziv complexity is applied to the significant sources binary activity, the spatial and temporal versions of the measure will be correlated. In fact, our results obtained from 116 TMS sessions (see Figure 4.9 below) demonstrate that both versions of PCI estimate the same measure of **spatiotemporal** complexity of the TMS-evoked cortical activity. When this correlation is observed, PCI is high only if many regions of the cerebral cortex react to the initial perturbation quickly and in different ways.

4.1.4 – *The Conditional Entropy Perturbational Complexity Index* (PCI_{CE})

We implemented a second and independent measure of complexity to test whether our results were crucially affected by the specific manner in which complexity was calculated. The complexity of a string can also be understood statistically, in terms of the entropic information content of the string. Such measures are based on the conditional entropy or entropy rate, i.e., the average information required to specify a new state with respect to the information content of previous states. For a binary string of n bits $s = s(1)s(2)s(3)\ldots s(n)$, patterns of length L are sequences of L bits $s_L(i) = s(i)s(i+1)\ldots s(i+L-1)$. If $p_L(i)$ is the joint probability of the occurrence of pattern $s_L(i)$, and $H(p_L)$ is the Shannon entropy calculated from the probabilities p_L, the conditional entropy at length L is calculated as the variation of the Shannon entropy with respect to L: $CE(L) = H(p_L) - H(p_{L-1})$. We estimated the probabilities p_L from the binary matrix **SS**, using the second dimension (samples) as the "string dimension" and counting the temporal patterns $s_L(i)$ in all sources

(first dimension). Conditional entropy was then calculated for $1 < L < 10$, correcting the bias introduced in the CE estimation due to the limited number of samples (Porta, 1998), and the complexity was equated with the minimal CE (which occurred between $L = 3$ and $L = 6$ for all TMS/hd-EEG sessions analyzed). Because strictly random strings of length n have a constant conditional entropy equal to their source entropy H, we then defined the Conditional Entropy Perturbational Complexity Index (PCI_{CE}) as the minimal estimated conditional entropy of **SS**, normalized by H as given by Eq. (4.1).

4.2 – Testing PCI

PCI, as a measure of the information content of the spatiotemporal cortical activity significantly evoked by TMS, is low if causal interactions among cortical areas are reduced because the matrix of activation triggered by TMS will be reduced. PCI is also low if all cortical areas react to the perturbation in a similar way because in this case the large matrix of activation is highly redundant and can be compressed using little information. PCI will only be high if many regions of the cerebral cortex react to the initial perturbation quickly and in different ways. Therefore, the considerations and results discussed in the previous chapters support PCI_{LZ} as a possible index correlated with the brain's capacity for consciousness. In this section, we test this hypothesis, calculating PCI_{LZ} in a total of 116 TMS sessions obtained across 36 subjects under the following different conditions: alert wakefulness, slow-wave sleep, REM sleep and anesthesia in healthy subjects and vegetative state (VS), minimally conscious state (MCS), emergence from the minimally conscious state (EMCS) and locked-in syndrome (LIS) in brain-injured patients. PCI values were computed for **SS** matrices obtained from the first 300 ms of the post-TMS interval (matrix dimension of 3004 sources x 106 time samples), for all subjects in all conditions.

4.2.1 – *PCI allowed the discrimination of alert wakefulness from NREM sleep and anesthesia-LOC, independent of the strength and the extent of cortical activation*

We calculated PCI_{LZ} from 13 TMS/hd-EEG sessions recorded in 5 healthy subjects progressing from wakefulness to NREM sleep. The first TMS-EEG session was collected while subjects were awake, with stimuli targeted to the rostral portion of the right premotor

cortex and a maximum electrical field at the cortical target of 90 V/m. The second TMS-EEG session was collected after subjects entered a consolidated period (> 5 min) of NREM stage 2 sleep. In three of the five subjects, we were able to register a third session, where TMS, targeted to the midline sensorimotor area with a maximum electric field of 160 V/m, induced responses resembling complete slow waves.

PCI_{LZ} was also calculated from 12 TMS/hd-EEG sessions performed in 6 healthy subjects during anesthesia. The first TMS-EEG session was collected in each subject before midazolam injection, with stimuli targeted to the rostral portion of the right premotor cortex at an intensity of 120V/m. Midazolam was then given until the subject was unresponsive (level 1 of the OAA/S), with a maximum dose of 0.2 mg/kg. A second TMS session was then collected during deep unresponsiveness (see section 3.3 for details of the experimental protocol).

In addition to PCI, SCD and SCS were calculated for each subject in all conditions. Neither SCD nor SCS unambiguously captured the cortical changes during loss of consciousness. While SCS performed better and discriminated between wakefulness and LOC induced by anesthesia (Figure 3.12), it failed to do so when TMS induced a global cortical slow wave during NREM sleep (Figure 4.2 B). PCI, however, was able to unambiguously distinguish between consciousness and unconsciousness (Figure 4.2 C). During alert wakefulness, all subjects responded to TMS with a complex pattern of cortical activation corresponding to a PCI_{LZ} ranging between 0.45 and 0.61 (mean = 0.52 ± 0.04), whereas during NREM sleep and unconsciousness induced by anesthesia, all subjects responded to TMS with a PCI_{LZ} below 0.30 (mean = 0.22 ± 0.03). Moreover, as a measure of complexity normalized by the source entropy of the binary data, PCI_{LZ} treated the global, long-lasting stereotypical response observed during the induction of slow waves in NREM sleep in the same manner as the local, short-lasting responses observed during sleep and anesthesia. High PCI values were seen only if the activation evoked by TMS was a balanced, long-lasting and widespread spatiotemporal pattern, such as those observed during alert wakefulness.

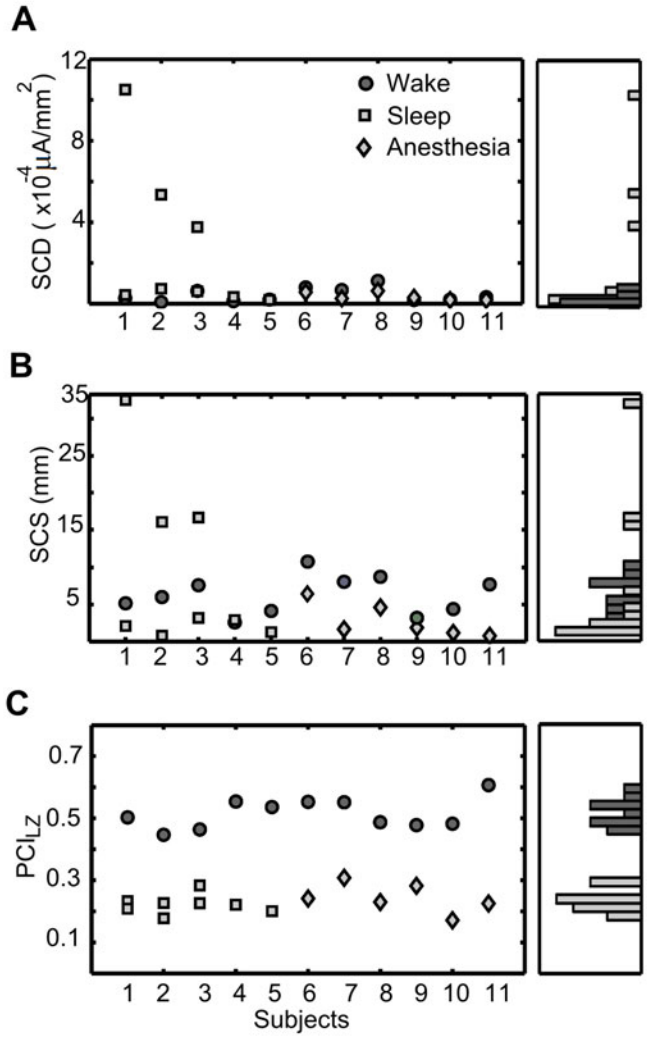

Figure 4.2: Complexity of the cortical activity evoked by TMS (PCI_{LZ}) but not reactivity (SCD) or connectivity (SCS) unambiguously discriminated alert wakefulness from NREM sleep and anesthesia-LOC. The individual values of significant current density (**A**), significant current scattering (**B**) - both cumulative over the post-stimulus interval (0-300 ms) - and perturbational complexity (**C**) are shown for 11 healthy subjects during alert wakefulness, NREM sleep and anesthesia. The panels on the right display the distributions of SCD, SCS and PCI_{LZ} across subjects during alert wakefulness (dark gray bars) and loss of consciousness (light gray bars). PCI distributions reliably discriminated between consciousness and unconsciousness, independent of the strength and the extent of cortical activation ($p < 0.00005$, Wilcoxon rank sum test).

4.2.2 – *PCI allowed a clear-cut distinction between conscious and unconscious healthy subjects, independent of the stimulated area and stimulation intensity*

To test the variability of PCI with respect to the stimulation parameters, TMS was targeted to different cortical areas with different stimulation intensities in 7 awake, healthy subjects. Sessions of ~200 stimuli were collected while subjects were lying on an ergonomic chair, relaxed and with eyes open, looking at a fixed point on a screen. The maximum electrical field at the cortical target ranged from 80 to 160 V/m. The coil was targeted to the superior occipital gyrus (BA19), middle superior frontal gyrus (BA08), superior parietal gyrus (BA07) and the left hand motor area (BA06). Data corrupted by muscular artifacts or with a low signal-to-noise ratio were excluded, resulting in a total of 47 TMS sessions

Figure 4.3 exhibits the PCI_{LZ} results for these 47 sessions collected during wakefulness together with the results of the sleep/anesthesia protocols (Figure 4.2 C). Applying TMS to different cortical areas with different intensities in different subjects did not have a reductive impact on the ability of PCI to discriminate between consciousness and unconsciousness. In fact, if the stimulation parameters were effective in producing a cortical response to TMS, PCI was not sensitive to specific targets or the stimulation intensity. While the maximum PCI_{LZ} was 0.3 during NREM sleep and anesthesia, the PCI_{LZ} for all TMS sessions during alert wakefulness were in the same range observed previously. Altogether, 58 TMS sessions from 18 subjects during alert wakefulness presented a markedly higher PCI (PCI_{LZ} in the range of 0.44-0.65 with a mean value of 0.55 ± 0.05) relative to the 14 TMS sessions from 11 subjects during sleep/anesthesia-LOC (PCI_{LZ} in the range of 0.17-0.30 with a mean value of 0.22 ± 0.03). The p-value for the comparison between the PCI_{LZ} values for the conscious and unconscious groups was less than 10^{-8} (Wilcoxon rank sum test).

4.2.3 – *The discrimination of PCI between consciousness and unconsciousness did not depend on the particular method used to calculate complexity*

We calculated the perturbational complexity based on conditional entropy (PCI_{CE}) (see section 4.1.4) for all 72 TMS sessions displayed in Figure 4.3 to confirm that our results, which allowed a distinction between conscious and unconscious subjects, remained valid for

changes in the method used for calculating the informational content of significant activity evoked by TMS (**SS**(***x***,*t*)).

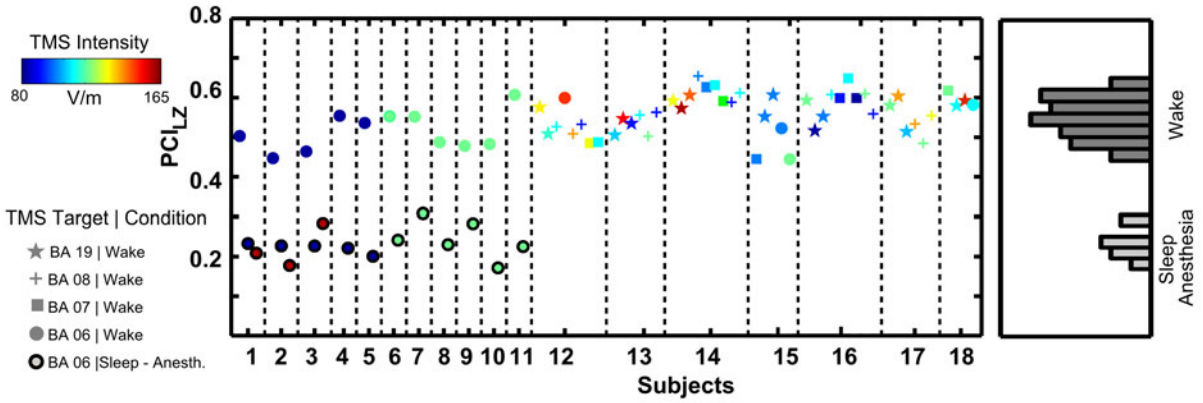

Figure 4.3: PCI_{LZ} discriminated between conscious and unconscious healthy subjects, independently of the stimulated area and the stimulation intensity. The results of Figure 4.2 C (subjects 1-11) are displayed together with PCI_{LZ} values of an additional 47 TMS sessions collected from 7 healthy subjects with stimulations of different intensities (color scale on the left) targeted to different areas (BA 19, 08, 07 and 06) during alert wakefulness (subjects 12-18). As shown on the right, PCI_{LZ} generated disjoint distributions between conscious (dark gray bars) and unconscious subjects (light gray bars). The PCI_{LZ} calculated in conscious subjects ranged between 0.44 and 0.65 (mean value of 0.55 ± 0.05), whereas the PCI_{LZ} calculated after loss of consciousness ranged between 0.17 and 0.30 (mean value of 0.22 ± 0.03). The p-value for the comparison (Wilcoxon rank sum test) between the PCI_{LZ} values for the conscious and unconscious groups was less than 10^{-8} (adapted from Casali et. al, 2012).

Figure 4.4 demonstrates how the reliable discrimination between alert wakefulness and sleep/anesthesia-LOC was also achieved using the conditional entropy index. The correlation between PCI_{CE} and PCI_{LZ} across all sessions was near 90%. PCI_{CE} ranged from 0.3 to 0.48 during alert wakefulness (mean value of 0.39 ± 0.04) and from 0.1 to 0.25 during sleep/anesthesia-LOC (mean value of 0.19 ± 0.04). The p-value for the comparison of PCI_{CE} between the conscious and unconscious groups had the same order of magnitude as the p-value for the comparison of PCI_{LZ} ($p < 10^{-8}$, Wilcoxon rank sum test).

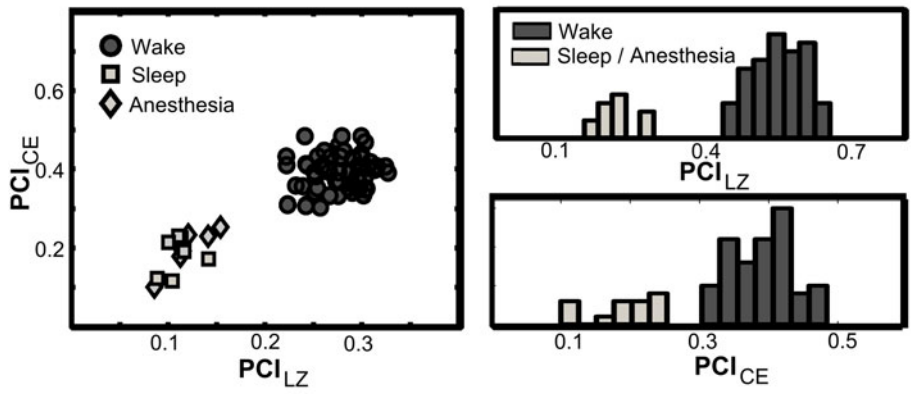

Figure 4.4: The discrimination between conscious and unconscious healthy subjects was preserved using the PCI calculated based on conditional entropy (PCI_{CE}). Left: The linear correlation between PCI_{CE} and PCI_{LZ} across all sessions displayed in Figure 4.3 is shown (correlation coefficient = 88.7%). Right: The distributions of PCI_{LZ} (top) and PCI_{CE} (bottom) in wakefulness and sleep/anesthesia-LOC. Both PCI measures generated disjoint distributions between conditions ($p < 10^{-8}$, Wilcoxon rank sum test; figure adapted from Casali et al., 2012).

4.2.4 – *The PCI time-course revealed a reproducible time-scale related to the emergence of complex TMS-evoked cortical activation during alert wakefulness*

PCI_{LZ} increases with the number of spatial patterns occurring at a given time that have not occurred previously. By keeping track of the Lempel-Ziv complexity ($c(n)$) at each sample time, we constructed a time-series for the complexity during both wakefulness and sleep/anesthesia-LOC (Figure 4.5).

For each condition, the rate at which complexity increased was reproducible across subjects. During sleep and anesthesia, $c(n)$ reached its maximum value in less than 120 ms for all subjects. PCI_{LZ} during wakefulness continued to increase after these latencies, rising to maximum values at times beyond 200 ms. We used the maximum complexity observed during unconsciousness ($PCI_{LZ} = 0.30$) as the threshold above which complex patterns of cortical activation were only observed during alert wakefulness. PCI_{LZ} exceeded this complexity level in all subjects during wakefulness after a time-scale τ ranging from 80 to 115 ms (mean $\tau = 102 \pm 11$ ms). These results demonstrate that the high spatiotemporal complexity observed in cortical activations evoked by TMS during wakefulness is the result

of causal processes in the thalamocortical system occurring later than the processes involved in the immediate local response to TMS.

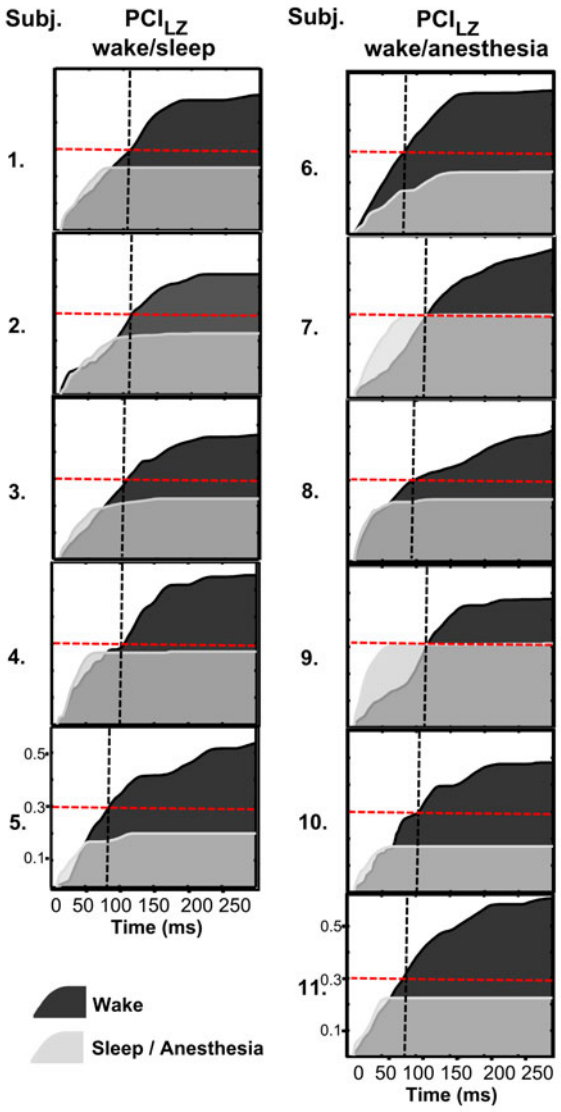

Figure 4.5: The complexity time-course indicates a reproducible time-scale related to the emergence of complex TMS-evoked cortical activations during alert wakefulness. The Lempel-Ziv complexity time-courses are displayed for all subjects in the sleep/anesthesia protocols (section 4.2.1). The temporal evolution of complexity was highly reproducible across subjects in both conditions: wakefulness (dark gray areas) and sleep/anesthesia-LOC (light gray areas). Red dashed lines indicate the maximum PCI_{LZ} (0.30) observed across all subjects during sleep/anesthesia-LOC. Black dashed lines mark the latency (τ) above which an individual's PCI_{LZ} calculated during alert wakefulness exceeds the maximum complexity during unconsciousness, resulting in an average time-scale across subjects of τ = 102 ms ± 13 ms (adapted from Casali et al., 2012).

4.2.5 – *PCI increased progressively, in parallel with the level of consciousness*

The results obtained thus far demonstrate that by measuring the complexity of the cortical response to an external perturbation, it is possible to objectively probe the capacity of the thalamocortical system to dynamically engage different cortical areas in an integrated process of causal interactions. This objective and quantifiable attribute of cortical activation reliably discriminates between conditions in which consciousness is clearly present and conditions in which it is unambiguously absent. All healthy subjects submitted to TMS/hd-EEG during loss of consciousness presented cortical responses to the stimulation with a $PCI_{LZ} \leq 0.3$, whereas all awake and alert healthy subjects responded to TMS with a significantly more complex spatiotemporal pattern of cortical activation ($PCI_{LZ} > 0.44$; Figure 4.3).

What should we expect when the levels of consciousness are vague and uncertain? For instance, during REM sleep, subjects are not awake and are almost paralyzed but they usually experience long and vivid dreams, which suggests that consciousness levels are close to wakefulness. We calculated the PCI_{LZ} from the four TMS sessions described in section 3.2.2 that were collected while the subject transitioned from wakefulness through stage 1 sleep to NREM (stages 2-3) and REM sleep. Figure 4.6 shows the significant sources binary matrix (**SS**(***x***,*t*)) and the corresponding PCI_{LZ} values during wakefulness (A), sleep stage 1 (B), sleep stages 2-3 (C) and REM sleep (D). Complexity gradually decreased from wakefulness to NREM sleep and rose to wakefulness levels during REM sleep.

In addition to healthy subjects during REM sleep, patients affected by consciousness disorders expose an important instance of clinical ambiguity, where the level of consciousness is not easily evaluated. Perturbational complexity was calculated from 40 TMS sessions with 17 patients (12 patients displayed in Figure 3.14 and five additional patients who were diagnosed as EMCS) who evolved from coma into different clinical states.

Figure 4.7 shows the PCI_{LZ} results for the brain injured patients, together with data from the 19 healthy subjects exhibited in Figures 4.6 (one subject in distinct sleep stages) and 4.3 (PCI distributions during wakefulness and sleep/anesthesia–LOC for 18 subjects). The dashed line, transecting all panels of the figure horizontally, marks the maximum PCI_{LZ} observed during loss of consciousness in healthy subjects ($PCI_{LZ} = 0.3$). When PCI was calculated for VS patients who were aroused but unaware, the results were in the range of

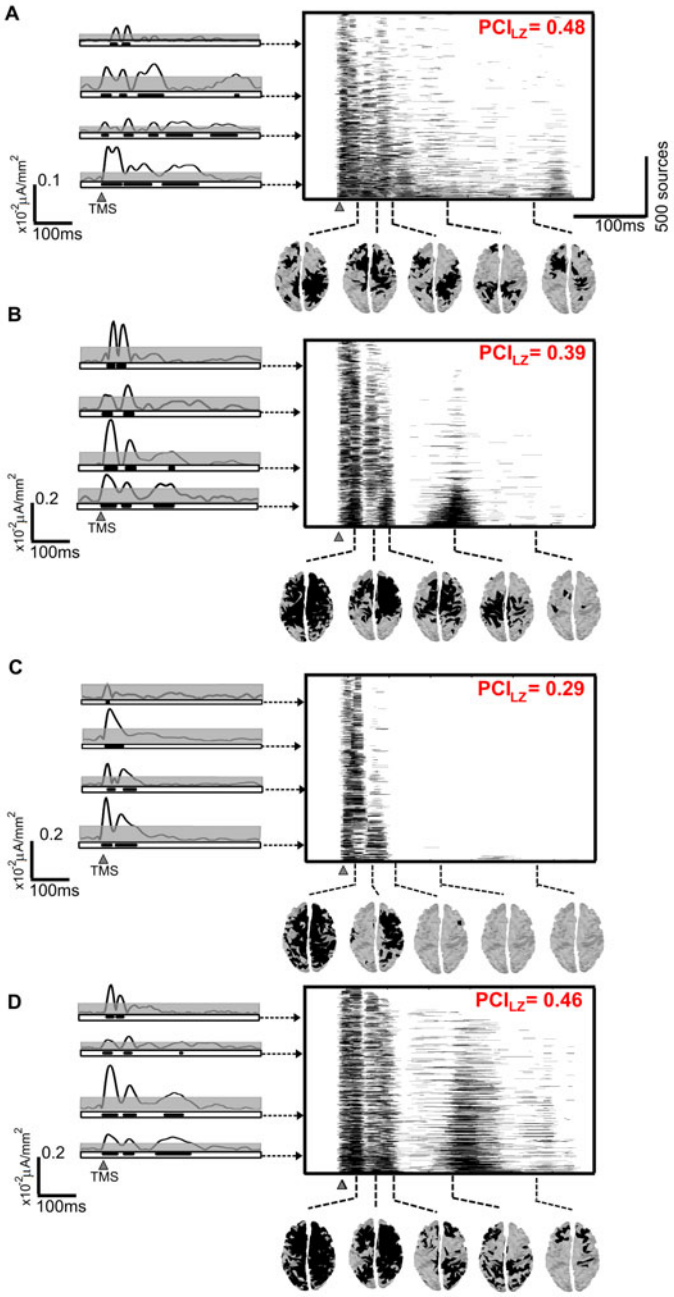

Figure 4.6: Perturbational complexity was progressively reduced during the transition to NREM sleep and partially recovered during REM sleep. For each condition described in Figure 3.8 (A, B, C and D), matrix plots of **SS(*x*, *t*)** are shown (large squares) with TMS-evoked currents for some representative sources (time-series on the left), temporal patterns of significant cortical activations (black area in the rectangles below TMS-evoked currents) and spatial distributions of significant cortical activations (cortical topographical maps on the bottom right – see also Figure 4.1). **SS(*x*,*t*)**, when calculated during alert wakefulness (**A**), appeared as a complex binary matrix with PCI_{LZ} = 0.48. Complexity progressively decreased when the subject was allowed to sleep, from PCI_{LZ} = 0.39 during sleep stage 1 (**B**) to PCI_{LZ} = 0.29 during sleep stages 2-3 (**C**). Finally, during REM sleep (**D**), complexity recovered to wakefulness levels (PCI_{LZ} = 0.46).

sleep/anesthesia-LOC (PCI_{LZ} for VS patients ranged from 0.19 to 0.29 with a mean value of 0.25 ± 0.04). At the other extreme, the two patients who could communicate reliably but were diagnosed with LIS responded with a PCI in the range of the wakefulness distribution (mean value of 0.52 ± 0.06).

When PCI was calculated for the intermediate levels, corresponding to MCS and EMCS, the results followed the trends of the CRS-R, remaining between the complexity distributions obtained when consciousness was unambiguously present and when consciousness was unambiguously absent. The PCI_{LZ} calculated for MCS patients ranged from 0.22 to 0.49 (mean = 0.38 ± 0.07), and all MCS patients exhibited TMS/hd-EEG sessions with a complexity above the maximum value observed during loss of consciousness. Finally, all sessions collected from patients diagnosed with EMCS were above the sleep/anesthesia-LOC distribution, with PCI_{LZ} ranging from 0.37 to 0.52 (mean = 0.43 ± 0.05). Therefore, PCI was able to reliably discriminate intermediate but highly ambiguous states of consciousness (MCS and EMCS) from both lower (VS and unconscious healthy subjects) and higher levels of consciousness (LIS and alert healthy subjects) (Figure 4.8).

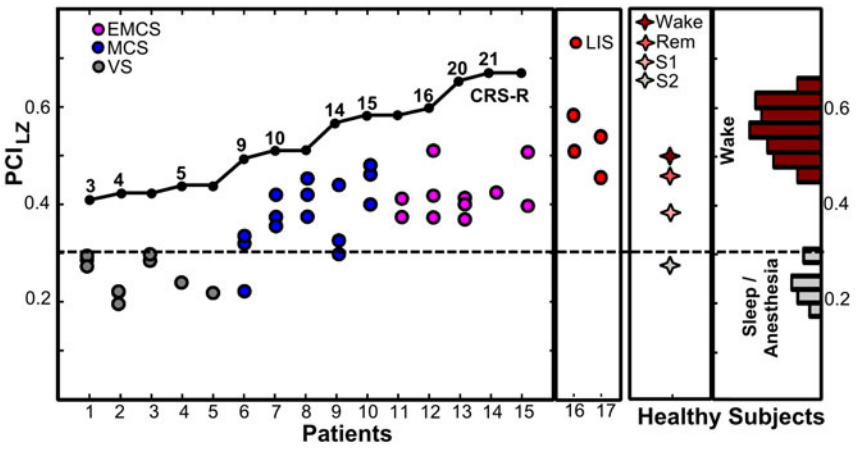

Figure 4.7: PCI_{LZ} increased progressively in parallel with the level of consciousness. The Lempel-Ziv perturbational complexity results are shown when calculated for all 116 TMS sessions collected from 17 brain-injured patients and 19 healthy subjects. PCI_{LZ} produced disjoint distributions between alert wakefulness (dark red) and sleep/anesthesia-LOC (light gray) for healthy subjects (**Right**) (see also Figure 4.3). Unconscious healthy subjects responded to TMS with a $PCI_{LZ} \leq 0.3$ (dashed horizontal line). When calculated in a healthy subject transitioning from wakefulness to NREM and REM sleep, PCI gradually dropped to unconscious levels

during sleep stages 2-3, and increased again during REM sleep (see also Figure 4.6). In brain-injured patients who progressed from coma to different clinical states (**Left**), PCI_{LZ} increased progressively, following the CRS-R trend from VS through MCS and EMCS, attaining similar levels to healthy awake subjects when calculated for LIS patients (adapted from Casali et al., 2012).

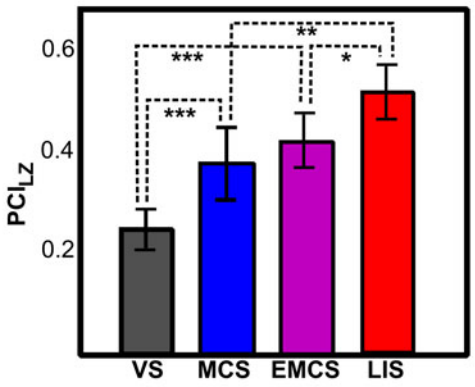

Figure 4.8: PCI discriminated intermediate levels of consciousness (MCS and EMCS) from both lower (VS) and higher (LIS) levels of consciousness in brain-injured patients. The group averages and standard deviations for PCI_{LZ} for each condition are presented with their statistical significance. * Wilcoxon rank sum test $p < 0.05$, ** $p < 0.005$, *** $p < 0.0005$ (adapted from Casali et al., 2012).

4.2.6 – *PCI is a measure of the spatiotemporal complexity of the cortical activity evoked by TMS*

PCI_{LZ} increases with the number of different spatial patterns occurring in a given time sample (the vertical **SS** dimension) that do not occur in previous samples (the horizontal **SS** dimension). In order to test for the sensitivity of PCI_{LZ} to temporal complexity, we calculated the transposed Lempel-Ziv complexity, PCI^{T}_{LZ}, which measures the number of different **temporal** patterns contained in the spatiotemporal binary distribution **SS**(***x***,*t*). Figure 4.9 shows the strong correlation between PCI^{T}_{LZ} and PCI_{LZ} when calculated for all 116 TMS/hd-EEG sessions (correlation coefficient > 97%). These results confirm the sensitivity of the Lempel-Ziv complexity to both spatial and temporal patterns of cortical activity evoked by TMS (see section 4.1.3).

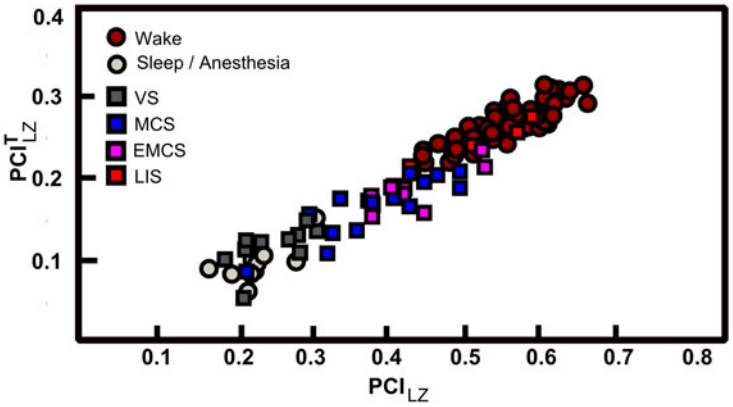

Figure 4.9: PCI^{T}_{LZ} and PCI_{LZ} are strongly correlated. The linear correlation of PCI^{T}_{LZ} and PCI_{LZ} across all 116 TMS sessions is displayed (correlation coefficient = 97.9%). These results show that PCI^{T}_{LZ} and PCI_{LZ} are equally sensitive to the spatiotemporal complexity of the cortical activity evoked by TMS (adapted from Casali et al., 2012).

While PCI_{LZ} is capable of providing an estimate of the complexity time-course, PCI^{T}_{LZ} can be calculated with a computation time significantly shorter than the time required by PCI_{LZ}, without influencing the potential of PCI to discriminate between consciousness and unconsciousness. All unconscious subjects exhibited a $PCI^{T}_{LZ} \leq 0.15$ (mean = 0.09 ± 0.02), whereas during alert wakefulness, PCI^{T}_{LZ} ranged from 0.22 to 0.31 (mean = 0.27 ± 0.02) (see also figure A.3 in the Appendix). Finally, the correlation between PCI_{LZ} and PCI^{T}_{LZ} was preserved when the measures were calculated from brain-injured patients: similarly to PCI_{LZ}, PCI^{T}_{LZ} also increased progressively in parallel with the level of consciousness (see also figure A.4 in the Appendix).

Together, the results of this chapter demonstrate that PCI, a measure of the spatiotemporal complexity of the cortical activity evoked by transcranial magnetic stimulation, was able to reliably and robustly discriminate between conscious and unconscious healthy subjects, producing disjoint distributions that were independent of the stimulation parameters or the specific method used to calculate the complexity. Moreover, PCI was able to detect progressive changes in consciousness, such as those that occur while a subject is falling asleep, and to discriminate between intermediate consciousness levels in

patients suffering from DOC (MCS and EMCS) versus both lower (VS, slow-wave sleep/anesthesia-LOC) and higher (LIS, healthy alert wakefulness) levels of consciousness. It is worth highlighting that PCI is a single number that can be calculated at the bedside with little *a priori* information. This technique bypasses sensory pathways and subcortical structures and directly probes the thalamocortical system, independent of the integrity of the sensory and motor systems. In addition, PCI does not depend on the willingness of the patient to participate, estimating the capacity of thalamocortical circuits to sustain causal activity independent of behavior. All of these considerations support PCI as an appropriate tool to approximate an objective measure of the neural correlate of consciousness, with the potential to assist the diagnosis and prognosis in patients with disorders of consciousness.

Chapter 5.

Complexity and the Brain's Capacity for Consciousness

"Hay ríos metafísicos, ella los nada como esa golondrina está nadando en el aire, girando alucinada en torno al campanario, dejándose caer para levantarse mejor con el impulso. Yo describo y defino y deseo esos ríos, ella los nada. Yo los busco, los encuentro, los miro desde el puente, ella los nada. Y no lo sabe, igualita a la golondrina. No necesita saber como yo, puede vivir en el desorden sin que ninguna conciencia de orden la retenga. Ese desorden que es su orden misterioso, esa bohemia del cuerpo y el alma que le abre de par en par las verdaderas puertas."(Julio Cortazar, "Rayuela")

Neuroscience faces the challenging task of developing and implementing objective measures of consciousness that can be applied to patients who are unable to interact with their external environment. Here, we approached this difficult problem from a novel perspective. General considerations about what constitutes the content of conscious experience and the assumption that the brain must be involved in sustaining consciousness led us to hypothesize that consciousness depends on the brain's capacity to sustain causal interactions between different areas of the thalamocortical system. To investigate this hypothesis, we employed TMS/hd-EEG to directly perturb the cortex and measure the cortical response generated exclusively by causal interactions. We found that when consciousness was absent, such as during slow-wave sleep (section 3.2) and deep anesthesia (section 3.3), the thalamocortical system responded to TMS with qualitatively simpler, slower and non-specific patterns of activation as compared to alert wakefulness (section 3.1). Moreover, employing TMS/hd-EEG at the bedside of patients emerging from coma after severe brain-injury allowed for the reliable discrimination between VS and MCS patients in

terms of the qualitative complexity of their responses to the perturbation (section 3.4). To create an objective evaluation of this capacity for consciousness, we developed and tested the Perturbational Complexity Index (PCI), a measure of spatiotemporal complexity of cortical activity evoked by TMS that is high only if many regions of the cerebral cortex react to the initial perturbation quickly and in different ways. Figures 4.3 and 4.7 summarize the results obtained from calculating PCI on 19 healthy subjects and 17 brain-injured patients. PCI values were remarkably reproducible across subjects in the same group, and the difference between consciousness and unconsciousness was not affected by varying stimulation parameters (Figure 4.3). The results of this objective measure applied to brain-injured patients (Figure 4.7) indicate that it is possible to reliably discriminate between different levels of consciousness in a way that is completely independent of the subject's ability to exchange information with the external world.

The complexity of cortical activity was considered to be a possible correlate of consciousness in a number of other studies. Lyapunov exponents (Fell et al., 1993), mutual information (Tononi and Edelman, 1994), Granger causality (Seth, 2005), approximate entropy and algorithmic complexity of spontaneous EEG (Gu 2003) have all been used to estimate the complexity of cortical activity. However, our measure allows, for the first time to our knowledge, the unambiguous empirical distinction, at the single subject level, of consciousness and different models of unconsciousness. Furthermore, the technique implemented here does not suffer from the major theoretical and practical limitations that affect previously developed measures. Because PCI is based on the significant response evoked by TMS, random noise is considerably reduced by using stimulus-locked averaging and statistics and, therefore, in contrast to most attempts to evaluate algorithmic complexity, our index is not significantly susceptible to confusing complexity with randomness. Additionally, instead of measuring activity that is indirectly related to complexity in the thalamocortical system, such as high-frequency coherence of spontaneous EEGs, which can occur even when the underlying causal cortical activity is spatiotemporally simple, PCI is a direct measure of brain complexity. Finally, PCI measures causality instead of correlation and is capable of distinguishing complex activity generated by interactions between parts of the system from the activity generated by a common exterior input. Thus, because PCI is a complexity measure associated with a perturbational approach, it can be employed to directly investigate and quantify the patterns of causal interactions in the thalamocortical system that

are relevant for consciousness, instead of random noise or activity originating from sources others than causal interactions between cortical areas.

Figure 4.9 prove that PCI is a measure that shows sensitivity to the complexity of both the spatial and temporal components of the response to TMS. This result reinforces the causal origin of the significant response to TMS: activity complex in time can not fail to involve several areas of the thalamocortical system, and a response complex in space can not fail to extend in time. Because PCI is calculated on data obtained after stimulus-locked averaging from multiple trials, cortical activity produced by a non-deterministic process does not contribute to complexity as measured by PCI. Considering then the known difficulties in generating a random-like signal from deterministic causal processes, it is reasonable to expect that temporally complex, pseudorandom activations will depend on a complex network of interactions with other regions of the thalamocortical system: temporal patterns contributing to PCI require the involvement of spatially distributed areas. On the other hand, the cortical activity which is time-locked to the stimulus onset can only be statistically significant directly in the area stimulated or, indirectly, through causal connections with the area stimulated. Therefore, considering the time of neural conduction and synaptic transmission, a complex pattern in space should also depend on an activity that is extensive in time: there is no instantaneous spatial complexity of causal origin.

The results displayed in Figure 4.5 also support this relationship between space and time: the time-course of complexity clearly suggests a reproducible temporal scale (approximately 100 ms), above which causal interactions in the brain account for the complex spatial patterns of cortical activity observed during alert wakefulness. This figure also illustrates how we can extract characteristics of the dynamic processes relevant to consciousness from the complexity of the cortical activity evoked by TMS. Since the works of Libet (Libet, 1982) we know that conscious perception requires processes in the brain that are extended in time. Our results support this finding and those of other recent studies indicating that conscious perception is based on information processing lasting beyond 100 ms (Lamme and Roelfsema, 2000; Harrison et al., 2007; Huang et al. 2007). In this type of longer delay, recurrent connections are likely involved, possibly via cortico-thalamo-cortical re-entries (Garrido et al., 2007; Guillery and Sherman, 2002; Lamme and Roelfsema, 2000), once the feedforward sweep of activity is completed in the first 100 ms. Interestingly, Boly and colleagues (Boly et al., 2011) recently observed that the impairment of feedback

processes involved in the generation of long-latency electrophysiological responses may be a reliable correlate of consciousness in severely brain-damaged patients. Figure 4.5 suggests that PCI may be capable of detecting this type of impairment.

Furthermore, our results indicate a correlation between consciousness and causal interactions in the brain and strongly support recent clinical evidence suggesting that consciousness depends less on specific circuits and more on the capacity of distributed regions of the brain to interact through divergent cortico-cortical and cortico-thalamo-cortical connections. For instance, studies of patients with frontal brain injuries of both traumatic (Mataro et al., 2001) and non-traumatic (Markowitsch and Kessler, 2000) etiologies have reported clear evidence that specific lesions of the frontal cortex, which is an association area related to superior cognitive capacities, may produce neuropsychological deficits in executive functioning, memory and visuo-constructive domains while preserving intact gross cognitive functioning. Consistent with these results, other studies report intact functional modules with impaired cortico-cortical and cortico-thalamo-cortical communication in vegetative patients who do not exhibit evidence of superior cognitive function (Schiff, 2002; Laureys 2004).

High complexity values, as measured by PCI, are only possible if the activity of a specific cortical area is capable of causally inducing activity in another cortical area in an effective and recursive way. Therefore, the ability to attain high PCI values in response to TMS requires the integrity of cortico-cortical and cortico-thalamo-cortical pathways and a balanced level of neuronal excitability in the thalamocortical system. We hypothesize that alterations in any one of these properties may inhibit the brain's ability to sustain the degree of cortical communication required for consciousness.

5.1 – Possible Physiological Mechanisms Underlying Changes in the Spatiotemporal Complexity of Cortical Activity

The complexity of cortical activity evoked by TMS was reduced when subjects transited from alert wakefulness to slow-wave sleep or during midazolam-induced LOC (Figure 4.2) and rose to higher levels during REM sleep (Figure 4.6). These results suggest that the cortical mechanisms that underlie the loss of complexity in healthy subjects may be

related to mechanisms that dissociate wakefulness and awareness and underlie slow-wave sleep and $GABA_A$-mediated anesthesia.

The first intracellular sign of the transition from waking to slow-wave sleep is the appearance of cyclic, long-lasting, high-amplitude hyperpolarization periods in all types of neocortical neurons, during which cortical neurons stop firing (Steriade et al., 1993; 2001). The transition from slow-wave sleep to either REM sleep or wakefulness is invariably associated with the abolition of these long-lasting hyperpolarizing potentials. Several studies based on computer simulations (Bazhenov et al., 2002) or on *in vivo* (Steriade et al., 2001; Timofeev et al., 2001) and *in vitro* recordings (Sanchez-Vives and McCormick, 2000) have shed light on the mechanisms involved in the generation and interruption of the hyperpolarized states associated with slow EEG rhythms (<1 Hz) and other thalamocortical rhythms that are particularly prominent during sleep (McCormick and Bal, 1997).

The long-lasted hyperpolarization, also called the down-state, observed during slow-wave sleep (Steriade et al., 2001) and under specific anesthetic drugs (Steriade et al., 1993), is not produced by active inhibitory post-synaptic currents. Instead, intracellular recordings indicate that these potentials are probably due to the activity of K^+ leak currents dominating the membrane potential when synaptic currents are largely absent, a form of inhibition known as disfacilitation (Timofeev et al., 2001). It has been suggested that the diminished responsiveness of cortical neurons during long-lasting hyperpolarization is related to an altered balance between excitation and inhibition, which ultimately originates from the hyperpolarization of thalamic relays that do not transfer pre-thalamic information to further relays under conditions such as low cholinergic activity (Timofeev et al., 1996; Gil et al., 1997; Timofeev et al., 2001). During this condition, cortical, thalamocortical, and reticular thalamic neurons are hyperpolarized, and the combination of intrinsic (e.g., hyperpolarization-activated and transient cation currents) and extrinsic (reticulo-thalamo-cortical, cortico-cortical, and cortico-thalamo-cortical interactions) neuronal properties sustain synchronized rhythmic activity in the form of delta, spindle, and other slow waves observed in the EEG (McCorkmick and Bal, 1997; Sherman and Guillery, 2002).

In the particular case of slow oscillations, the hyperpolarized down-state of neocortical neurons is associated with very strong synaptic and intrinsic depolarizing pressures, which ultimately produce depolarized or up-states (Timofeev et al., 2001). During

the down-state, intrinsic neuronal properties, such as the presence of hyperpolarization-activated channels, limit membrane polarization and the summation of random miniature excitatory post-synaptic potentials originating from cortico-cortical and cortico-thalamo-cortical connections ultimately produce strong discharges associated with an increase in the input resistance of cortical cells (Steriade et al., 2001). During these up-states, both excitatory and inhibitory neurons fire synchronously at extremely high firing rates. The amplitude of the depolarization is locally controlled by the activation of GABAergic inhibitory neurons. When outward conductances, such as Na^+-activated K^+ currents, overcome positive feedback from cortical circuits, the depolarized network rapidly fails, returning to a silent hyperpolarized phase (Sanchez-Vives and McCormick, 2000). This cortical bi-stability – where cortical neurons oscillate between an "up" state produced by recurrent cortico-cortical and cortico-thalamo-cortical networks of excitatory and inhibitory connections, followed by a long-lasting (>0.5 s) down-state where synaptic barrages cease – is reflected in the low frequency (<1 Hz) synchronous activity observed in the EEG during slow-wave sleep and during the maintenance period of $GABA_A$-mediated anesthesia (Franks, 2008).

It is likely that the reversible loss in the capacity to sustain communication in the thalamocortical system during slow-wave sleep and midazolam-induced anesthesia may be caused by the bi-stability of neocortical neurons. During the bi-stable period, local activation may either be ineffective in overcoming the hyperpolarization of connected neighbors or may trigger an up-state, which invariably will be followed by a long-lasting period of silence, thus disrupting effective communication between cortical areas. Moreover, in conditions of cortical bi-stability, thalamic cells are likely to be found in the burst mode of oscillation, blocking the trans-thalamic pathways proposed to be crucial for cortico-cortical communications (Sherman and Guillery, 2002). Therefore, the hyperpolarization of neocortical and thalamic cells that occurs during slow-wave sleep and in conditions of $GABA_A$-mediated inhibition is likely to prevent effective and specific communication between different cortical areas, thus obstructing the emergence of complex patterns of activity in response to TMS. In fact, when TMS-hd/EEG was applied to healthy subjects during slow-wave sleep (section 3.2.1) or midazolam-induced LOC (section 3.3), cortical responses to the stimulation either remained limited to the target area or evolved into a full-fledged stereotypical slow wave. In both cases, PCI values were significantly lower than values measured during alert wakefulness (Figure 4.2). Other qualitative characteristics of

cortical responses evoked by TMS during LOC, beside the late breakdown of long-range interactions among cortical areas, also point to mechanisms that may involve the bi-stability of cortical neurons and changes in the corticothalamocortical dynamics. The initial, large positive–negative wave evoked by TMS in sleep and midazolam-induced anesthesia may be accounted by the strong depolarizing pressure exerted on neocortical neurons during a bi-stable condition. Finally, the responses with low PCI values, both during NREM sleep (Figure 3.7) and midazolam-induced anesthesia (Figure 3.10), are associated to an obliteration of the characteristic natural frequencies observed in healthy alert subjects (section 3.1), which suggests that alterations in corticothalamic modules may also underlie the loss of spatiotemporal complexity.

In conditions in which wakefulness is lost, but the ascending cholinergic systems are active, as during REM sleep, communication in the thalamocortical system is partially recovered. The activation of components of the cholinergic system during REM sleep invariably results in the depolarization of thalamic neurons and an increase in the excitability of cortical networks, thereby suppressing disfacilitation and setting both thalamic and neocortical cells back to excitable tonic mode (McCormick and Bal, 1997, Steriade et al., 1993b; 1993c; 2001). In this condition, TMS evoked a long-range balanced pattern of cortical activation, similar to that observed in alert wakefulness (section 3.2.2, Massimini et al., 2010). Our results indicate that the complexity of cortical responses to the stimulation may be at least partially recovered during this condition (Figure 4.6). Data suggest that the selectivity of cortical inhibition may play a significant role in the differences between REM sleep and wakefulness (Steriade et al., 1979; Llinás and Paré, 1991; Timofeev et al., 2001), and we speculate that the less selective action of inhibitory currents in the neocortex during REM sleep as compared to the waking state may be responsible for possible differences in complexity between these conditions.

This mechanism involving long-lasting silent periods that inhibit cortico-cortical and thalamo-cortical communication could explain why PCI is high during both alert wakefulness and REM sleep in healthy subjects but low during slow-wave sleep and midazolam-induced anesthesia. However, because the complexity of cortical activity mainly depends on the availability of rapid causal interactions between different cortical areas through feedforward and feedback sweeps, we also expect PCI to detect conditions in which cortical communication is impaired by other pharmacological and pathological mechanisms.

The main structural abnormalities of brain-injured VS patients, e.g., diffuse axonal injuries and thalamic lesions, are related to alterations in causal interactions within the thalamocortical system (Adams et al., 2000). In particular, the thalamus is likely to play an essential role in cortical activation, not only by promoting the fine-tuning of excitation and inhibition required for balanced neuronal excitability in the neocortex but also by actively participating in the mutual connections between cortical areas (Adams et al., 2000; Guillery and Scherman, 2002). For this reason, we expect severe thalamic lesions to yield low PCI values in response to TMS. In addition, activation deficits in association centers of the cortex, such as those observed in permanent VS patients who retain isolated remnants of functional networks (Schiff et al., 2002) or those predicted to be associated with complex partial seizures (Blumenfeld et al., 2003), may block both bottom-up and top-down pathways, impairing the causal processes involved in complex responses to TMS. Finally, in addition to the suppression of cortical activity, the hyper-activation of specific cortical areas can be expected to produce simple patterns of causal responses to controlled cortical stimulation in conditions such as generalized tonic-clonic seizures (Blumenfeld and Taylor, 2003) or in response to specific anesthetics (i.e., the NMDA-antagonist ketamine (Maksimow et al., 2006; Franks, 2008)) due to the presence of stereotyped patterns of cortical activity that impede normal communication in the thalamocortical system.

The loss of the ability to sustain conscious experience can be caused by diverse physiological (slow-wave sleep), pharmacological (anesthesia) or pathological (coma, vegetative state, seizures) mechanisms. As these mechanisms appear to be all related to the impairment or alterations of the feedforward and feedback circuits that integrate specialized modules in the thalamocortical system, we suggest that a measure of the complexity of the cortical activity evoked by an effective perturbation of the system may be sensitive to all these conditions in which consciousness is lost. To better interpret the underlying mechanisms related to changes in the complexity of cortical activity, it is crucial that future studies calculate PCI under different anesthetic agents (ketamine, xenon, opioids) and on a greater number of patients with different neuropathological conditions.

5.2 – Limitations of the Study and Future Developments

The results summarized in Figure 4.7 comply with what is required by a quantitative measure of consciousness: PCI is high during alert conscious wakefulness and low during unconscious states such as slow-wave sleep, anesthesia and vegetative state. Although PCI values fall into disjointed distributions between consciousness and unconsciousness and a single value (PCI_{LZ} = 0.3) appears to represent a threshold above which signs of consciousness are present in all subjects and patients, this measure still needs to be tested on a greater number of subjects and patients under different conditions before conclusions about the potential clinical applications of PCI can be drawn. In this work, we tested PCI under anesthesia induced by midazolam, which exclusively targets $GABA_A$ receptors. It will be essential to replicate these findings with other anesthetics, including volatile and i.v. agents, which are commonly used in surgical procedures and are likely to have different mechanisms of action than increased GABAergic receptor activity. In addition, the interpretation and possible clinical application of the measure require our results to be replicated in a greater number of subjects in intermediate levels of vigilance, such as in different sleep stages and intermediate levels of anesthesia. It may also be interesting to test PCI as a function of the behavioral state (e.g., as effective connectivity was found to correlate with specific task-related paradigms during alert wakefulness (Johnson et al., 2011)).

Testing PCI on a greater number of patients recovering from coma will be essential to explore the potential of employing PCI in the diagnosis of brain-injured patients. Our results obtained from 17 critically ill patients indicate that this measure is capable of reliably discriminating intermediate states of consciousness (MCS and EMCS) from both lower (VS and unconscious healthy subjects) and higher levels of consciousness (LIS and alert healthy subjects). However, PCI was not able to distinguish between MCS and EMCS, suggesting that patients in these conditions present similar capacities to sustain cortical communication in the thalamocortical system. In addition to address this issue with a greater number of patients, future studies could also consider calculating complexity on a higher order basis. PCI, as introduced in this work, is calculated on a binary basis (on/off source activity), and it is possible that relevant information needed to accurately distinguish between similar but unequal conditions is lost in such a paradigm. Moreover, even if PCI is not sensitive to specific targets or to the stimulation intensity, systematic studies aimed at optimizing the measure, possibly by integrating PCI with high-resolution structural neuroimaging

techniques to control for differences in the coil's orientation with respect to the direction of axons, may uncover optimal stimulation parameters that increase the specificity and sensitivity with which the measure discriminates between consciousness levels.

Absolute values of PCI should not be considered, regardless of the parameters employed in the analysis. The data presented in this work have specific temporal (~2.75 ms) and spatial (~20 mm^2) resolutions, corresponding to a matrix $\mathbf{SS}(\boldsymbol{x}_j, t)$ with dimensions of 3004 x 106. The dimensions of the input matrix for the complexity algorithm affect the normalization and the absolute value of the measure and the clarification of the exact dependence of relative values of PCI on spatial and temporal resolutions requires further investigation. In particular, given the intrinsic limitations of EEG for spatial resolution, the patterns of connectivity detected by TMS/hd-EEG are necessarily coarse. However, coarse resolution does not necessarily mean worse resolution: as with any measure of spatiotemporal complexity, resolutions that are too fine may impair the identification of similarities between conditions. On the other hand, a resolution that is too coarse may be insufficient to retain the amount of information needed to distinguish distinct patterns. Because the accurate activation of small areas is not as relevant for PCI as the activity produced by large-scale interactions between functional cortical regions, we opted to adopt a coarse spatial resolution that was previously tested in studies of large-scale interactions between thalamocortical areas (Casali et al., 2010; Ferrarelli et al., 2010). Finally, a better comprehension of the complexity measure will likely require a detailed study of PCI's dependence on the forward model and on the accuracy of the inverse solution, which will help to clarify the index's dependence on spatial resolution. In particular, qualitative comparisons between different conditions at the sensor level suggest that PCI could be calculated directly from the EEG. Further studies may explore this possibility, considering that volume conduction and spatial smoothing affecting the EEG may require new ways to calculate complexity at the sensor level.

PCI is calculated from significant cortical activity, and the index is therefore not significantly affected by biases confounding complexity with randomness. However, random fluctuations around the statistical threshold may cause over-estimation of the complexity of cortical activity evoked by TMS under specific conditions. In particular, when a noisy signal (corrupted by strong harmonic distortions or with an insufficient number of trials) is modulated by a slow wave in response to TMS, as observed during slow-wave sleep (Figure

3.6C), the slow return to baseline accompanied by fast random oscillations above and below the statistical threshold may introduce noise into the **SS** matrix, resulting in false positives for complexity. When this type of modulation of noisy data is present, it may be advisable to use lowpass filters, thus removing those components that are normally disregarded in TMS-evoked potentials (>40 Hz).

In this work, we employed PCI in combination with TMS, but the measure introduced here may also be calculated from data acquired from somatosensory stimulation. However, as we will discuss in the next section, the unique and unprecedented interpretation of PCI combined with direct cortical perturbation implies profound theoretical, technical and practical advantages of this measure as a possible complement to other ERP and fMRI protocols in the diagnosis of brain-injured patients.

5.3 – The Brain's Capacity for Consciousness and the Science of the Mind

This work is not a thesis about the ultimate nature of consciousness, or the essence of intentionality, but rather about a quantifiable ability of the brain without which, by hypothesis, consciousness as we normally understand it can not be sustained. Nevertheless, the possibility explored here of an empirical investigation of the brain's capacity for consciousness, rather than the neural correlates of specific contents, fits naturally into the more general debate about the possibility of a scientific search for the ultimate nature of consciousness. This last section is reserved for a discussion of some theoretical aspects implicit in our study and highly overlooked in the scientific approach to the question, but which are likely to receive special attention with the development of empirical measures like the one proposed here.

5.3.1 – *The Search for the Contents of Consciousness*

The question of the level of consciousness of critically ill patients diagnosed in VS or MCS and who are incapable of functional communications is ultimately the question for what it is like to be in such states. Facing a patient lying with eyes open, staring at the ceiling, but unable to communicate and with no signs of voluntary movement, the family and

the clinicians are puzzled and without answers to their concerns about the world experienced by that person under their care, about the possibility and quality of what could be defined as his "point of view", his own perspective: what it is like to be him? When looking at the ceiling, dos he see something, is the roof something to him? And pain, is he able to feel it as we do? And music, the image of his family, the sound of his name, are really part of something we might call his experience? Can we tell if he has some kind of experience at all?

Can we actually assume that one day we will have the answers to these questions? Most cognitive scientists believe that an affirmative answer must be based on a definite and *a priori* position about the ultimate nature of the content of consciousness. In particular, the majority of researchers in the field expect some progress only if we agree on the ontological identification of mental contents as internal **representations**, formed ultimately by neural states. Assuming that the content of our thinking is in fact a neural state, by observing relevant properties of such neural states we could then envisage the possibility of to identify whether the patient has an experience associated with pain, if the music sound is something that he is actually experiencing, if the reaction to his name is more than an automatic reaction and even if the ceiling above his bed is actually something for him. It would then be possible to identify categorically, from the "third-person" perspective of science, the situations in which, from the point of view of the patient himself, there is something that corresponds to what it is like to be him, a self perspective of "first-person", full of intentional content.

But could we really make this unambiguous transition between what is objective (the neural activity) and what is subjective (the quality of the conscious content)? That is unlikely to happen. It is conceivable that an "internal representation" associated with an intentional content (a representation that is "something **for**" the subject) is formed, caused or made by neural states. But it is very difficult to justify that such a subjective intentional content can be **ontologically reduced** to objective neural states (Searle, 2004; Putnam, 1988; Nagel, 1974; Wittgenstein, 1953). Even if this is the case, however, even if we admit the existence of a wide bridge linking the two huge gaps implicit in the argument - that the content of consciousness is an internal representation and that this representation can be identified with a neural state - if the question to be made refers to the content of experience as an internal representation, then the answer is in essence accessible only to the subject himself. A practical epistemological problem remains (Nagel, 1974): the distinct nature of the third- and first-person speech seems to preclude the determination of the necessary and sufficient link

between a specific property of a certain neural state and a specific qualitative property of the subjective experience. There is an "explanatory gap" (Levine, 1983) between the different modes of access to such content of consciousness that, assuming the fundamental separation between internal representation and the objects in the world, seems to be insurmountable. In fact, although techniques based on fMRI and ERP protocols significantly advance our empirical knowledge about what is associated with conscious perception, the problems of interpretation, coupled with the ambiguity of the results, reinforce the idea that there is an epistemological impasse at the basis of such empirical research.

Let us then step back. Why an empirical investigation of consciousness should start from the problematic reduction of the content of consciousness to certain internal neural states? Historically, this "internalist" paradigm appears as a natural reaction to the mind/body dualism of Descartes: the mind is also part of the body and everything related to consciousness has also a material basis. Particularly since the second half of the twentieth century, however, it has been understood that the reduction of the content of consciousness to neural states is not the only way to combat the Cartesian dualism. Recent empirical studies (Wexler, B., 2006; O'Regan and Noë, 2001; Libedinsky and Livingstone, 2011) reinforce an "externalist" paradigm, which finds its theoretical basis in the different but converging philosophies of Martin Heidegger (1927) and Ludwig Wittgenstein (1953) and aims to combat dualism in its bases[3]. In this scenario, consciousness, rather than a **representation**, would consist more precisely in a **disclosure**, a "disclosing activity" which is not necessarily representational, but concerns primarily real external objects that, only then, could be "internalized" and, in a similar but unequal way, "revealed" in imagination or memory (Rowlands, 2010; Noë, 2004). Amid this debate, it is reasonable to expect that a science of consciousness that lives up to its name might consider the ultimate nature of the content of the experience not as a matter of principle, but as an open question to be clarified by empirical research itself.

[3] In fact, the "internalism" can be seen as a remnant of Descartes' dualistic paradigm. The "internalist" program intends to eliminate the dualism between *res extensa* and *res cogitans* by acceptance of the starting point of the Cartesian argument in defense of such metaphysical dichotomy: the dualism between mental states as internal representations and the objects of the outside world. The "externalist" position does not accept the Cartesian premise affirming that the fundamental domain of thought is formed by internal "representations" or mental states, which are linked to the external environment causally but not cognitively, while the world, in principle, is "out" of our direct reach and accessible only through inference from the perceived "impressions". This idea that there has to be an interface between our cognitive capabilities and the external world, whether in the Cartesian version or in the version of contemporary cognitive sciences, is likely to be at least partially responsible for all the major metaphysical and epistemological problems of Western philosophy since the seventeenth century (Putnam, 1999; McDowell, 1996).

This question about the ultimate nature of the objects of consciousness, whether essentially internal representations or including also real objects of the world, leads us to an aspect of particular theoretical importance to which our results are noteworthy. In the present thesis, we intend to address the problem of implementing objective ways to evaluate the level of consciousness in patients unable to communicate functionally, searching for neural correlates not of **specific contents of consciousness**, but directly of the **level of consciousness**. From general considerations about the phenomenological properties of the intentional content of consciousness, we set a reference to the object of study without committing ourselves to one or another paradigm about the ultimate nature of such contents[4]. We could then formulate a hypothesis that is justified from an empirical point of view: the brain must be able to sustain a balanced level of causal interactions between parts of the thalamocortical system to give access to the contents of consciousness. This path led us finally to a direct measure of causal interactions in the thalamocortical system that was sensitive to the level of consciousness under the conditions tested here. The relative success of this attempt to directly investigate the possible conditions that support awareness suggests that the search for neural correlates of consciousness should not necessarily be initiated by an investigation of the nature of the contents of consciousness. Instead, rather than based upon the awkward reduction of something that has a subjective and private character to something that is objective in nature, a science of the mind should start with the objective aspects of consciousness itself. And if there is a fundamental objective aspect with regard to the intentionality of consciousness, it does not relate so much to its content, but to the vehicles that provide such content.

5.3.2 – *The Objective Character of Intentionality*

The consideration of intentionality as the realm of subjectivity is very popular among some philosophers and scientists. Those who are attracted by such subjective realm can use

[4] The basic aspects of phenomenology considered in our argument (section 1.2) are largely consistent with most of the "internalist" and "externalist" paradigms. It should be noted, however, that these aspects of intentionality are obviously incompatible with a specific "internalist" paradigm, not treated here, that intends to delete or deny the intentional content of consciousness. Without a phenomenological guide to fix reference to the object under study, this "eliminativism" can only result in a science of some kind of information processing in the brain, but not in a science of what we usually understand as consciousness. Citing Nagel (1974) "*any reductionist program has to be based on an analysis of what is to be reduced. If the analysis leaves something out, the problem will be falsely posed.*"

this identification to establish a radical dualism between mind and matter and a skepticism about the prospects for a science of consciousness. On the other hand, scientists who consider such subjectivity as repulsive end up by flirting with the radical idea of denying the existence of intentional phenomena as if they were mere illusions, which results only in changing the name of the search for the correlates of consciousness – instead of asking whether the patient is conscious or not, we would then ask whether he/she is "deluded" or not –, without any simplification of the challenges involved in obtaining the answer. Possibly, perhaps even likely, there is a core of subjectivity that can not be eliminated in the content of consciousness, but at the same time, it does not follow that everything essentially involved with consciousness is subjective and accessible by introspection.

For what is likely the most relevant contribution of the discussion between externalists and internalists for a science of consciousness, that has gone largely unnoticed in most scientific discourse, refers to a constitutive aspect of intentionality that is decidedly **not** of private and introspective access, but requires a scientific, third-person approach. This aspect refers to the fact that the contents of consciousness - whether internal or external - depend on a component that is not part of the experienced content but, instead, is that **by virtue of which** we experience it. This idea, which echoes from an insight of Gottlob Frege (1892) and was developed in both philosophical traditions of the twentieth century, points to consciousness as involving a kind of **movement toward** its content that, as a condition of possibility for such content, can not itself be an object of consciousness.

We can illustrate this idea by considering the activity of a pianist. As a student of piano, the execution of a piece of music is marked by concentration on the position of the fingers, in the sequence of keys that must be played and other characteristics of this activity that is not yet mastered. It is only when the pianist is able to completely transcend these points, and precisely because he is able to do so, that his conscience, "travelling through" what was once opaque, gives him a content that is unique to the master: the particular experience of being engaged in the masterful execution of a piece. During the performance, keys and the fingers are no longer the objects of consciousness, they "disappear": the consciousness of him who gained the expertise to use them "goes through them". Still, there is a sense in which the keys are present: the pianist does not play them as being incapable of knowing that they are where they are. Instead, even without being the **object** of

consciousness, they are present **in** consciousness as the vehicles that enable access of the pianist to **what it is like to perform a play**.

If what characterizes an object of consciousness is its intentional character associated with a "*what-it-is-like-ness*", this phenomenological analysis suggests that such intentional character should be seen as the **vehicle** for the contents of consciousness and, as such, must be at the same time understood as a noneliminable component of conscious experience. First, vehicles, as conditions of possibility for their content, are not secondary, but, as stated by Martin Heidegger (1927), are "primordial": they found the phenomenon[5]. Second, any attempt to turn inward to the alleged internal representations and search for the vehicle of our thoughts inevitably turn it into the content of consciousness. This is how the phenomenology of intentionality itself points to a structure that is out of phenomenological access: the vehicle, as something opposed to the content, necessarily escapes the kind of knowledge accessible to the first-person perspective[6]. It is then suggested that the fundamental characteristic of consciousness, as opposed to its objects, is, from a first-person perspective, the movement present in the intentional act which is absent of what we are aware of, but is part of what it is like to be conscious of it. Intentionality itself, as something distinct from its object, appears as that "metaphysical river" where the character in Cortazar's novel swims: in the river of consciousness one swims without knowing and without needing to know; one finds themself so dipped in it that it is impossible to see the river from the swimmer's perspective, and therefore describe or much less define it. On the basis of intentionality there is something which, paraphrasing the poet, can be said an "unknowing" that, exactly for being unknown, is able to provide the "door opening" of consciousness.

In this scenario, all intentional content, whether or not accessible by exclusively introspective roads, presupposes an activity **external to such content**, not eliminable and that can not be accessed by introspection. That is, consciousness as an intentional act, as a

[5] This primordial character depends essentially on the observation that it is possible to form a representation of something from the use we make of it, but the use can not be derived from the mere representation. This point appears in Heidegger (1927; see also Hall, 1993), and is also explored by Wittgenstein (1953). In particular, the author of the "Philosophical Investigations" observed how we are not able to give a definition for simple words, such as "game" ("*Spiele*"), that can explain all the uses we make of them in common language. This however is not a problem: we do not need such definitions precisely insofar as we are able to **use** the words correctly. The use, which is more fundamental, depends on a learning of rules that is irreducible to the formation of a set of internal representations, but involves, instead, a "grasping" in an "open texture" of external criteria: "*an 'inner process' stands in need of outward criteria*"(Wittgenstein, 1953, § 580).

[6] It is likely that this pointing of phenomenology to something that is external to introspection is precisely what lies at the convergence of analytical and continental traditions in the second half of the twentieth century.

"go-to", is not primarily characterized by that to which it is addressed, but by the way through which it goes. But precisely because such vehicles of experience are inaccessible to introspection, the search for them can not in any way be answered by the first-person speech and it becomes a clear question for the scientist: what supports this "unknowing", the essential movement of intentionality toward its object?

If science largely ignored this objective character of intentionality, focusing instead on the search for correlates of specific contents of consciousness, it is probably because the empirical investigation of this characteristic that is out of the mental content would depend on a technique previously unavailable. In general, it is not possible to establish the properties of the vehicle based on the properties of the content to which it lead us (Dennett, 1991), and this makes the distinction between vehicle and content crucial to claims of an empirical science of consciousness. John Searle points out this important aspect with regard to investigating the neural correlates of consciousness (NCC):

> *"Most of the discussions that I have seen of the NCC are confused because the researchers are looking for an NCC for a particular element of the conscious field, such as, for example, the experience of the color red. But that experience occurs in a subject who is already conscious. So the NCC could not possibly give us sufficient conditions for consciousness because the subject has to be already conscious in order that the NCC in question can cause a particular perceptual experience. The basic insight is this: we should not think of perception as creating consciousness, but as modifying the preexisting conscious field. That is why most of the research that I have seen does not give an NCC but what we might call us a NCPP – a neural correlate of a particular perception, given that the subject is already conscious"* (Searle, 2004b).

Clearly, the empirical research of this objective character of intentionality, which is separate from the content of experience, requires a technique capable of distinguishing the presence or absence of consciousness in a way that does not depend significantly on the presence or absence of specific mental contents in the subject observed. As discussed below, this practical independence of the technique on the content experienced by the subject is exactly what we suggest to have found by measuring the complexity of cortical activation evoked by TMS.

5.3.3 – *PCI and the Science of the Mind*

PCI, combined with TMS/hd-EEG, can measure a possible condition for conscious experience **in general** and can be applied to the human thalamocortical system without a need for the subject to be engaged in a specific representational content. Unlike studies based on ERP or fMRI protocols, the response evoked by TMS in a conscious subject is not characterized by the awareness of a particular stimulus, much less awareness of the magnetic stimulation itself. In some specific circumstances, TMS can cause phosphenes (Kastner et al., 1998; Meyer et al., 1991) or motor activation (Barker et al., 1985), but with the stimulation parameters described herein, TMS is not accompanied by any conscious experience directly related to the magnetic stimulus in the cortex. In addition, the use of "noise masking" (section 2.1) prevents the formation of time-locked auditory potentials (Massimini et al., 2005), and the only experience that can be indirectly locked to the stimulus is a weak somatosensory stimulation on the surface of the scalp due to electrical current induced by the magnetic pulse. It is unlikely, however, that the intensity of this potential can be comparable to the response evoked by TMS to a point that it determines the differences observed between consciousness and unconsciousness. On the contrary, several studies combining TMS with PET and fMRI suggest that the neural activation induced by TMS is fundamentally characterized by the spreading of activity to specific brain regions that are connected with the stimulation site (Paus et al., 1997; 2001, Siebner et al., 1998) instead of an invariant event-related potential. Indeed, our results, after a systematic investigation of the potential evoked by TMS with different stimulation parameters, strongly suggest that the presence of a SEP in the scalp does not interfere significantly in the signal evoked by TMS (Casarotto et al., 2010; Casali et al., 2010). While the SEP does not depend significantly on the angle of stimulation or the stimulated area, having little variation even in relation to the intensity of stimulation, the potential evoked by TMS is highly sensitive to each of these variables (section 2.2). We found that stimuli in different areas during alert wakefulness produce responses with completely different characteristics (Figure 3.1 and 3.4) and engage different circuits at different times (Figure 2.11). These characteristics of effective connectivity in the thalamocortical system are precisely the properties of the response evoked by TMS that have decisive influence on what PCI was designed to measure.

Therefore, it is likely that PCI, combined with TMS/hd-EEG, can open an as yet unexplored path toward a neuroscience of consciousness, by allowing the empirical

investigation of the integrity of causal processes in the thalamocortical system that are supposedly required for awareness in general and not for the consciousness of this or that stimulus. If there is a characteristic directedness of intentionality towards its contents, our results suggest that this dynamic access of the conscious subject to the world formed by the contents of consciousness depends on the support to complex spatiotemporal cortical activity originated causally in the brain. This measure of the brain's capacity for consciousness would brings us closer to what can be properly understood as neural correlates of consciousness: we are not dealing any more with the reduction of something subjective into something objective, but with the direct investigation of the nature of an essentially objective character of the intentionality of consciousness.

It is worth noting at this point that some recent theoretical formulations about consciousness also pointed to certain aspects involved in the measure as implemented here, such as the necessity of engaging **diverse areas** of the thalamocortical system (Dehaene and Naccache, 2001), the **complexity** of cortical activation (Edelman and Tononi, 1994), or even the possibility of quantifying an **objective condition** of conscious experience through the "**integration** of information"[7](Tononi, 2004; 2008). However, such formulations have not been able to propose an objective measure capable of being tested, in a non-ambiguous and

[7] In particular, the Information Integration Theory of Consciousness (IITC) (Tononi, 2004; 2008) is an attempt to propose an objective measure of consciousness, regardless of specific mental contents and based on the integration of information in the thalamocortical system. However, the similarities with our work are superficial: such formulation presupposes a sense of the word "information" that is distinct from the sense implied by its use in the phenomenology of consciousness. We refer to the "informative" characteristics of consciousness when we say that it is the vehicle that provides us with a particular degree of access to certain contents and objects experienced. The theory, however, defines "information" as the reduction of the uncertainty about the possible states of the system: it is therefore information about the **internal** (neural) states of the conscious subject and not about the **objects of experience**. These two things, internal states and what is represented by such states, can only contain the same "information" if a certain internal state is equivalent to a given content of consciousness. However, this further depends on a "theory of meaning": on how a given internal representation is able to refer to things or states of affairs outside of it. The intentional relationship between representation and represented, which is the core of the "informativeness" property of consciousness and which can not be formalized (Searle, 2004; Putnam, 1999), is precisely what should be aimed at in attempting to solve the theoretical problem of consciousness (Putnam, 1988; Edelman, 2006; Rowlands, 2010). In the absence of a proper "theory of meaning", there is no convincing evidence, whether of a phenomenological or empirical basis, to justify the identification of consciousness with the ability to integrate "information" in the purely "syntactical" sense advocated by IITC. Moreover, beside the fact that this definition of "information" corresponds to an undesirable *petitio principii*, the proposed measure is in practice impossible to be calculated, for it depends on all the internal states that are available to the system via causal interactions among its parts. And finally, the double reduction present in the theory (from the content of consciousness to its neural representation, and from consciousness itself to a measure of the "syntactical" information contained in such representations) eventually leads to serious difficulties of interpretation, such as the confusion between the "amount of consciousness of one thing" and the "consciousness of a number of things" and the problematic identification of consciousness as a **capacity**, implying the attribution of consciousness to a subject even when, in his brain, "*no neurons were activated*" (Tononi, 2004).

successful way, on different models of unconsciousness, independent of the subject's engagement in specific mental contents. Moreover, it is crucial to note that these formulations turn out to have limited impact in the theoretical debate for a science of consciousness, since they do not describe clearly the kind of theoretical paradigm that such a science must presuppose. On the contrary, they usually rely on theoretical hypotheses based in a reductionist paradigm, begging the question not only about the nature of the contents of consciousness, by identifying them with internal representations of the subject, but also about the peculiar relationship between representation and represented that can give support to the phenomenal properties of intentionality.

The object of our empirical study, on the other hand, is not a **specific mental content**, but the brain's **vehicles** that give access to any conscious content. It is this crucial feature of the technique here implemented that provides the practical, technical and theoretical possibilities that have no precedent in the protocols generally used in the empirical study of consciousness. Just because this technique is potentially capable of empirically overpassing the content of consciousness, and therefore can directly test what is allegedly required to sustain such content, it is possible to avoid the problems of interpretation that affect the protocols based on ERP and fMRI: we do not measure something that can be considered an object of introspection, something supposedly accessible only by means purely subjective, but something that can only be accessed through an objective third-person perspective. For the same reason, several practical problems related to the application of the measure in cases of major clinical relevance, as in brain-injured patients, are overcome: this technique does not require the integrity of sensory pathways and the subject's ability or willingness to participate in the cognitive task associated with the experiment, because there is no specific cognitive task associated with the measure. Finally, if future developments confirm the potential of PCI and TMS in the investigation of neural processes relevant to consciousness, to the same extent that understanding the vehicle can clarify the nature of what it may lead us to, the complexity of cortical activation may establish new paradigms in the empirical investigation of the ultimate nature of the content of consciousness, contributing distinctly to the enormous challenge of establishing a science of the mind.

Appendix

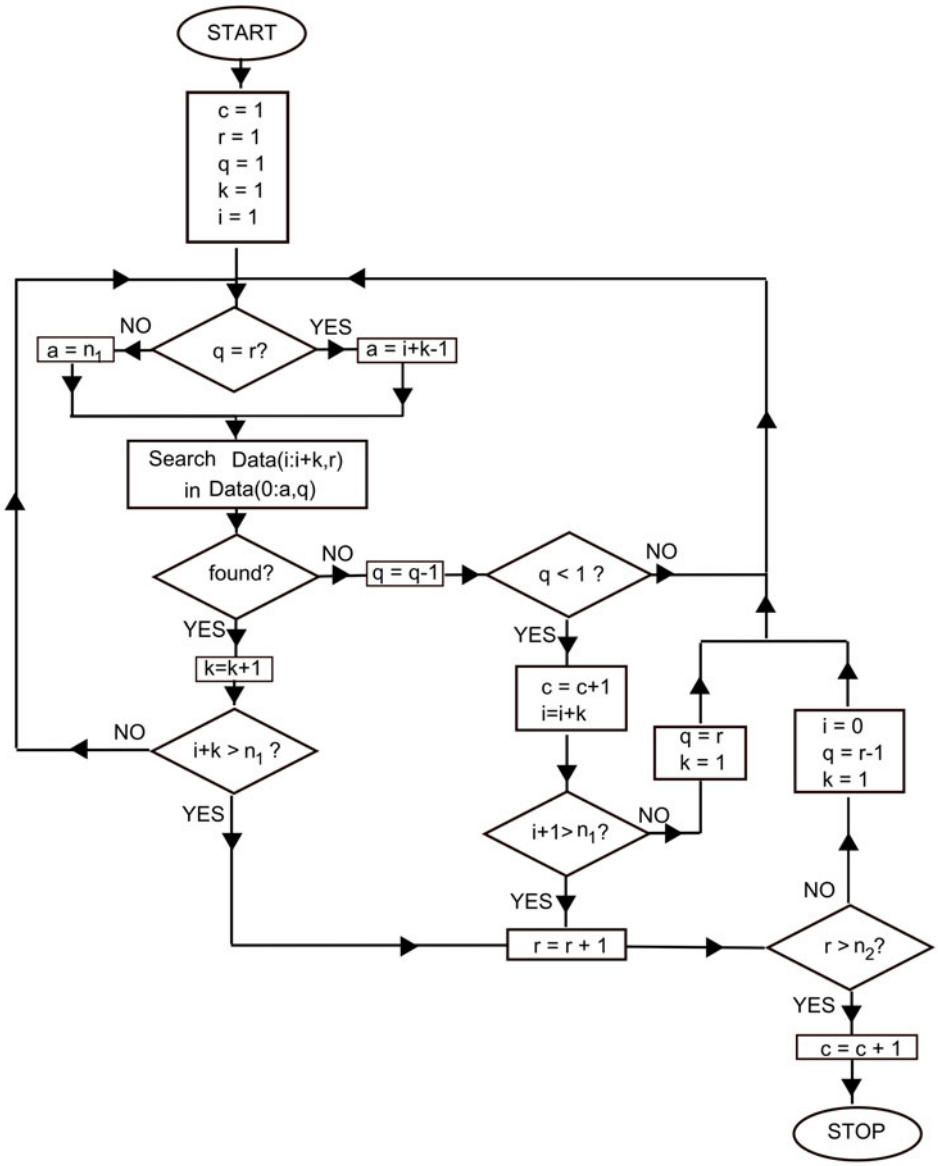

Figure A.1: Diagram of the algorithm used to calculate the Lempel-Ziv complexity of a binary matrix. The input of the algorithm is a binary $n_1 \times n_2$ matrix "Data", formed by entries $s(i, j) = 0$ or 1 for rows $i = 1\ldots n_1$ and columns $j = 1\ldots n_2$. Patterns of length k are sequences of k bits, Data($i{:}k$, j) = $s(i+1, j)s(i+2, j)\ldots s(i+k, j)$, with $0 < j+k \leq n_1$. The algorithm outputs the Lempel-Ziv complexity c of the matrix (adapted from Casali et al., 2012).

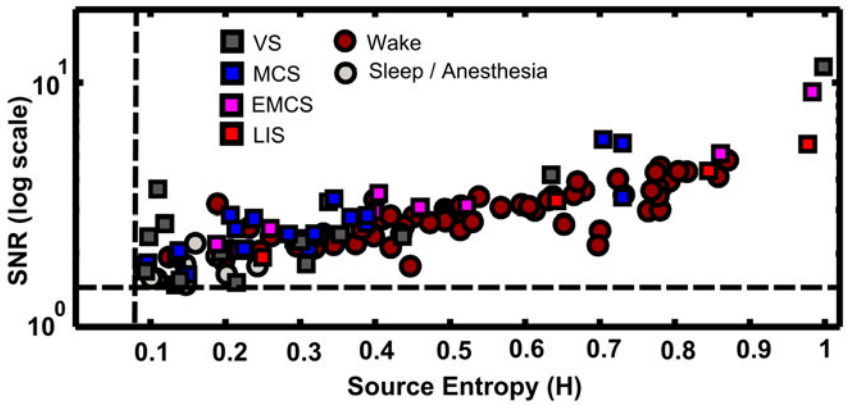

Figure A.2: Source Entropy (H) and Signal-to-noise ratio (SNR) are displayed for all sessions in which PCI was calculated. SNR is a measure of the quality of response to TMS calculated at the sensors level as the ratio of mean absolute amplitude of EEG in post-stimulus over the range of pre-stimulus. H is the average information contained in the spatial and temporal distribution of the response to TMS statistically significant relative to pre-stimulus (see Eq. 4.1). SNR and H were calculated based on the post-TMS interval of 0-300 ms. SNR> 1.5 (horizontal dashed line) and H > 0.08 (vertical dashed line) for all sessions. Neither the amplitude of the response (SNR) nor the total amount of statistical significance (H) seem to be correlated with the level of awareness in both healthy subjects (circle) and in brain injured patients (squares) (adapted from Casali et al., 2012).

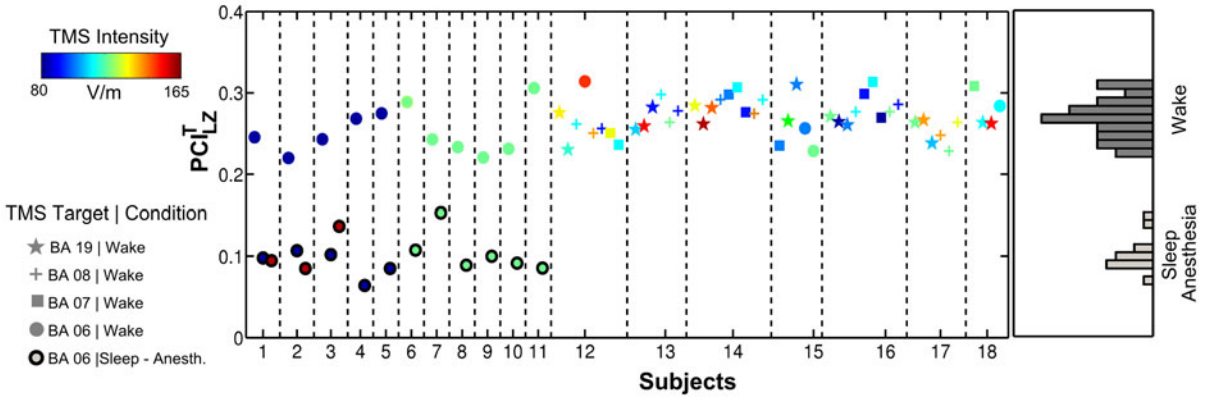

Figure A.3: PCI^{T}_{LZ} discriminated between conscious and unconscious healthy subjects, independently of the stimulated area and the stimulation intensity. PCI^{T}_{LZ} values for the same TMS sessions displayed in Figure 4.3. Similarly to PCI_{LZ}, PCI^{T}_{LZ} generated disjoint distributions between conscious (dark gray bars) and unconscious subjects (light gray bars). The PCI^{T}_{LZ} calculated in conscious subjects ranged between 0.22 and 0.31 (mean value of 0.27 ± 0.02), whereas the PCI^{T}_{LZ} calculated after loss of consciousness ranged between 0.06 and 0.15 (mean value of 0.09 ± 0.02). The p-value for the comparison between the PCI^{T}_{LZ} values for the conscious and unconscious groups was less than 10^{-8} (Wilcoxon rank sum test).

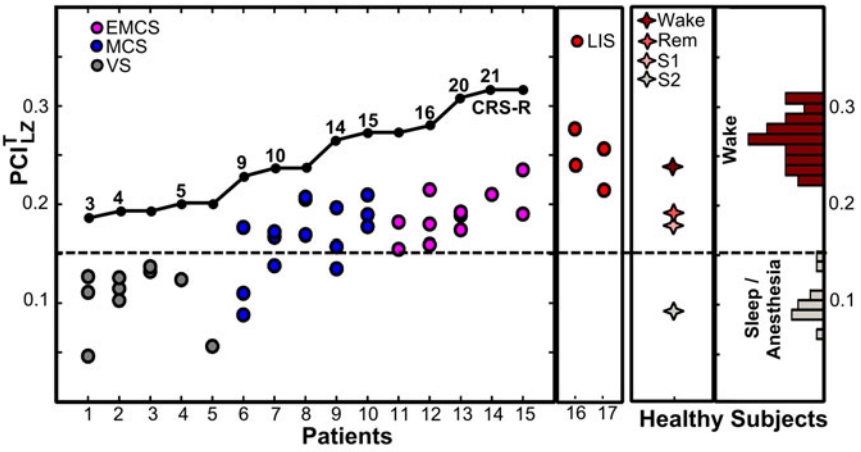

Figure A.4: PCI^{T}_{LZ} increased progressively in parallel with the level of consciousness. PCI^{T}_{LZ} values are shown for all TMS sessions displayed in Figure 4.7.

Table A.1: **Patients recovering from coma - Group I**

Sex (age; cause)	Time of TMS/EEG (days after insult)	CRS-R							
		Diagnosis at time of TMS/EEG	Auditory Function	Visual Function	Motor Function	Oromotor and Verbal Function	Communication	Arousal	Total score
Male (81; CVA)	19	VS	None	None	Flexion to pain	Oral reflexes	None	Without stimulation	5
Male (68; trauma)	21	VS	None	None	Flexion to pain	None	None	Without stimulation	4
Female (83; trauma)	14	VS	Startle reflex	None	Abnormal posturing	None	None	With stimulation	3
Female (15; CRA)	23	VS	Startle reflex	None	Abnormal posturing	Oral reflexes	None	Without stimulation	5
Male (19; trauma)	172	VS	Startle reflex	None	Abnormal posturing	Oral reflexes	None	With stimulation	4
Female (76; SaHem)	28	MCS	Startle reflex	Visual pursuit	Flexion to pain	Vocalization and oral movements	None	Without stimulation	10
Male (72; CVA)	38	MCS	Reproducible movement to command	Object Recognition	Flexion to pain	Vocalization and oral movements	Non-functional: intentional	With stimulation	14
Male (20; trauma)	1334	MCS	Reproducible movement to command	Visual pursuit	Abnormal posturing	Oral reflexes	None	Without stimulation	10
Female (38; trauma)	12	MCS	Reproducible movement to command	Visual pursuit	Automatic motor response	Vocalization and oral movements	Non-functional: intentional	With stimulation	15

Male (62; SaHem)	20	MCS	Startle reflex	Visual pursuit	Flexion to pain	Oral reflexes	None	Without stimulation	9
Male (45; CVA)	35	LIS	Systematic movement to command	Object recognition	None	Oral reflexes	Functional accurate	Attention	15
Female (25; CVA)	62	VS	Systematic movement to command	Object recognition	Flexion to pain	Vocalization and oral movements	Functional accurate	Attention	18

CRA: cardio respiratory arrest; CVA: cerebrovascular accident; SaHem: Subarachnoid Hemorrhage (adapted from Rosanova et al., 2012).

Table A.2: Patients recovering from coma - Group II

Sex (age; cause)	TMS/EEG session number	Time of TMS/EEG (days after insult)	CRS-R							
			Diagnosis at time of TMS/EEG	Auditory Function	Visual Function	Motor Function	Oromotor and Verbal Function	Communication	Arousal	Total score
Female (60; trauma)	1	15	VS	None	None	Flexion to pain	Vocalization and oral movements	None	With stimulation	5
	2	23	MCS	Reproducible movement to command	None	Flexion to pain	Oral reflexes	Functional intention	With stimulation	8
	3	30	EMCS	Systematic movement to command	Object Recognition	Functional use of objects	Vocalization and oral movements	Functional Accurate	With stimulation	20
Male (16; trauma)	1	12	VS	None	None	Flexion to pain	None	None	With stimulation	3
	2	19	MCS	Reproducible movement to command	None	Automatic motor reaction	Vocalization and oral movements	None	With stimulation	12
	3	40	EMCS	Systematic movement to command	Object Recognition	Automatic motor reaction	Vocalization and oral movements	Functional Accurate	Attention	21
Female (60; CVA)	1	35	VS	None	None	Abnormal posturing	Oral reflexes	None	With stimulation	3
	2	46	VS*	None	Blink to threat	Flexion to pain	Oral reflexes	None	With stimulation	5
	3	56	EMCS	Reproducible movement to command	Visual pursuit	Object manipulation	Intelligible verbalizations	Functional Accurate	With stimulation	16
Female (77; SaHem)	1	25	VS	Startle reflex	None	Abnormal posturing	Oral reflexes	None	With stimulation	3
	2	44	VS	Startle reflex	None	Flexion to pain	Oral reflexes	None	With stimulation	5

Male (43; CRA)	1	22	VS	None	None	None	Oral reflexes	None	With stimulation	3
	2	62	VS	Startle reflex	None	Abnormal posturing	Oral reflexes	None	With stimulation	5

*The day before this patient was diagnosed MCS with reproducible movements and visual pursuit (total score of 9).
CRA: cardio respiratory arrest; CVA: cerebrovascular accident; SaHem: Subarachnoid Hemorrhage (adapted from Rosanova et al., 2012).

References

Adams, J. H., Graham, D. I. and Jennett, B., "The neuropathology of the vegetative state after an acute brain insult." *Brain*. **123**:1327-38, 2000.

Andrews, K. "Medical decision making in the vegetative state: withdrawal of nutrition and hydration". *Neuro Rehabilitation* **19:**299-304, 2004.

Andrews, K., Murphy, L., Munday, R. and Littlewood, C., "Misdiagnosis of the vegetative state: retrospective study in a rehabilitation unit". *BMJ* **313:**13-16, 1996.

Austin, J., "Sense and Sensibilia", Oxford: *Oxford Univ. Press*, 1962.

Baars, J. B., "Global workspace theory of consciousness: toward a cognitive neuroscience of human experience", *Prog Brain Res.* **150**:45-53, 2005.

Babiloni, C., Vecchio, F., Miriello, M., Romani, G. L. and Rossini, P. M., "Visuo-spatial consciousness and parieto-occipital areas: a high-resolution EEG study". *Cereb. Cortex* **16**: 37–46, 2006.

Babiloni, F., Babiloni, C., Locche, L., Cincotti, F., Rossini, P. M. and Carducci, F., "High-resolution electro-encephalogram: source estimates of Laplacian-transformed somatosensory-evoked potentials using a realistic subject head model constructed from magnetic resonance images." *Med. Biol. Eng. Comput*. **38**:512-9, 2000.

Baillet, S., Mosher, J. C. and Leahy, R. M., "Electromagnetic brain mapping." *IEEE Signal Process. Mag.* **18**:14–30, 2001.

Bardin, J. C., Fins, J. J., Katz, D. I., Hersh, J., Heier, L. A., Tabelow, K., Dyke, J. P., Ballon, D. J., Schiff, N. D. and Voss, H. U., "Dissociations between behavioural and functional magnetic resonance imaging-based evaluations of cognitive function after brain injury." *Brain* **134**: 769-82, 2011.

Barker, A. T. "The history and basic principles of magnetic nerve stimulation" *Electroencephalogr. Clin. Neurophysiol. Suppl*. **51**:3-21, 1999.

Barker, A. T., Jalinous, R. and Freeston, I. L., "Non-invasive magnetic stimulation of human motor cortex", *Lancet* **1**:1106–7, 1985.

Bartels, A. and Zeki, S., "Functional brain mapping during free viewing of natural scenes." *Hum. Brain. Mapp.* **21**:75-85, 2004.

Bartels, A. and Zeki, S., "The chronoarchitecture of the cerebral cortex." *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **360**:733-50, 2005.

Bazhenov, M., Timofeev, I., Steriade, M. and Sejnowski, T. J., "Model of Thalamocortical Slow-Wave Sleep Oscillations and Transitions to Activated States." *J. Neurosci.* **22**:8691–704, 2002.

Bekinschtein, T. A., Dehaene, S., Rohaut, B., Tadel, F., Cohen, L. and Naccache, L., "Neural signature of the conscious processing of auditory regularities. *Proc. Natl. Acad. Sci. USA*, **106**: 1672–77, 2009.

Berg, P. and Scherg, M., "A fast method for forward computation of multiple-shell spherical head models." *Electroencephalogr. Clin. Neurophysiol.* **90**: 58–64, 1994.

Blumenfeld, H., "Consciousness and epilepsy: why are patients with absence seizures absent?" *Prog. Brain Res.* **150**:271-86, 2005.

Blumenfeld, H. and Taylor, J. "Why do seizures cause loss of consciousness?" *Neuroscientist.* **9**:301–10, 2003.

Blumenfeld, H., Westerveld, M., Ostroff, R. B., Vanderhill, S. D., Freeman, J., Necochea, A., Uranga, P., Tanhehco, T., Smith, A., Seibyl, J. P., Stokking, R., Studholme, C., Spencer, S. S. and Zubal, I. G., "Selective frontal, parietal, and temporal networks in generalized seizures." *Neuroimage* **19**:1556-66, 2003.

Boly, M., Coleman, M. R., Davis, M. H., Hampshire, A., Bor, D., Moonen, G., Maquet, P. A., Pickard, J. D., Laureys, S. and Owen, A. M., "When thoughts become action: An fMRI paradigm to study volitional brain activity in non-communicative brain injured patients". *Neuroimage* **36**:979–92, 2007.

Boly, M., Faymonville, M. E., Schnakers, C., Peigneux, P., Lambermont, B., Phillips, C., Lancellotti, P., Luxen, A., Lamy, M., Moonen, G., Maquet, P. and Laureys, S., "Perception of pain in the minimally conscious state with PET activation: an observational study." *Lancet Neurol.* 7**:**1013-20, 2008.

Boly, M., Garrido, M. I., Gosseries, O., Bruno, M. A., Boveroux, P., Schnakers, C., Massimini, M., Litvak, V., Laureys, S. and Friston, K., "Preserved feedforward but impaired top-down processes in the vegetative state". *Science* **332**: 858-62, 2011.

Bonato, C., Miniussi, C. and Rossini, P. M., "Transcranial magnetic stimulation and cortical evoked potentials: a TMS/EEG co-registration study". *Clin. Neurophysiol.* **117**: 1699–07, 2006.

Bonebakker, A., Bonke, B., Klein, J., Wolters, G., Stijnen, T., Passchier, J. and Merikle, P. M., "Information processing during general anaesthesia: evidence for unconscious memory". In: Bonke, B., Bovill, J. G. W. and Moerman, N. (eds), "Memory and Awareness in Anaesthesia", Amsterdam: *Swets and Zeitlinger*, 101–109, 1996.

Brentano, F. "Psychology from an Empirical Standpoint", London: *Routledge and Kegan Paul*, 1973 [1874].

Brooks, D. H., Ahmad, G. F., MacLeod, R. S. and Maratos, G. M., "Inverse electrocardiography by simultaneous imposition of multiple constraints". *IEEE Trans. Biomed. Eng.* **46**:3 – 17, 1999.

Bullock, T. H., McClune, M. C., Achimowicz, J. Z., Iragui-Madoz, V. J., Duckrow, R. B. and Spencer, S. S., "Temporal fluctuations in coherence of brain waves". *Proc. Natl. Acad. Sci. USA*: **92**, 11568–72, 1995.

Casali, A. G., Casarotto, S., Rosanova, M., Mariotti, M. and Massimini, M. "General indices to characterize the electrical response of the cerebral cortex to TMS". *NeuroImage*, **49**: 1459-68, 2010.

Casali, A. G., Rosanova, M., Gosseries, O., Ferrarelli, F., Casarotto, S., D'Oliveira, T., Casali, K. R., Boly, M., Laureys, S., Tononi, G. and Massimini, M.,. "Gauging brain complexity in wakefulness, sleep, anesthesia and coma" (*submitted for publication*), 2012.

Casarotto, S., Romero Lauro, L. J., Bellina, V., Casali, A. G., Rosanova, M., Pigorini, A., Defendi, S., Mariotti, M. and Massimini, M., "EEG Responses to TMS Are Sensitive to Changes in the Perturbation Parameters and Repeatable over Time". *Plos ONE*, **5**:e10281, 2010.

Chaitin, G. "On the length of programs for computing finite binary sequences". *Journal of the ACM*, **13**:547-69, 1966.

Chernik, D. A., Gillings, D., Laine, H., Hendler, J., Silver, J. M., Davidson, A. B., Schwam, E. M. and Siegel, J. L. "Validity and reliability of the Observer's Assessment of Alertness/Sedation Scale: study with intravenous midazolam". *J. Clin. Psychopharmacol.* **10**:244–51, 1990.

Colby, C. L., Gattass, R., Olson, C. R. and Gross, C. G., "Topographical organization of cortical afferents to extrastriate visual area PO in the macaque: a dual tracer study". *J. Comp. Neurol*. **269**:392–413, 1988.

Crick, F. and Koch, C. "Towards a neurobiological theory of consciousness". *Sem. Neurosci.* **2**:263-75, 1990.

Crick, F. and Koch, C. "A framework for consciousness." *Nat. Neurosci.* **6**, 119–26, 2003.

Cudeiro, J. and Sillito, A. M., "Looking back: corticothalamic feedback and early visual processing." *Trends Neurosci.* **29**:298-306, 2006.

Damasio, A. R. "Descartes' Error: Emotion, Reason and the Human Brain*"*. New York: *Grosset/Putnam*, 1994.

Davis, M. H., Coleman, M. R., Absalom, A. R., Rodd, J. M., Johnsrude, I. S., Matta, B. F., Owen, A. M. and Menon, D. K., "Dissociating speech perception and comprehension at reduced levels of awareness". *Proc. Natl Acad. Sci*. *USA* **104**, 16032–7, 2007.

Dehaene, S. and Naccache, L., "Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework." *Cognition* **79:** 1–37, 2001.

Dehaene, S., Naccache, L., Le Clec, H. G., Koechlin, E., Mueller, M., Dehaene-Lambertz, G., van de Moortele, P. F. and Le Bihan, D., "Imaging unconscious semantic priming". *Nature* **395**: 597–600, 1998.

Dehaene, S., Naccache, L., Cohen, L., Bihan, D. L., Mangin, J. F., Poline, J. B., and Riviere, D. "Cerebral mechanisms of word masking and unconscious repetition priming". *Nat. Neurosci.* **4**: 752–8, 2001b.

Dehaene, S., Sergent, C. and Changeux, J. P., "A neuronal network model linking subjective reports and objective physiological data during conscious perception." *Proc. Natl. Acad. Sci. USA*. **100**:8520-5, 2003.

Del Cul, A., Baillet, S. and Dehaene, S. "Brain dynamics underlying the nonlinear threshold for access to consciousness." *Plos Biol*. **5**:e260, 2007.

Delorme, A. and Makeig, S., "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis". *J. Neurosci. Methods* **134**:9-21, 2004.

Dennett, D.C., "Consciousness Explained", London: *Penguin*, 1991.

Descartes, R., "Meditations on First Philosophy", John Cottingham (ed.), Cambridge: *Cambridge Press*, 1996 [1641].

Destexhe, A., Hughes, S. W., Rudolph, M. and Crunelli, V.,"Are corticothalamic 'up' states fragments of wakefulness?" *Trends Neurosci*. **30**:334-42, 2007.

Diekhof, E. K., Biedermann, F., Ruebsamen, R. and Gruber, O. "Top-down and bottom-up modulation of brain structures involved in auditory discrimination". *Brain Res.* **1297**: 118-23, 2009.

Doesburg, S. M., Green, J. J., McDonald, J. J. and Ward, L .M., "Rhythms of consciousness: binocular rivalry reveals large-scale oscillatory network dynamics mediating visual perception." *Plos ONE* **4**:e6142, 2009.

Eccles, J.C., "The Human Psyche", New York: *Springer*, 1980.

Edelman, G., "Second Nature: Brain, Science and Human Knowledge", New Haven: *Yale Univ. Press*, 2006.

Engel, A. K. and Singer, W., "Temporal binding and the neural correlates of sensory awareness." *Trends Cogn. Sci*. **5**:16-25, 2001.

Esser, S. K., Huber, R., Massimini, M., Peterson, M. J., Ferrarelli, F., and Tononi, G., "A direct demonstration of cortical LTP in humans: a combined TMS/EEG study". *Brain Res. Bull.* **69**: 86–94, 2006.

Etcoff, N. L., Freeman, R. and Cave, K. R., "Can we lose memories of faces? Content specificity and awareness in a prosopagnosic". *J. Cog. Neurosci.* **3**:25–41, 1991.

Faugeras, F., Rohaut, B., Weiss, N., Bekinschtein, T. A., Galanaud, D., Puybasset, L., Bolgert, F., Sergent, C., Cohen, L., Dehaene, S. and Naccache, L., "Probing consciousness with event-related potentials in the vegetative state". *Neurology* **77**:264-8, 2011.

Fell, J., Roschke, J. and Beckmann, P. "Deterministic chaos and the rst positive Lyapunov exponent: A nonlinear analysis of the human electroencephalograms during sleep." *Biol. Cybern.* **69**:139-146, 1993.

Fellinger, R., Klimesch, W., Schnakers, C., Perrin, F., Freunberger, R., Gruber, W., Laureys, S. and Schabus, M., "Cognitive processes in disorders of consciousness as revealed by EEG time-frequency analyses". *Clinical neurophysiol.*, 2011 [Epub ahead of print].

Fernandez-Duque, D., Grossi, G., Thornton, I. M. and Neville, H. J., "Representation of change: separate electrophysiological markers of attention, awareness, and implicit processing". *J. Cogn. Neurosci.* **15**: 491–507, 2003.

Ferrarelli, F., Massimini, M., Sarasso, S., Casali, A. G., Riedner, B. A., Angelini, G., Tononi, G. and Pearce, R. A., "Breakdown in cortical effective connectivity during midazolam-induced loss of consciousness". *Proc. Natl. Acad. Sci. USA*, **107**: 2681-86, 2010.

Fischer, C., Luaute, J. and Morlet, D., "Event-related potentials (MMN and novelty P3) in permanent vegetative or minimally conscious states." *Clin. Neurophysiol.* **121**:1032-42, 2010.

Franks, N. P. "General anaesthesia: from molecular targets to neuronal pathways of sleep and arousal". *Nat. Rev. Neurosci.* **9**:370–86, 2008.

Frege, G., "On Sense and Reference." In: Geach, P. and Black, M. (eds.) "Translations from the Philosophical Writings of Gottlob Frege", Oxford: *Blackwell*, 1960 [1892].

Friston, K., "Statistics I: Experimental Design and Statistical Parametric Mapping", in: Toga, A. W. and Mazziota, J. C., "Brain Mapping: The Methods" (2nd ed.) *Academic Press*, 2002.

Friston, K. J., Penny, W., Phillips, C., Kiebel, S., Hinton, G. and Ashburner, J., "Classical and Bayesian inference in neuroimaging: theory". *NeuroImage* **16**:465–83, 2002.

Friston, K. J., Henson, R., Phillips, C. and Mattout, J., "Bayesian estimation of evoked and induced responses." *Hum. Brain Mapp.* **27**:722–35, 2006.

Fründ, I., Busch, N. A., Schadow, J., Körner, U. and Herrmann, C. S., "From perception to action: phase-locked gamma oscillations correlate with reaction times in a speeded response task". *BMC Neurosci.* **8**, 27 – 37, 2007.

Gaillard, R., Dehaene, S., Adam, C., Clémenceau, S., Hasboun, D., Baulac, M., Cohen, L. and Naccache, L., "Converging intracranial markers of conscious access." *Plos Biol.* **7**:e61, 2009.

Gallagher, S. and Zahavi, D., "The Phenomenological Mind", New York: *Routledge*, 2008.

Gan, T. J., "Pharmacokinetic and pharmacodynamic characteristics of medications used for moderate sedation". *Clin. Pharmacokinet.* **45**:855-69, 2006.

Garrido, M. I., Kilner, J. M., Kiebel, S. J. and Friston, K. J., "Evoked brain responses are generated by feedback loops". *Proc. Natl. Acad. Sci. U.S.A,* **104**:20961-6, 2007.

Giacino, J. T., "The vegetative and minimally conscious states: consensus-based criteria for establishing diagnosis and prognosis." *Neuro Rehabilitation* **19**:293-8, 2004.

Giacino, J. T., Ashwal, S., Childs, N., Cranford, R., Jennett, B., Katz, D. I., Kelly, J. P., Rosenberg, J. H., Whyte, J., Zafonte, R. D. and Zasler, N. D., "The minimally conscious state: definition and diagnostic criteria". *Neurology* **58**: 349-53, 2002.

Giacino, J. T., Kalmar, K. and Whyte, J., "The JFK Coma Recovery Scale-Revised: measurement characteristics and diagnostic utility". *Arch. Phys. Med. Rehabil.* **85**: 2020-9, 2004.

Giacino, J. T., Schnakers, C., Rodriguez-Moreno, D., Kalmar, K., Schiff, N. and Hirsch, J., "Behavioral assessment in patients with disorders of consciousness: gold standard or fool's gold?" *Prog. Brain Res*. **177**: 33-48, 2009.

Gill-Thwaites, H., "Lotteries, loopholes and luck: misdiagnosis in the vegetative state patient." *Brain Inj.* **20:**1321-8, 2006.

Gil, Z., Connors, B. W. and Amitai, Y., "Differential regulation of neocortical synapses by neuromodulators and activity." *Neuron* **19:**679–86, 1997.

Gottlieb, J. P., Kusunoki, M. and Goldberg, M. E., "The representation of visual salience in monkey parietal cortex". *Nature* **391**:481–4, 1998.

Greenfield, S. A. and Collins, T. F., "A neuroscientific approach to consciousness". *Prog. Brain. Res.* **150**:11-23, 2005,

Gu, F., Meng, X., Shen, E. and Cai, Z., "Can We Measure Consciousness with EEG Complexities?" *I. J. Bifurcation and Chaos* **13**: 733-742, 2003.

Guariglia, C., Padovani, A., Pantano, P. and Pizzamiglio, L., "Unilateral neglect restricted to visual imagery." *Nature* **364**:235-7, 1993.

Guillery, R. W. and Sherman, S. M., "Thalamic Relay Functions and Their Role in Corticocortical Communication: Generalizations from the Visual System". *Neuron*, **33**: 163–75, 2002.

Halgren, E., Marinkovic, K. and Chauvel, P. "Generators of the late cognitive potentials in auditory and visual oddball tasks." *Electroencephalogr. Clin. Neurophysiol.* **106**: 156–64, 1998.

Hall, H., "Intentionality and World: Division I of Being and Time", In: Guignon, C. (ed.), "The Cambridge Companion to Heidegger", Cambridge: *Cambridge Press*, pp. 122-139, 1993.

Hallett, M. "Transcranial magnetic stimulation and the human brain". *Nature* **406**:147-50, 2000.

Hamzei, F., Knab, R.,Weiller, C. and Röther J., "The influence of extra- and intracranial artery disease on the BOLD signal in FMRI". *Neuroimage* **20**:1393–9, 2003.

Harrison, L. M., Stephan, K. E., Rees, G. and Friston, K. J., "Extra-classical receptive field effects measured in striate cortex with fMRI." *Neuroimage* **34**:1199-208, 2007.

Hauk, O., "Keep it simple: a case for using classical minimum norm estimation in the analysis of EEG and MEG data." *Neuroimage* **21**:1612-21, 2004.

Haynes, J. D., Driver, J. and Rees, G., "Visibility reflects dynamic changes of effective connectivity between V1 and fusiform cortex." *Neuron* **46**: 811–821, 2005.

Haynes, J. D., Deichmann, R. and Rees, G., "Eye-specific effects of binocular rivalry in the human lateral geniculate nucleus." *Nature* **438**: 496–499, 2005b.

Heidegger, M. "Being and Time: a revised edition of the Stambaugh translation", *State Univ. of New York Press*, 1995 [1927].

Hobbes, T., "Leviathan", R. Tuck (ed.). Cambridge: *Cambridge Press*, 1997 [1651].

Hobson, J. A., Pace-Schott, E. F. and Stickgold, R., "Dreaming and the brain: toward a cognitive neuroscience of conscious states". *Behav. Brain. Sci.* **23**:793-842, 2000.

Hobson, J. A. and Pace-Schott, E. F. "The cognitive neuroscience of sleep: neuronal systems, consciousness and learning." *Nat. Rev. Neurosci.* **3**:679–93, 2002.

Holler, Y., Bergmann, J., Kronbichler, M., Crone, J. S., Schmid, E. V., Golaszewski, S. and Ladurner, G., "Preserved oscillatory response but lack of mismatch negativity in patients with disorders of consciousness." *Clin. Neurophysiol.* **122**:1744-54, 2011.

Huang, X., Albright, T. D. and Stoner G. R., "Adaptive surround modulation in cortical area MT", *Neuron* **53**:761-70, 2007.

Huber, R., Mäki, H., Rosanova, M., Casarotto, S., Canali, P., Casali A. G., Tononi, G. and Massimini, M., "Human cortical excitability increases with time awake", *Cereb. Cortex* (in press), 2012.

Hughes, S. W., Lorincz, M., Cope, D. W., Blethyn, K. L., Kekesi, K. A., Parri, H. R., Juhasz, G. and Crunelli, V., "Synchronized oscillations at alpha and theta frequencies in the lateral geniculate nucleus". *Neuron* **42**:253-68, 2004.

Hupé, J. M., James, A. C., Payne, B. R., Lomber, S. G., Girard, P. and Bullier, J., "Cortical feedback improves discrimination between figure and background by V1, V2 and V3 neurons". *Nature* **394**:784-7, 1998.

Husserl, E., "The Shorter Logical Investigations". New York: *Routledge*, 2001 [1913].

Ilmoniemi, R. J., Virtanen, J., Ruohonen, J., Karhu, J., Aronen, H. J., Näätänen, R. and Katila, T., "Neuronal responses to magnetic stimulation reveal cortical reactivity and connectivity". *Neuroreport* **8**: 3537–40, 1997.

Imas, O. A., Ropella, K. M., Ward, B. D., Wood, J. D. and Hudetz, A. G., "Volatile anesthetics enhance flash-induced gamma oscillations in rat visual cortex." *Anesthesiology* **102**:937-47, 2005.

Iramina, K., Maeno, T., Nonaka, Y. and Ueno, S., "Measurement of evoked electroencephalography induced by transcranial magnetic stimulation." *J. Appl. Phys.* **93**: 6718–20, 2003.

James, W., "The Principles of Psychology", New York: *Henry Holt*, 1890.

James, W., "Essays in Radical Empiricism", In: Bowers, F. and Skrupskelis, I. (eds) "The Work of William James", Cambridge: *Harvard Univ. Press*, 1976 [1912].

Johnson, J., Kundu, B., Casali, A. G. and Postle, B., "Effects of Physiological State on Temporal and Spectral Properties of the TMS-Evoked Response" (*submitted for publication*), 2011.

Kahan, T. L. and Laberge, S. P., "Dreaming and waking: Similarities and differences revisited." *Conscious Cogn.* **20**:494-514, 2011.

Kähkönen, S., Wilenius, J., Komssi, S. and Ilmoniemi, R. J., "Distinct differences in cortical reactivity of motor and prefrontal cortices to magnetic stimulation". *Clin. Neurophysiol.* **115**: 583–8, 2004.

Kähkönen, S. "MEG and TMS combined with EEG for mapping alcohol effects". *Alcohol* **37**: 129–33, 2005.

Kähkönen, S., Komssi, S., Wilenius, J. and Ilmoniemi, R. J., "Prefrontal TMS produces smaller EEG responses than motor-cortex TMS: implications for rTMS treatment in depression". *Psychopharmacology* (Berl) **181**: 16–20, 2005.

Kähkönen, S., Komssi, S., Wilenius, J. and Ilmoniemi, R. J., "Prefrontal transcranial magnetic stimulation produces intensity-dependent EEG responses in humans". *NeuroImage* **24**: 955–60, 2005b.

Kakigi, R., Naka, D., Okusa, T., Wang, X., Inui, K., Qiu, Y., Tran, T. D., Miki, K., Tamura, Y., Nguyen, T. B., Watanabe, S. and Hoshiyama, M., "Sensory perception during sleep in humans: a magnetoencephalograhic study." *Sleep Med.* **4**:493-507, 2003.

Kandel, E. R., Schwartz, J. H. and Jessell, T. M., "Principles of Neural Science", New York: *McGraw-Hill* (4th ed.), 2000.

Kant, I. "Critique of Pure Reason", in: Guyer, P. and Wood, A. W. (eds), "The Cambridge Edition of the Works of Immanuel Kant", Cambridge: *Cambridge Press*, 1998 [1787].

Kaspar, F. and Schuster, H. G., "Easily calculable measure for the complexity of spatiotemporal patterns". *Phys. Rev. A* **36**:842–8, 1987.

Kastner S., Demmer I. and Ziemann U., "Transient visual field defects induced by transcranial magnetic stimulation over human occipital pole." *Exp. Brain Res.* **118**:19-26, 1998.

Khan, A. Z., Pisella, L., Vighetto, A., Cotton, F., Luauté, J., Boisson, D., Salemme, R., Crawford, J. D. and Rossetti, Y., "Optic ataxia errors depend on remapped, not viewed, target location". *Nat. Neurosci.* **8**:418-20, 2011.

Koch, C., "The Quest for Consciousness: a neurobiological approach", Englewood: *Roberts & Co*, 2004.

Koivisto, M., Lähteenmäki, M., Sørensen, T. A., Vangkilde, S., Overgaard, M. and Revonsuo, A., "The earliest electrophysiological correlate of visual awareness?" *Brain Cogn*. **66**: 91–103, 2008.

Kolb, B. and Whishaw, I., "Fundamentals of Human Neuropsychology." New York: *W.H. Freeman and Co*., 1990.

Kolmogorov, A. "Three approaches to the quantitative definition of information". *Problems of Information Transmission*, **1**:1-11, 1965.

Komssi, S., Aronen, H. J., Huttunen, J., Kesäniemi, M., Soinne, L., Nikouline, V. V., Ollikainen, M., Roine, R. O., Karhu, J., Savolainen, S. and Ilmoniemi, R. J., "Ipsi- and contralateral EEG reactions to transcranial magnetic stimulation". *Clin. Neurophysiol.* **113**: 175–84, 2002.

Komssi, S., Kähkönen, S. and Ilmoniemi, R.J., "The effect of stimulus intensity on brain responses evoked by transcranial magnetic stimulation". *Hum. Brain Mapp.* **21**:154–64, 2004.

Komssi, S., Huttunen, J., Aronen, H. J. and Ilmoniemi, R. J., "EEG minimum-norm estimation compared with MEG dipole fitting in the localization of somatosensory sources at S1." *Clin. Neurophysiol*. **115**:534-42, 2004b.

Komssi, S. and Kähkönen, S., "The novelty value of the combined use of electroencephalography and transcranial magnetic stimulation for neuroscience research." *Brain Research Reviews* **52**:183–92, 2006.

Komssi, S., Savolainen, P., Heiskala, J. and Kähkönen, S., "Excitation threshold of the motor cortex estimated with transcranial magnetic stimulation electroencephalography." *NeuroReport* **18**:13–6, 2007.

Kosslyn, S. M., Pascual-Leone, A., Felician, O., Camposano, S., Keenan, J. P., Thompson, W. L., Ganis, G., Sukel, K. E. and Alpert, N. M., "The role of area 17 in visual imagery: convergent evidence from PET and rTMS." *Science* **284**:167-70, 1999.

Kotchoubey, B., Lang, S., Mezger, G., Schmalohr, D., Schneck, M., Semmler, A., Bostanov, V. and Birbaumer, N., "Information processing in severe disorders of consciousness: vegetative state and minimally conscious state". *Clin. Neurophysiol*. **116**: 2441-53, 2005.

Kouider, S. and Dehaene, S., "Levels of processing during non-conscious perception: A critical review of visual masking." *Philos. Trans. R .Soc. Lond. B Biol. Sci.* **362**: 857–75, 2007.

Kraemer, D. J., Macrae, C. N., Green, A. E. and Kelley, W. M., "Musical imagery: sound of silence activates auditory cortex." *Nature* **434**:158, 2005.

Kreiman, G., Koch, C. and Fried, I., "Imagery neurons in the human brain." *Nature* **408**:357-61, 2000.

Kroeger, D. and Amzica, F., "Hypersensitivity of the anesthesia-induced comatose brain." *J. Neurosci.* **27**:10597-607, 2007.

Laarne, P. H., Tenhunen-Eskelinen, M. L., Hyttinen, J. K. and Eskola, H. J., "Effect of EEG electrode density on dipole localization accuracy using two realistically shaped skull resistivity models". *Brain Topogr.* **12**: 249–54, 2000.

Lamme, V. A. and Roelfsema, P. R., "The distinct modes of vision offered by feedforward and recurrent processing". *Trends Neurosci.* **23**:571-9, 2000.

Lamy, D., Salti, M., and Bar-Haim, Y., "Neural correlates of subjective awareness and unconscious processing: an ERP study." *J. Cogn. Neurosci.* **21**: 1435–46, 2009.

Långsjö, J. W., Maksimow, A., Salmi, E., Kaisti, K., Aalto, S., Oikonen, V., Hinkka, S., Aantaa, R., Sipilä, H., Viljanen, T., Parkkola, R. and Scheinin, H., "S-ketamine anesthesia increases cerebral blood flow in excess of the metabolic needs in humans." *Anesthesiology* **103**:258-68, 2005.

Laureys, S. and Boly, M., "What is it like to be vegetative or minimally conscious?" *Curr. Opin. Neurol.* **20**:609–13, 2007.

Laureys, S., Owen, A. M. and Schiff, N. D., "Brain function in coma, vegetative state and related disorders". *Lancet neurology* **3**: 537-46, 2004.

Lehman, D. and Skrandies, W., "Reference-free identification of components of checkerboard-evoked multichannel potential fields". *Electroencephalogr. Clin. Neurophysiol.* **48**:609-21, 1980.

Lempel, A. and Ziv, J.,"On complexity of finite sequences", *IEEE Trans. Inform. Theory* **22**:75-88, 1976.

Levine, J. "Materialism and Qualia: The Explanatory Gap", *Pacific Philosophical Quarterly* **64**: 354-61, 1983.

Li, M. and Vitányi, P., "An Introduction to Kolmogorov Complexity and its Applications", New York: *Springer* (3rd ed.), 2008.

Libedinsky, C. and Livingstone, M. "Role of prefrontal cortex in conscious visual perception." *J. Neurosci.* **31**:64-9, 2011.

Libet, B. “Neural processes in the production of conscious experience”, in M. Velmans (ed.), “The Science of Consciousness: Psychological, Neuropsychological and Clinical Reviews”, London: *Routledge*, 1996.

Libet, B. “Brain stimulation in the study of neuronal functions for conscious sensory experiences”. *Human Neurobiology* **1**:235-242, 1982.

Lioumis, P., Kicić, D., Savolainen, P., Mäkelä, J. P. and Kähkönen, S., “Reproducibility of TMS-Evoked EEG responses”. *Hum. Brain Mapp.* **30**: 1387–96, 2009.

Llinás, R. R. and Paré, D., “Of dreaming and wakefulness.” *Neuroscience* **44,** 521–35, 1991.

Llinás, R., Ribary, U., Contreras, D. and Pedroarena, C. “The neuronal basis for consciousness”. *Philos Trans R Soc Lond B Biol Sci.* **353**:1841-9, 1998.

Llinás, R. R., Choi, S., Urbano, F. J. and Shin, H. S., “Gamma-band deficiency and abnormal thalamocortical activity in P/Q-type channel mutant mice”. *Proc. Natl. Acad. Sci. USA* **104**:17819-24, 2007.

Lv J., Simpson, D. M. and Bell, S. L. “Objective detection of evoked potentials using a bootstrap technique”. *Med. Eng. Phys*. **29**: 191-8, 2007.

Maccabee, P. J., Amassian, V. E., Eberle, L. P. and Cracco, R. Q., “Magnetic coil stimulation of straight and bent amphibian and mammalian peripheral nerve in vitro: locus of excitation”. *J. Physiol.* **460**: 201–19, 1993.

Majerus, S., Gill-Thwaites, H., Andrews, K. and Laureys, S. “Behavioral evaluation of consciousness in severe brain damage”. *Prog. Brain Res.* **150**: 397-413, 2005.

Maksimow, A., Särkelä, M., Långsjö, J. W., Salmi, E., Kaisti, K. K., Yli-Hankala, A., Hinkka-Yli-Salomäki, S., Scheinin, H. and Jääskeläinen, S. K., “Increase in high frequency EEG activity explains the poor performance of EEG spectral entropy monitor during S-ketamine anesthesia.” *Clin Neurophysiol.* **117**:1660-8, 2006.

Maldjian, J. A., Laurienti, P. J., Kraft, R. A. and Burdette, J. H., “An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets.” *NeuroImage* **19**:1233–9, 2003.

Maldjian, J. A., Laurienti, P. J. and Burdette, J. H., “Precentral gyrus discrepancy in electronic versions of the Talairach atlas.” *NeuroImage* **21**: 450–5, 2004.

Mantini, D., Corbetta, M., Perrucci, M. G., Romani, G. L., and Del Gratta, C., “Large-scale brain networks account for sustained and transient activity during target detection.” *NeuroImage* **44**: 265–74, 2009.

Markowitsch, H. J. and Kessler, J., “Massive impairment in executive functions with partial preservation of other cognitive functions: the case of a young patient with severe degeneration of the prefrontal cortex”. *Exp. Brain. Res.* **133**: 94-102, 2000.

Massimini, M., Ferrarelli, F., Huber, R., Esser, S. K., Singh, H. and Tononi, G., “Breakdown of cortical effective connectivity during sleep”. *Science* **309**: 2228–32, 2005.

Massimini, M., Ferrarelli, F., Esser, S. K., Riedner, B. A., Huber, R., Murphy, M., Peterson, M. J. and Tononi, G., "Triggering sleep slow waves by transcranial magnetic stimulation". *Proc. Natl. Acad. Sci. USA*. **104**: 8496–501, 2007.

Massimini, M., Boly, M., Casali, A. G., Rosanova, M. and Tononi, G., "A perturbational approach for evaluating the brain's capacity for consciousness". *Progress Brain Res.*, **177**: 201-14, 2009.

Massimini, M., Ferrarelli, F., Murphy, M., Huber, R., Riedner, B., Casarotto, S. and Tononi G., "Cortical reactivity and effective connectivity during REM sleep in humans." *Cogn. Neurosci*. **1**:176-83, 2010.

Mataro, M., Jurado, M. A., Garcia-Sanchez, C., Barraquer, L., Costa-Jussa, F. R., Junque, C., "Long-term effects of bilateral frontal brain lesion: 60 years after injury with an iron bar". *Arch. Neurol.* **58**: 1139-42, 2001.

Mattout, J., Phillips, C., Penny, W. D., Rugg, M. D. and Friston, K. J., "MEG source localization under multiple constraints: an extended Bayesian framework". *NeuroImage* **30**:753–67, 2006.

McCormick, D. A. and Bal, T., "Sleep and Arousal: Thalamocortical Mechanisms", *Annu. Rev. Neurosci.* **20**:185–215, 1997.

McCubbin, J., Yee, T., Vrba, J., Robinson, S. E., Murphy, P., Eswaran, H. and Preissl, H., "Bootstrap significance of low SNR evoked response". *J. Neurosci. Methods* **168**: 265-72, 2008.

McDowell, J., "Mind and World", Cambridge: *Harvard Univ. Press*, 1996.

McRobbie, D. and Foster, M. A., "Thresholds for biological effects of time-varying magnetic fields". *Clin. Phys. Physiol. Meas.* **5**: 67–78, 1984.

Melloni, L., Molina, C., Pena, M., Torres, D., Singer, W. and Rodriguez, E., "Synchronization of neural activity across cortical areas correlates with conscious perception." *J. Neurosci.* **27**: 2858–65, 2007.

Menary, R. (editor), "The Extended Mind", Cambridge: *MIT Press*, 2010.

Merleau-Ponty, M., "Phenomenology of Perception", New York: *Routledge*, 1996 [1945].

Meyer B. U., Diehl, R., Steinmetz, H., Britton, T. C. and Benecke, R., "Magnetic stimuli applied over motor and visual cortex: influence of coil position and field polarity on motor responses, phosphenes, and eye movements.", *Electroencephalogr. Clin. Neurophysiol. Suppl.* **43**:121-34, 1991.

Miyashita, Y., "How the brain creates imagery: projection to primary visual cortex." *Science* **268**:1719-20, 1995.

Monti, M. M., Vanhaudenhuyse, A., Coleman, M. R., Boly, M., Pickard, J. D., Tshibanda, L., Owen, A. M. and Laureys, S., "Willful modulation of brain activity in disorders of consciousness." *N. Engl. J. Med.* **362**: 579-89, 2010.

Monti, M. M., Coleman, M. R. and Owen, A. M., "Neuroimaging and the vegetative state: resolving the behavioral assessment dilemma?" *Ann. N.Y. Acad. Sci.* **1157**:81-9, 2009.

Moran, J. and Desimone, R., "Selective Attention Gates Visual Processing in the Extrastriate Cortex". *Science* **229**:782-4, 1985.

Mosher, J. C., Leahy, R. M. and Lewis, P. S., "EEG and MEG: forward solutions for inverse methods." *IEEE Trans. Biomed. Eng.* **46**: 245– 59, 1999.

Motter, B. C., "Neural correlates of attentive selection for color or luminance in extrastriate area V4". *J. Neurosci.* **14**: 2178–89, 1994.

Nagel, T., "What is it like to be a bat?", *Philosophical Review* **83**:435–51, 1974.

Noë, A., "Action in Perception", Cambridge: *MIT Press*, 2004.

Nowak, L. G. and Bullier, J., "The timing of information transfer in the visual system". In: Kaas, J. H., Rockland, K. and Peters, A. (eds), "Cerebral cortex - volume 12: Extrastriate Cortex in Primates". New York: *Plenum Press*, 205–41, 1998.

Nunez, P. L. and Silberstein, R. B., "On the relationship of synaptic activity to macroscopic measurements: does co-registration of EEG with fMRI make sense?" *Brain Topogr.* **13**: 79– 96, 2000.

O'Regan, J. K. and Noë, A., "A sensorimotor account of vision and visual consciousness." *Behav. Brain. Sci.* **24**:939-1031, 2001.

Owen, A. M. and Coleman, M. R., "Functional neuroimaging of the vegetative state". *Nat. Rev. Neurosci.* **9**:235-43, 2008.

Owen, A. M., Coleman, M. R., Boly, M., Davis, M. H., Laureys, S. and Pickard, J. D. "Detecting awareness in the vegetative state". *Science* **313**:1402, 2006.

Palva, S., Linkenkaer-Hansen, K., Näätänen, R. and Palva, J. M., "Early neural correlates of conscious somatosensory perception". *J. Neurosci.* **25**:5248–58, 2005.

Pantazis, D., Nichols, T. E., Baillet, S. and Leahy, R. M., "Spatiotemporal localization of significant activation in MEG using permutation tests". *Inf. Process. Med. Imaging* **18**: 512–23, 2003.

Pascual-Leone, A., Gates, J. R. and Dhuna, A., "Induction of speech arrest and counting errors with rapid-rate transcranial magnetic stimulation". *Neurology* **41**:697-702, 1991.

Paus, T., Jech, R., Thompson, C. J., Comeau, R., Peters, T. and Evans, A. C., "Transcranial magnetic stimulation during positron emission tomography: a new method for studying connectivity of the human cerebral cortex." *J. Neurosci.* **17**:3178-84, 1997.

Paus, T., Castro-Alamancos, M. and Petrides, M., "Cortico-cortical connectivity of the human mid-dorsolateral frontal cortex and its modulation by repetitive transcranial magnetic stimulation." *Eur. J. Neurosci.* **14**:1405–11, 2001.

Paus, T., “Inferring causality in brain images: A perturbation approach.” *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **360**:1109–14, 2005.

Pigorini, A., Casali, A. G., Casarotto, S., Ferrarelli, F., Baselli, G., Mariotti, M., Massimini, M. and Rosanova, M., “Time-frequency spectral analysis of TMS-evoked EEG oscillations by means of Hilbert-Huang transform”, *J. Neurosci. Methods,* **198**:236-45, 2011.

Phillips, C., Mattout, J., Rugg, M. D., Maquet, P. and Friston, K. J., “An empirical Bayesian solution to the source reconstruction problem in EEG”. *NeuroImage* **24**:997–1011, 2005.

Polonsky, A., Blake, R., Braun, J. and Heeger, D. J., “Neuronal activity in human primary visual cortex correlates with perception during binocular rivalry”. *Nat. Neurosci*. **3**: 1153–9, 2000.

Porta, A., Baselli, G., Liberati, D., Montano, N., Cogliati, C., Gnecchi-Ruscone, T., Malliani, A. and Cerutti, S. “Measuring regularity by means of a corrected conditional entropy in sympathetic outflow.” *Biol. Cybern.* **78**:71-8, 1998.

Putnam, H., “Representation and Reality”, Cambridge: *MIT Press*, 1988.

Putnam, H., “Functionalism: Cognitive Science or Science Fiction?”, in: Johnson, D. M. and Erneling, C. (eds.), “The Future of the Cognitive Revolution”, Oxford: *Oxford Univ. Press*, 1997.

Putnam, H., “The Threefold Cord: Mind, Body and World”, New York: *Columbia Univ. Press*, 1999.

Qiu, F. T. and von der Heydt, R., “Figure and ground in the visual cortex: v2 combines stereoscopic cues with gestalt rules.” *Neuron* **47**:155-66, 2005.

Quiroga, R. Q., Reddy, L., Kreiman, G., Koch, C. and Fried, I., “Invariant visual representation by single neurons in the human brain”. *Nature* **435**:1102-7, 2005.

Rizzolatti, G. and Matelli, M.,“Two different streams form the dorsal visual system: anatomy and functions”. *Exp. Brain Res.* **153**:146–57, 2003.

Rodriguez, E., George, N., Lachaux, J. P., Martinerie, J., Renault, B., and Varela, F. J., “Perception’s shadow: Long-distance synchronization of human brain activity”. *Nature* **397**: 430–3, 1999.

Roelfsema, P. R., Lamme V. A. and Spekreijse, H., “Object-based attention in the primary visual cortex of the macaque monkey”. *Nature* **395**: 376–38, 1998.

Rosanova, M., Casali, A. G., Bellina, V., Resta, F., Mariotti, M. and Massimini, M., “Natural Frequencies of Human Corticothalamic Circuits”. *J. Neurosci.*, **29**:7679-85, 2009.

Rosanova, M., Gosseries, O., Casarotto, S., Boly, M., Casali A. G., Bruno, M., Mariotti, M., Boveroux, P., Tononi, G., Laureys, S. and Massimini, M., “Recovery of cortical

effective connectivity and recovery of consciousness in vegetative patients". *Brain* (in press), 2012.

Rossini, P. M., Narici, L., Martino, G., Pasquarelli, A., Peresson, M., Pizzella, V., Tecchio, F. and Romani, G. L., "Analysis of interhemispheric asymmetries of somatosensory evoked magnetic fields to right and left median nerve stimulation". *Electroencephalogr. Clin. Neurophysiol.* **91**: 476–82, 1994.

Rossini, P. M., Altamura, C., Ferretti, A.,Vernieri, F., Zappasodi, F., Caulo, M., Pizzella, V., Del Gratta, C., Romani, G. L. and Tecchio, F. "Does cerebrovascular disease affect the coupling between neuronal activity and local haemodynamics?" *Brain* **127**:99–110, 2004.

Royal College of Physicians, "Medical aspects of the persistent vegetative state (1). The Multi-Society Task Force on PVS". *N. Engl. J. Med.* **330**: 1499-508, 1994.

Ruohonen, J. and Ilmoniemi, R. J., "Modelling of the stimulating field generation in TMS" *Electroencephalogr. Clin. Neurophysiol. Suppl*. **51**:30-40, 1999.

Russell, B., "Our Knowledge of the External World", New York: *Barnes and Noble*, 2008 [1914].

Sadaghiani, S., Hesselmann, G. and Kleinschmidt, A. "Distributed and antagonistic contributions of ongoing activity fluctuations to auditory stimulus detection. *J. Neurosci.* **29**: 13410–17, 2009.

Salin, P. A. and Bullier, J., "Corticocortical connections in the visual system: structure and function". *Physiol. Rev.* **75**:107-54, 2005.

Sanchez-Vives, M. V. and McCormick, D. A., "Cellular and network mechanisms of rhythmic recurrent activity in neocortex". *Nature Neurosci.* **3**: 1027–34, 2000.

Schiff, N. D., Ribary, U., Moreno, D. R., Beattie, B., Kronberg, E., Blasberg, R., Giacino, J., McCagg, C., Fins, J. J., Llinás, R. and Plum, F., "Residual cerebral activity and behavioural fragments can remain in the persistently vegetative brain." *Brain.* **125**:1210-34, 2002.

Schmolesky, M. T., Wang, Y., Hanes, D. P., Thompson, K. G., Leutgeb, S., Schall, J. D. and Leventhal A. G., "Signal timing across the macaque visual system." *J. Neurophysiol.* **79**: 3272–8, 1998.

Schnakers, C., Perrin, F., Schabus, M., Majerus, S., Ledoux, D., Damas, P., Boly, M., Vanhaudenhuyse, A., Bruno, M. A., Moonen, G. and Laureys, S., "Voluntary brain processing in disorders of consciousness". *Neurology* **71**: 1614-20, 2008.

Schnakers, C., Vanhaudenhuyse, A., Giacino, J., Ventura, M., Boly, M., Majerus, S., Moonen, G. and Laureys, S., "Diagnostic accuracy of the vegetative and minimally conscious state: Clinical consensus versus standardized neurobehavioral assessment". *BMC Neurol.* **9**:35-40, 2009.

Schnakers, C., Perrin, F., Schabus, M., Hustinx, R., Majerus, S., Moonen, G., Boly, M., Vanhaudenhuyse, A., Bruno, M. A. and Laureys, S., "Detecting consciousness in a total locked-in syndrome: an active event-related paradigm." *Neurocase* **15**:271-7, 2009b.

Schroeder, C. E., Mehta, A. D. and Givre, S. J., "A spatiotemporal profile of visual system activation revealed by current source density analysis in the awake macaque." *Cereb. Cortex* **8**:575–92, 1998.

Schurger, A., Cowey, A., and Tallon-Baudry, C., "Induced gamma-band oscillations correlate with awareness in hemianopic patient GY". *Neuropsychologia* **44**: 1796–803, 2006.

Searle, J. R., "Minds and brains without programs". In: Blakemore, C. and Greenfield, S. (eds.), "Mindwaves", Oxford: *Blackwell*, 1987.

Searle, J. R., "Mind: A Brief Introduction", New York: *Oxford Univ. Press*, 2004.

Searle, J. R., Comments on "Are there neural correlates of consciousness?" *Journal of Consciousness Studies*, **1**: 80-82, 2004b.

Seth, A. K., "Causal connectivity of evolved neural networks during behavior." *Network.* **16**:35-54, 2005.

Seth, A. K., Izhikevich, E., Reeke, G. N. and Edelman, G. M., "Theories and measures of consciousness: an extended framework."*Proc. Natl. Acad. Sci. USA* **103**:10799-804, 2006.

Sherman, S. M. and Guillery, R. W., "The role of the thalamus in the flow of information to the cortex." *Philos. Trans. R. Soc. Lond. B Biol. Sci*. **357**:1695-708, 2002.

Siebner, H. R., Willoch, F., Peller, M., Auer, C., Boecker, H., Conrad, B. and Bartenstein, P., "Imaging brain activation induced by long trains of repetitive transcranial magnetic stimulation". *NeuroReport.* **9**:943–948, 1998.

Sillito, A. M., Cudeiro, J. and Jones, H. E., "Always returning: feedback and sensory processing in visual cortex and thalamus." *Trends Neurosci.* **29**:307-16, 2006.

Simpson, T. P., Manara, A. R., Kane, N. M., Barton, R. L., Rowlands, C. A. and Butler, S. R., "Effect of propofol anaesthesia on the event-related potential mismatch negativity and the auditory-evoked potential N1". *Br. J. Anaesth.* **89**:382-8, 2002.

Sinkkonen, J., Tiitinen, H. and Näätänen, R., "Gabor filters: an informative way for analyzing event-related brain activity". *J. Neurosci. Methods* **56**: 99–104, 1995.

Solomonoff, R., "A formal theory of inductive inference". (Parts I and II) *Information and Control*, **7**:1-22 and 224-254, 1964.

Sperry, R., "Science and Moral Priority: Merging Mind, Brain and Human Values", New York: *Praeger*, 1985.

Steriade M., Kitsikis, A. and Oakson, G., "Excitatory-inhibitory processes in parietal association neurons during reticular activation and sleep-waking cycle." *Sleep* **1,** 339–355, 1979.

Steriade, M., Nuñez, A. and Amzica, F., "A novel slow (<1 Hz) oscillation of neocortical neurons in vivo: depolarizing and hyperpolarizing components. *J. Neurosci*. **13**: 3252–65, 1993.

Steriade, M., Amzica, F. and Nuñez, A. "Cholinergic and noradrenergic modulation of the slow (<0.3 Hz) oscillation in neocortical cells". *J. Neurophysiol.* **70**: 1384–1400, 1993b.

Steriade, M., McCormick, D. A. and Sejnowski, T. J., "Thalamocortical oscillations in the sleeping and aroused brain". *Science* **262**: 679–85, 1993c.

Steriade, M., Timofeev, I. and Grenier, F. "Natural waking and sleep states: a view from inside neocortical neurons". *J. Neurophysiol.* **85**:1969–85, 2001.

Tallon-Baudry, C., "Attention and awareness in synchrony." *Trends Cogn. Sci.* **8**:523-5, 2004.

Tallon-Baundry, C., Bertrand, O., Delpuech, C. and Pernier, J., "Stimulus specificity of phase-locked and non-phase-locked 40 Hz visual responses in human". *J. Neurosci.* **16**: 4240–9, 1996.

Tanaka, N., Hämäläinen, M. S., Ahlfors, S. P., Liu, H., Madsen, J. R., Bourgeois, B. F., Lee, J. W., Dworetzky, B. A., Belliveau, J. W. and Stufflebeam, S. M., "Propagation of epileptic spikes reconstructed from spatiotemporal magnetoencephalographic and electroencephalographic source analysis." *NeuroImage* **50**:217-22, 2010.

Tanelian, D. L., Kosek, P., Mody, I. and MacIver, M. B. "The role of the GABAA receptor/chloride channel complex in anesthesia". *Anesthesiology* **78**:757–76, 1993.

Tanné, J., Boussaoud, D., Boyer-Zeller, N. and Rouiller, E. M., "Direct visual pathways for reaching movement in the macaque monkeys". *NeuroReport* **7:** 267–72, 1995.

Thut, G., Ives, J. R., Kampmann, F., Pastor, M. A. and Pascual-Leone, A., "A new device and protocol for combining TMS and online recordings of EEG and evoked potentials." *J. Neurosci. Methods* **141**: 207–17, 2005.

Tikhonov, A. N. and Arsenin, V. Y., "Solutions of ill-posed problems". Washington: *Winston*, 1977.

Timofeev, I., Contreras, D. and Steriade M., "Synaptic responsiveness of cortical and thalamic neurones during various phases of slow sleep oscillation in cat." *J. Physiol.*, **494**: 265-78, 1996.

Timofeev, I., Grenier, F. and Steriade, M., "Disfacilitation and active inhibition in the neocortex during the natural sleep-wake cycle: An intracellular study", *Proc. Natl. Acad. Sci. USA* **98**: 1924–9, 2001.

Tomita, H., Ohbayashi, M., Nakahara, K., Hasegawa, I. and Miyashita, Y., "Top-down signal from prefrontal cortex in executive control of memory retrieval." *Nature* **401**: 699-703, 1999.

Tononi, G. and Edelman, G. M., "Consciousness and complexity". *Scienc*e, **282**:1846-51, 1998.

Tononi, G., "An information integration theory of consciousness". *BMC Neuroscience*, **5**:42-63, 2004.

Tononi, G., "Consciousness as integrated information: A provisional manifesto." *Biol. Bull.* **215**: 216–42, 2008.

Vallar G., "Spatial neglect, Balint-Homes' and Gerstmann's syndrome, and other spatial disorders". *CNS Spectr*. **12**:527-36, 2007.

Van de Velde, M., Van Erp, G. and Cluitmans, P. J. M., "Detection of muscle artefact in the normal human awake EEG". *Electr. Clin. Neurophysiol.* **107**: 149–158, 1998.

Van den Bussche, E., Van den Noortgate, W., Reynvoet, B., "Mechanisms of masked priming: a meta-analysis." *Psychol. Bull*. **135**:452–77, 2009.

Van Der Werf, Y. D., Sadikot, A. F., Strafella, A. P. and Paus, T. "The neural response to transcranial magnetic stimulation of the human motor cortex. II. Thalamocortical contributions." *Exp. Brain Res.* **175**:246-55, 2006.

Van Dort, C. J., Baghdoyan, H. A. and Lydic, R. "Neurochemical modulators of sleep and anesthetic states." *Int. Anesthesiol. Clin.* **46**:75–104, 2008.

Vanderwolf, C. H., "Are neocortical gamma waves related to consciousness?" *Brain Res.* **855**: 217–24, 2000.

Vanhaudenhuyse, A., Schnakers, C., Bredart, S. and Laureys, S., "Assessment of visual pursuit in post-comatose states: use a mirror". *J. Neurol. Neurosurg. Psychiatry* **79**:223, 2008.

Varela, F., Lachaux, J. P., Rodriguez, E. and Martinerie, J., "The brain-web: Phase synchronization and large-scale integration." *Nat. Rev. Neurosci.* **2**: 229–39, 2001.

Velmans, M., "Understanding Consciousness", London: *Routledge*, 2000.

Verbny, Y. I., Merriam, E. B. and Banks, M. I., "Modulation of gamma-aminobutyric acid type A receptor-mediated spontaneous inhibitory postsynaptic currents in auditory cortex by midazolam and isoflurane". *Anesthesiology* **102**:962–9, 2005.

Virtanen, J., Ruohonen, J., Näätänen, R. and Ilmoniemi, R. J., "Instrumentation for the measurement of electric brain responses to transcranial magnetic stimulation." *Med. Biol. Eng. Comput.* **37**: 322–6, 1999.

Walsh, V. and Pascual-Leone, A., "Transcranial Magnetic Stimulation: A Neurochronometrics of Mind", Cambridge: *Bradford MIT Press*, 2003.

Welch, T., “A technique for high-performance data compression”. *IEEE Computer*. **17**:8-19, 1984.

Wexler, B., “Brain and Culture: Neurobiology, Ideology and Social Science”, Cambridge: *MIT Press*, 2006

Williams, M. A., Visser, T. A., Cunnington, R. and Mattingley, J. B., “Attenuation of neural responses in primary visual cortex during the attentional blink”. *J. Neurosci.* **28**: 9890–4, 2008.

Winkowski, D. E. and Knudsen, E. I., “Top-down gain control of the auditory space map by gaze control circuitry in the barn owl”. *Nature* **439**: 336-9, 2006.

Wittgenstein, L. “Philosophical Investigations”, Oxford: *Blackwell* (2nd edition), 1997 [1953].

Wyart, V. and Tallon-Baudry, C., “How ongoing fluctuations in human visual cortex predict perceptual awareness: Baseline shift versus decision bias.” *J. Neurosci.* **29:** 8715-25, 2009.

Younden, W. J., “Index for rating diagnostic tests”. *Cancer* **3**: 32–35, 1950.

Zang, Z., “A fast method to compute surface potentials generated by dipoles within multilayer anisotropic spheres.” *Phys. Med. Biol*. **40**: 335-49, 1995.

Zeman, A., “What in the world is consciousness?” *Prog. Brain Res.* **150**:1-10, 2005.

Ziv, J. and Lempel, A., “An universal algorithm for sequential data-compression”. *IEEE Trans. Inf. Theory* **23**:337-43, 1977.

Acknowledgements

There are many persons who, directly or indirectly, were of fundamental importance for this work.

First, I thank my advisor, Professor Maurizio Mariotti, for his support and confidence; the Università degli Studi di Milano, for financial support of this project; and Professor Giulio Tononi, for receiving me in Madison and being a great inspiration in this investigation.

An empirical work like this involves the dedication and effort of a large number of collaborators. My special gratitude to the colleagues in the group of Milan: Andrea Pigorini, Silvia Casarotto, Matteo Fecchio, Caroline Szymanski, Martino Napolitani and Paola Canali; in Liege: Olivia Gosseries and Mélanie Boly; in Madison, Fabio Ferrarelli and Jeff Johnson; in Porto Alegre, Tiago D'Oliveira.

There are people who were present during this period and transcended the scientific sphere and deserve special thanks. Mario Rosanova and Valentina Bellina, for our scientific collaboration, for teaching me about TMS, and for their support and friendship, making our arrival and stay in Italy much more enjoyable. Dr. Alberto Porta, for his patience in introducing me in the analysis of physiological data. And so many other friends in Milan, who have made our life in Italy such a happy time: Marina, Enrica, Ele, Ricky, Luciano and Erne, Chiara, Vale, Enzo, Marta, Giovanni and Monica (e i bambini), il grande amico Marcello Bianchetti, Ale, Salvo and Marina, and all our Italian friends.

I would like to mention some persons who directly or indirectly enabled this thesis to take this multidisciplinary way. My philosopher friends, Cássio Vinícius Steiner de Souza and, particularly, Fernando Matias Valenzuela Carlucci, who also contributed by reviewing and discussing part of this text. Some of my teachers that are indirectly present, inspiring some passages of this thesis: in Physics, Prof. Vilson Zanchin and Prof. Elcio Abdalla; in

Mathematics, Prof. João Peneireiro; and in Philosophy, Prof. Silvia Altmann and Prof. Paulo Faria.

Thanks to my family in Brazil, to all who have supported me and believed it was possible. In particular, my wife, Karina Rabello Casali, for her patience in hearing about this project along all these years and for having been my faithful companion in so many ways, helping me to achieve my most precious gifts: what we are today together and who we have today in our lives.

In the end, I would like to acknowledge and record my profound gratitude to two persons who were instrumental to this work, as well as to my education as a person and as a scientist. To my friend and teacher, Professor Nicola Montano, I express my warm gratitude for his friendship, for sharing with us his view of the world, of science and life, and for teaching us to cultivate humanism in an increasingly inhumane world. And finally, I give my greatest thanks to Dr. Marcello Massimini, without whom this work would not be possible, not only for having opened the doors of his laboratory and inspired this investigation, teaching me the true nature of scientific work, but also for being a dear friend with whom I could always count on.

Dedico questa tesi a tutti i membri del TMS/hd-EEG lab della Università degli Studi di Milano, in modo speciale a Marcello e Mario, al loro instancabile impegno nel progresso scientifico e, soprattuto, per l'amicizia e la fiducia che hanno riposto in me di cui sarò loro eternamente grato.

Adenauer Girardi Casali