Estimating integrated information with TMS pulses during wakefulness, sleep, and under anesthesia

David Balduzzi

Abstract— **This paper relates a recently proposed measure of information integration to experiments investigating the evoked high-density electroencephalography (EEG) response to transcranial magnetic stimulation (TMS) during wakefulness, early non-rapid eye movement (NREM) sleep and under anesthesia. We show that bistability, arising at the cellular and population level during NREM sleep and under anesthesia, dramatically reduces the brain's ability to integrate information.**

I. INTRODUCTION

Consciousness fades every night during early NREM sleep and under dosages of anesthetic, although average neuronal firing rates differ little from those observed in wakefulness [1], [2]. Since neurons remain active across these different states, it is interesting to ask how physiological changes between sleep and wakefulness affect information processing in the brain, and to investigate whether information-theoretic measures can be developed that distinguish conscious from unconscious states.

Developing robust criteria for determining whether a patient is conscious is important since a small fraction of patients regain or remain conscious during surgery [3]. The simplest method for assessing conscious awareness, long used in clinical settings, is to check for responses to verbal commands. This is unsatisfactory since there are known instances of patients who are clinically unresponsive – due, for example, to the effect of paralyzing agents, loss of motivation to respond, or selective brain lesions – but appear to be conscious [4], [5]. Measures of conscious awareness based on the EEG signal such as the bispectral index have been used in clinical settings [2]. However, the bispectral index can classify responsive patients as unconsciousness [6], suggesting alternative measures are required.

Integrated information, which measures the obstruction to decomposing the information generated by a system of interacting components into independent parts, has been proposed as a measure of conscious awareness [7]. Experiments applying TMS pulses in different physiological states – wakefulness, early NREM sleep and under a dosage of anesthetic – show marked differences in evoked responses measured using EEG [8], [9], [10]. It has been argued that the differences in evoked responses are evidence that the brain generates more integrated information in waking than during sleep [11]. However, relating evoked responses to TMS pulses to the theoretical notion of integrated information [12], [13] is not completely straightforward. This note analyzes a minimal model connecting theory to experiment.

D. Balduzzi is with the MPI for Intelligent Systems, Tübingen, Germany. Email: balduzzi@tuebingen.mpg.de

II. INTEGRATED INFORMATION

From an engineering perspective the brain is an extraordinarily complex input/output device, composed of billions of smaller such devices (neurons and populations of neurons). During wakefulness, it selectively and intelligently responds to a *vast* number of external stimuli.

A clear difference between wakefulness and sleep is that a patient's repertoire of responses to, say, verbal commands decreases sharply. This is an unsuitable criterion for conscious awareness for many reasons – for example, paralyzing agents introduced together with dosages of anesthetic may prevent the patient from responding. It is therefore necessary to consider internal responses to internal perturbations.

This section introduces two perturbation based measures: effective information, which quantifies the selectivity of a devices responses to inputs, and integrated information, which quantifies the obstruction (in bits) to decomposing the information generated by a system of devices into information generated by a collection of independent subsystems. For more detailed treatments see [12], [13].

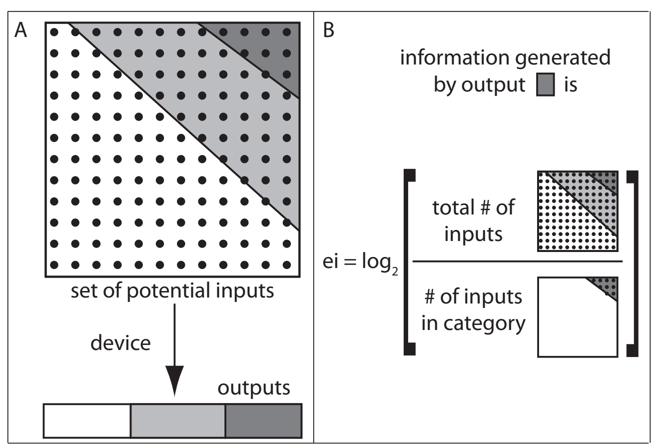

Fig. 1. **Effective information.** (A): A deterministic device can receive 144 inputs (black dots) and produce 3 outputs (grayscale). Each of the inputs is implicitly assigned to a category (shaded areas). (B): The effective information generated by a device choosing a particular output.

A. *Effective information*

A device generates information about its environment by responding selectively to inputs. A thermometer, for example, responds selectively to differences in temperature and is blind to differences in air pressure whereas the converse applies to a barometer. Observing a device's response to a single input tells us little about the differences it detects

– air pressure and temperature may both be high at once. Discovering how a device categorizes its inputs requires exhaustively and actively perturbing it with all of them and tracking its responses.

Let $\mathfrak{m}$ denote a device and $p^{\mathfrak{m}}(x_{out}|x_{in})$ the probability that it chooses output x_{out} given input x_{in}. We perturb the device and track its responses by applying Bayes' rule to the uniform distribution, to compute the *actual repertoire* $p^{\mathfrak{m}}(X_{in}|x_{out})$ of inputs that cause (lead to) x_{out}:

$$p^{\mathfrak{m}}\big(x_{in}|x_{out}\big) = \frac{p^{\mathfrak{m}}(x_{out}|x_{in})}{p(x_{out})} p^{unif}(x_{in}). \qquad (1)$$

For a deterministic device the actual repertoire assigns $p = \frac{1}{M}$ to the M inputs that cause x_{out} and $p = 0$ to the rest. In Fig. 1A, if the device's output is DARK GRAY, then the actual repertoire assigns non-zero probability to the inputs in the dark gray region. The actual repertoire thus captures the category of inputs leading to a given output.

An important feature of a device's output is how selectively it depends on its input. Intuitively, a device that always produces the same output, no matter the input, generates no information about its environment. Similarly, if a device is highly sensitive to its inputs, responding with a different output for any small change, then it generates considerable information about its environment.

Formally, we quantify selectivity via *effective information*

$$ei\Big(\mathfrak{m}, x_{out}\Big) = H\Big[p^{\mathfrak{m}}(X_{in}|x_{out})\Big\|p^{unif}(X_{in})\Big], \qquad (2)$$

the Kullback-Leibler divergence between the actual repertoire and the uniform distribution on the set of potential inputs.

For a deterministic device, Fig 1B, effective information admits a simple description as

$$ei\Big(\mathfrak{m}, x_{out}\Big) = \log_2 \frac{\text{total \# of inputs}}{\text{\# causing } x_{out}}. \qquad (3)$$

B. Integrated information

Any collection of input/output devices is itself an input/output device. In particular, the brain is a massively complex input/output device composed of subdevices which are densely interconnected so that activity within any area can rapidly affect distant areas. Indeed, if activity in one part of the brain were unable to influence other parts, this would introduce a gap in the organism's ability to relate aspects of its environment that competitors could exploit, reducing the organism's likelihood of survival. Integrated information quantifies the weakest link in the information-processing performed within a collection of devices.

The information generated by a system $\mathfrak{n}$ relative to a subsystem $\mathfrak{m}$ is

$$ei\Big(\mathfrak{m} \to \mathfrak{n}, x_{out}\Big) = H\Big[p^{\mathfrak{n}}(X_{in}|x_{out})\Big\|p^{\mathfrak{m}}(X_{in}|x_{out})\Big]. \qquad (4)$$

It quantifies the information the larger device generates over and above the subdevice. Integrated information is the information generated by the system relative to the minimum information partition:

$$\phi(\mathfrak{n}, x_{out}) = ei\big(\mathfrak{p}^{MIP} \to \mathfrak{n}, x_{out}\big). \qquad (5)$$

The minimum information partition is the decomposition of the system into disjoint parts that does the least "information-theoretic damage", see [12].

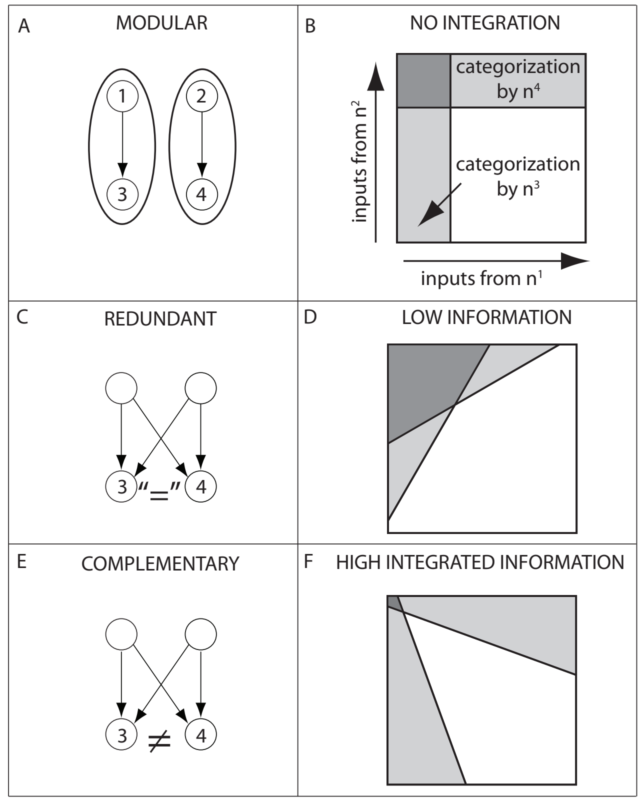

Fig. 2. **Integrated information.** (AB): A modular system of two non-interacting components generates independent (orthogonal) categorizations. (CD): Redundant categorizations overlap, so they generate little more information together than either taken individually. (EF): Complementary categorizations generated by functionally integrated, functionally specialized devices are *more* informative together than the sum of their sub-categorizations.

Fig. 2 contains three examples that we work through to build intuition which will prove useful when interpreting experimental results. Panels AB show a modular system: there are interactions within but not between the two ellipses. The categorizations performed by devices n^3 and n^4 are independent, as can be seen from the structure of the categories they generate: device n^3 is insensitive to perturbations of n^2 and similarly for n^4 and n^1. The system has $\phi = 0$ and decomposes into two completely independent subsystems. The information generated by the system is the sum of the information generated by its two subsystems.

Panels CD show a redundant system: devices n^3 and n^4 categorize inputs similarly implying their respective categories largely overlap. The information generated by the devices together is little more than either taken individually.

Finally, panels EF show two complementary categorizations. The devices treat their inputs quite differently so their categorizations have little overlap. The complementary devices generate far more information together than the sum of their individual contributions.

III. WAKEFULNESS, SLEEP AND ANESTHESIA

We review results distinguishing brain activity during wakefulness from early NREM sleep and under anesthesia. A prominent feature of the brain during early NREM sleep and under anesthesia is that neurons and populations of neurons are bistable which, expanding on [11], we argue dramatically reduces the brain's ability to generate integrated information.

A. Observed differences in brain activity

The brain is never inactive; neurons fire at leisurely rates even during sleep and under the influence of anesthetics. However, the dynamics of neural activity is markedly different in conscious and unconscious states.

1) Spontaneous activity: At a cellular level, during NREM sleep cortical neurons undergo a slow oscillation of about 1 Hz between depolarized up-states, characterized by tonic firing similar to wakefulness, and hyperpolarized down-states, where neurons are silent [1]. Anesthetic has a similar effect on cortical neurons, with higher doses increasing the duration of the hyperpolarized down-states [2].

At a population level, during NREM sleep small initial depolarizing events propagate through cortex by progressively depolarizing, and thus recruiting, populations of neurons into synchronized up-states spreading across large swathes of cortex that are visible as traveling slow waves in EEG recordings [14]. Similar slow waves have been observed under propofol anesthesia [15].

2) Evoked responses: The brain's evoked response to external perturbations in the form of TMS pulses is also quite different during wakefulness compared to early NREM sleep or under anesthesia.

Applying TMS pulses to different cortical areas during wakefulness elicits characteristic, reproducible sequences of evoked potentials that involve many cortical areas over a time period of around 300ms [16]. By contrast, responses during early NREM sleep tend to be either extremely spatiotemporally constrained, when pulses target motor cortex, resulting in a short ($<$ 150 ms) local response, or else, when pulses target mesial parietal regions which form a hub in cortical connectivity, the evoked response resembles spontaneous slow waves, propagating over much of the brain in a large, undifferentiated wave of depolarization.

Neurons are thus bistable during NREM sleep, alternating between depolarized up- and hyperpolarized down-states at the cellular level. The bistability also arises at the population level, appearing as large slow-waves of propagating depolarization clearly visible in EEG recordings. Finally, perturbing the brain during sleep and anesthesia yields either a localized response that fades rapidly, or else stimulates a global event closely resembling spontaneous slow-waves [8], [9], [10]. Responses to TMS pulses during wakefulness are far more complex, resulting in complex, reproducible sequences of evoked responses the precise nature of which depends on the location and nature of the stimulation.

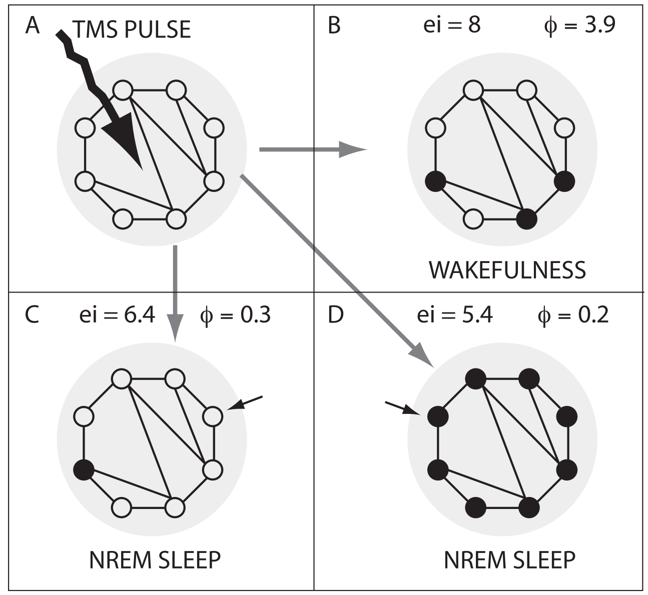

Fig. 3. **A minimal model of evoked EEG responses to TMS.** AND-gates firing for 2 or more spikes. During wakefulness (B), pulses result in variegated responses, differentially recruiting brain regions. During early NREM sleep, pulses either result in (C) a short, localized response that does not propagate or (D) trigger a global wave of activity, propagating across much of the brain, that resembles spontaneous slow waves [8].

B. Implications for integrated information

Integrated information is computed by (i) taking all possible partitions of a system, (ii) exhaustively perturbing the devices in each partition with all possible inputs to compute effective information and (iii) finding the (normalized) minimum. *Exactly* quantifying integrated information thus necessitates an extremely fine-grained analysis that is computationally expensive. By contrast, TMS pulses are coarse, non-physiological perturbations that can be performed only a few times. Relating TMS evoked responses to integrated information therefore requires some care.

1) A minimal model: Fig. 3 shows a system of 8 devices that fire upon receiving two or more spikes. Although the model is extremely simple, it captures the basic feature that excitatory neurons fire more if they receive more spiking inputs. Over larger spatial scales, neural populations contain a mix of mostly excitatory and fewer inhibitory neurons; however, the basic trend towards increasing firing activity in response to increasing firing inputs roughly holds over physiological ranges. Thus, rather than relate the AND-gates to individual neurons, we interpret them as minimal models of populations of neurons whose activity is either above or below a threshold in high-density EEG recordings.

2) Balanced states generate high integrated information: Fig. 3B shows a cartoon version of typical evoked responses during wakefulness: a select subset of brain regions exhibits activity levels above a threshold [8]. Integrated information is high (3.9 bits) and effective information generated by the entire system is the maximum (8 bits).

A system generates high effective information if it is extremely sensitive to change in its inputs. Recall, Fig. 1, high ei means that the category is very small, so almost any modification of the input changes the output. A system generates high integrated information if the effects of perturbations are not local: perturbing one part of the system alters the output of other parts. Compare Fig. 2B, where changing perturbations along the y and x-axes makes no difference to devices n^3 and n^4 respectively, with Fig. 2F, where there are no orthogonal axes respecting the categorizations. The waking brain, like the firing pattern in Fig. 3B, exhibits highly selective, functionally integrated responses to inputs.

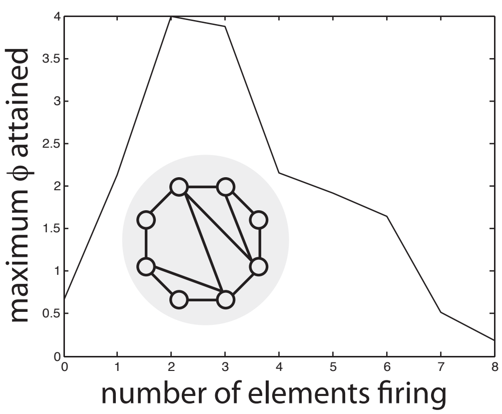

Fig. 4. **Integrated information as a function of population activity.** Nodes as in Fig. 3. Integrated information peaks for intermediary activity; if too many or too few elements are active, ϕ drops markedly.

3) Bistability reduces cortical integration: Figures 3CD show cartoon versions of evoked responses to TMS during early NREM sleep: there is either a short, localized response near the site of the pulses, or else pulses trigger global waves of activity propagating over most of the brain [8]. In both cases effective information is reduced slightly and integrated information drops dramatically. This reflects a general phenomenon: Fig. 4 shows the maximum values of integrated information across all firing rates. Integrated information peaks when two or three gates are active, and decreases substantially at very high or low firing rates.

Fig. 2CD explains the drop in effective information. The AND-gates have broadly similar mechanisms (though not identical since their connectivity differs), so when they all produce the same output there is substantial redundancy in the information they generate.

Although the system is not composed of independent subsystems as in Fig. 2A, the drop in integrated information can nevertheless be explained via Fig. 2B. When all AND-gates are firing the system is relatively insensitive to perturbations of individual gates. For example, whether or not the gate marked by a black arrow in Fig. 3D fired makes no difference to the system's response if all other gates are firing. Somewhat paradoxically, homogeneous global responses *generate little integrated information* since they are insensitive to local differences in activity. Similar considerations apply to extremely low firing rates, Fig. 3C.

IV. CONCLUSION

Directly computing how much integrated information a large system, such as the brain, generates is computationally challenging. We therefore considered a minimal model that captures a basic feature of cortical populations: excitability. The model suggests that neuronal bistability at the population level, resulting in homogeneous brain responses insensitive to local variations in activity, may dramatically reduce the integrated information generated by cortex. Bistable dynamics can decrease effective information by 20 to 30% and integrated information by more than 90%, suggesting that sleep (and also anesthesia) are characterized by a breakdown of effective connectivity between brain regions [8]. Better understanding the relationship between integrated information, conscious awareness, and neuronal bistability – particularly in borderline cases – requires further empirical studies and, crucially, tractable, biologically realistic models.

REFERENCES

[1] M. Steriade, I. Timofeev, and F. Grenier, "Natural waking and sleep states: a view from inside neocortical neurons," *J Neurophysiol*, vol. 85, no. 5, pp. 1969–85, May 2001.

[2] L. Voss and J. Sleigh, "Monitoring consciousness: the current status of EEG-based depth of anaesthesia monitors," *Best Pract Res Clin Anaesthesiol*, vol. 21, no. 3, pp. 313–25, Sep 2007.

[3] P. S. Sebel, T. A. Bowdle, M. M. Ghoneim, I. J. Rampil, R. E. Padilla, T. J. Gan, and K. B. Domino, "The incidence of awareness during anesthesia: a multicenter United States study," *Anesth Analg*, vol. 99, no. 3, pp. 833–9, Sep 2004.

[4] A. M. Owen, M. R. Coleman, M. Boly, M. H. Davis, S. Laureys, and J. D. Pickard, "Detecting awareness in the vegetative state," *Science*, vol. 313, no. 5792, p. 1402, Sep 2006.

[5] J. B. Posner, C. B. Saper, and N. D. Schiff, *Plum and Posner's Diagnosis of Stupor and Coma*. Oxford University Press, 2007.

[6] G. Schneider, A. W. Gelb, B. Schmeller, R. Tschakert, and E. Kochs, "Detection of awareness in surgical patients with EEG-based indices–bispectral index and patient state index," *Br J Anaesth*, vol. 91, no. 3, pp. 329–35, Sep 2003.

[7] G. Tononi, "An information integration theory of consciousness." *BMC Neurosci*, vol. 5, p. 42, 2004.

[8] M. Massimini, F. Ferrarelli, R. Huber, S. Esser, H. Singh, and G. Tononi, "Breakdown of cortical effective connectivity during sleep," *Science*, vol. 309, pp. 2228–2232, 2005.

[9] ——, "Triggering slow waves by transcranial magnetic stimulation," *Proc. Nat. Acad. Sci.*, vol. 104, pp. 8496–8501, 2007.

[10] F. Ferrarelli, M. Massimini, S. Sarasso, A. Casali, B. A. Riedner, G. Angelini, G. Tononi, and R. A. Pearce, "Breakdown in cortical effective connectivity during midazolam-induced loss of consciousness," *Proc Natl Acad Sci U S A*, vol. 107, no. 6, pp. 2681–6, Feb 2010.

[11] M. T. Alkire, A. G. Hudetz, and G. Tononi, "Consciousness and anesthesia," *Science*, vol. 322, no. 5903, pp. 876–80, Nov 2008.

[12] D. Balduzzi and G. Tononi, "Integrated Information in Discrete Dynamical Systems: Motivation and Theoretical Framework." *PLoS Comput Biol*, vol. 4, no. 6, p. e1000091, 2008.

[13] ——, "Qualia: The Geometry of Integrated Information," *PLoS Comput Biol*, vol. 5, no. 8, p. e1000462, 2009.

[14] M. Massimini, R. Huber, F. Ferrarelli, S. Hill, and G. Tononi, "The sleep slow oscillation as a travelling wave," *J. Neur.*, vol. 24, no. 32, 2004.

[15] M. Murphy, M.-A. Bruno, B. A. Riedner, P. Boveroux, Q. Noirhomme, E. C. Landsness, J.-F. Brichant, C. Phillips, M. Massimini, S. Laureys, G. Tononi, and M. Boly, "Propofol anesthesia and sleep: a high-density EEG study," *Sleep*, vol. 34, no. 3, pp. 283–91, 2011.

[16] S. Casarotto, L. J. Romero Lauro, V. Bellina, A. G. Casali, M. Rosanova, A. Pigorini, S. Defendi, M. Mariotti, and M. Massimini, "EEG responses to TMS are sensitive to changes in the perturbation parameters and repeatable over time," *PLoS One*, vol. 5, no. 4, p. e10281, 2010.