Why Does Consciousness Fade in Early Sleep?

GIULIO TONONI AND MARCELLO MASSIMINI

Department of Psychiatry, University of Wisconsin, Madison, Wisconsin, USA

Consciousness fades during deep nonrapid eye movement (NREM) sleep early in the night, yet cortical neurons remain active, keep receiving sensory inputs, and can display patterns of synchronous activity. Why then does consciousness fade? According to the integrated information theory of consciousness, what is critical for consciousness is not firing rates, sensory input, or synchronization *per se*, but rather the ability of a system to integrate information. If consciousness is the capacity to integrate information, then the brain should be able to generate consciousness to the extent that it has a large repertoire of available states (information), yet it cannot be decomposed into a collection of causally independent subsystems (integration). A key prediction stemming from this hypothesis is that such ability should be greatly reduced in deep NREM sleep; the dreamless brain either breaks down into causally independent modules, shrinks its repertoire of possible responses, or both. In this article, we report the results of a series of experiments in which we employed a combination of transcranial magnetic stimulation and high-density electroencephalography (TMS/hd-EEG) to directly test this prediction in humans. Altogether, TMS/hdEEG measurements suggest that the sleeping brain, despite being active and reactive, loses its ability of entering states that are both integrated and differentiated; it either breaks down in causally independent modules, responding to TMS with a short and local activation, or it bursts into an explosive and aspecific response, producing a full-fledged slow wave.

Key words: **NREM sleep; consciousness; TMS; thalamocortical system; slow waves**

Every night, sleep reminds us that consciousness is something that can come and go, grow and shrink, depending on how our brain is functioning. Everyone is familiar with the impression of nothingness that lies in between our falling into and awakening from dreamless sleep. Of course, blank reports upon awakening from sleep are not the rule, and many awakenings, especially from rapid eye movement (REM) sleep, yield dream reports. Dreams can be at times as vivid and intensely conscious as waking experiences. Dream-like consciousness also occurs during various phases of slow-wave sleep, especially at sleep onset and during the last part of the night. Nevertheless, there are always a certain proportion of awakenings that do not yield any dream report, suggesting a marked reduction of consciousness. Such "empty" awakenings typically occur during the deepest stages of non-REM (NREM) sleep (stages 3 and 4), especially during the first half of the night. Understanding why consciousness fades during certain phases of sleep is important not just with respect to brain function during sleep but first and foremost because it can help us in identifying what is really necessary and sufficient for the brain to give rise to conscious experience.

The relationships between sleep and consciousness are indeed interesting and puzzling. It was first thought that the fading of consciousness during sleep was a result of the brain shutting down. However, while metabolism is reduced, the thalamocortical system remains active also during NREM sleep stages 3 and 4, with mean firing rates comparable to those of quiet wakefulness.[9] It was also hypothesized that sensory inputs are blocked during sleep and that they are necessary to sustain conscious experience.[12] However, we now know that, even during deep sleep, sensory signals continue to reach the cerebral cortex where they are processed subconsciously.[5] Gamma activity and synchrony have been viewed as possible correlates of consciousness and they were found to be low in NREM sleep.[2] However, according to some studies, they were equally low in REM sleep, when conscious experience is usually vivid, and they can be high in anesthetized individuals.[11] Moreover, intracellular recordings show that gamma activity persists during NREM sleep,[13] and other studies report that gamma coherence is a local phenomenon that does not change between wakefulness and sleep.[1] Large-scale theta bands may also correlate with

Address for correspondence: Giulio Tononi, Department of Psychiatry, University of Wisconsin, 6001 Research, Park Blvd., Madison, WI 53719. Voice: +1-608-263-6063.
gtononi@wisc.edu

Ann. N.Y. Acad. Sci. 1129: 330–334 (2008). © 2008 New York Academy of Sciences.
doi: 10.1196/annals.1417.024

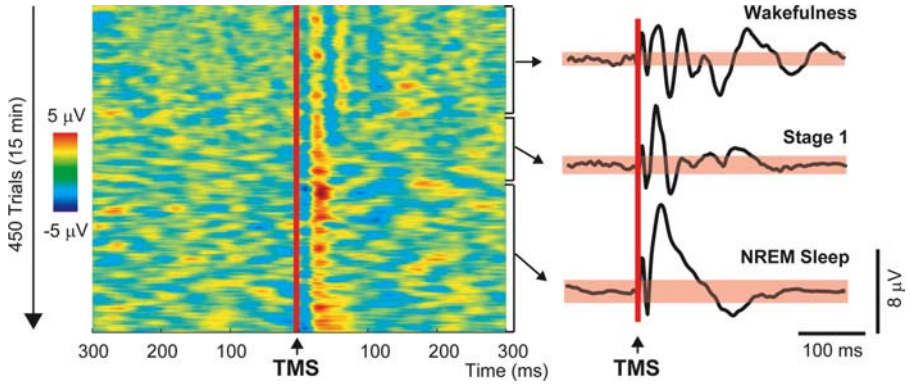

FIGURE 1. Changes in the transcranial magnetic stimulation (TMS)-evoked response during shifts in the state of vigilance. *Left*: Single trials recorded from one channel located under the stimulator while the subject transitioned from wakefulness through stage 1 to nonrapid eye movement (NREM) sleep. Single-trial electroencephalography (EEG) data (filtered 4–100 Hz) are color coded for voltage. *Right*: Averaged TMS-evoked responses (filtered 1–100 Hz) obtained during the three states of vigilance. The *horizontal pink bands* indicate the significance level (three SDs from the mean prestimulus voltage).

conscious perception during wakefulness, but synchrony in these frequency bands actually increases during NREM sleep.[3] Why, then, does consciousness fade?

According to the integrated information theory of consciousness,[10] what is critical for consciousness is not necessarily firing rates, sensory input, specific frequency bands, or synchronization *per se* but rather the amount of integrated information generated by a system. Specifically, the brain substrate of consciousness is thought to be a complex of neural elements, presumably located within the thalamocortical system, which is endowed with the following two properties: i) *information*—the system has a large repertoire of available states so that, when it enters a particular state through causal interactions among its elements, it rules out a large number of alternative states and therefore generates a large amount of information; ii) *integration*—the system cannot be decomposed into a collection of causally independent subsystems so that, when it enters a particular state, it generates information as a whole (i.e., *integrated information*, above and beyond the information generated independently by its parts). An exhaustive measure of complexes and the associated value of integrated information is currently feasible in simple artificial networks,[10] but it is a daunting proposition in a complex biological system, such as the human brain. Nonetheless, the theory makes clear-cut predictions that can be addressed experimentally at least at a gross level. Specifically, the fading of consciousness during early NREM sleep should be associated with either a reduction of integration within the main thalamocortical complex (for example, it could break down into causally independent modules) or a reduction of information (the repertoire of available states might shrink), or both.

The theory also suggests that, to evaluate integrated information, it is not enough to observe activity levels or patterns of temporal correlations among distant brain regions (*functional connectivity*). Instead, the ability to integrate information among distributed cortical regions must be examined from a *causal* perspective; one must employ a perturbational approach (*effective connectivity*) and examine to what extent cortical regions can interact causally (*integration*) and produce differentiated responses (*information*). One should probe effective connectivity by directly stimulating the cortex to avoid possible subcortical filtering and gating, and ideally one should do so in humans, as only with humans do we know that consciousness fades during early NREM sleep. The problem, of course, is that perturbing directly and noninvasively one particular cortical region while recording the overall response of the rest of the human brain is technically challenging.

In a series of recent experiments,[6] we employed a combination of navigated transcranial magnetic stimulation (TMS) and high-density electroencephalography (hd-EEG) to measure noninvasively and with good spatiotemporal resolution the brain response to the direct perturbation of a chosen brain region. Using a 60-channel TMS-compatible EEG amplifier, we recorded TMS-evoked brain responses while subjects, lying with eyes closed on a reclining chair, progressed from wakefulness to deep NREM sleep. Thanks to noise masking and other procedures, subjects were unaware of TMS.

The left panel of FIGURE 1 displays the single-trial responses recorded from one electrode located under

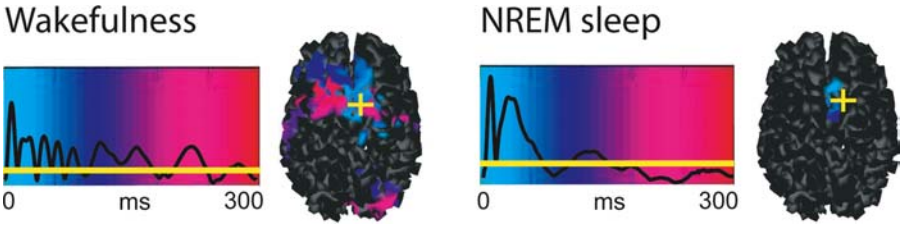

FIGURE 2. Spatiotemporal cortical current maps during wakefulness and NREM sleep. *Black traces* represent the global mean field powers, and the *yellow lines* indicate significance levels. For each significant time sample, maximum current sources were plotted and color coded according to their latency of activation (light blue, 0 ms; red, 300 ms). The yellow cross marks the TMS target on the cortical surface.

the stimulator during a transition from wakefulness through stage 1 to NREM (stages 3 and 4) sleep. In the right panel the averages calculated from the single trials collected in these three vigilance states are shown. During wakefulness, TMS induced a sustained response made of recurrent waves of activity. Specifically, a sequence of time-locked high-frequency (20–35 Hz) oscillations occurred in the first 100 ms and were followed by a few slower (8–12 Hz) components that persisted until 300 ms. As soon as the subjects transitioned into stage 1, the TMS-evoked response grew stronger at early latencies but became shorter in duration from dampening of subsequent fast waves. With the onset of NREM sleep, the brain response to TMS changed markedly. The initial wave doubled in amplitude and became slower. Following this large wave, no further TMS-locked activity could be detected, except for a negative rebound between 80 and 140 ms. Specifically, fast waves, still visible during stage 1, were completely obliterated, and all TMS-evoked activity had ceased by 150 ms.

To reveal the pattern of effective connectivity associated with each cortical perturbation, we performed source modeling and we calculated the spatiotemporal dynamics of the currents induced by TMS. FIGURE 2 shows the patterns of activation evoked by TMS over premotor cortex during wakefulness and NREM. Black traces represent the global mean field powers, and the yellow lines indicate significance levels. For each significant time sample, maximum current sources are plotted and color coded according to their latency of activation (light blue, 0 ms; red, 300 ms). The yellow cross marks the TMS target on the cortical surface. During wakefulness, the initial response to TMS was followed until about 300 ms by multiple waves associated with spatially and temporally differentiated patterns of activation that propagated along long-range ipsilateral and transcallosal connections. During NREM sleep, by contrast, the activity evoked by TMS did not propagate in space and time. Thus, although TMS during sleep elicits an initial response that is even stronger than during wakefulness, this response remains localized, does not propagate to connected brain regions, dissipates rapidly, and is stereotypical regardless of stimulation site.

These findings suggest that the fading of consciousness during certain stages of sleep may be related to a breakdown in cortical effective connectivity. What prevents the emergence of a specific long-range pattern of activation during sleep? During NREM sleep, cortical neurons are depolarized and fire tonically just as in quiet wakefulness, but these depolarized upstates are interrupted by short, hyperpolarized downstates when neurons remain silent.[7] This alternation involves large populations of cortical neurons and is reflected in the EEG as high-amplitude slow oscillations. The transition from upstate to downstate appears to be a result of depolarization-dependent potassium currents that increase with the amount of prior activation.[7] Perhaps, because of this bistability of cortical networks during NREM sleep,[4] any local activation, whether occurring spontaneously or induced by TMS, will eventually trigger a local downstate that prevents further propagation of activity.

The role of bistability in altering information processing during sleep has been corroborated in subsequent experiments in which TMS, applied with the appropriate parameters, triggered full-fledged slow oscillations.[14] In FIGURE 3, few cycles of the slow oscillation evoked by TMS in humans are compared to the spontaneous slow oscillations recorded in cats (field potential and membrane potential). Each TMS pulse, delivered with high intensity over the median parietal cortex, triggered a negative wave associated with global downstate in vast regions of the brain and was followed by a positive component crowned by spindles. Spatially, the TMS-evoked slow oscillation was associated with a broad and aspecific response: cortical

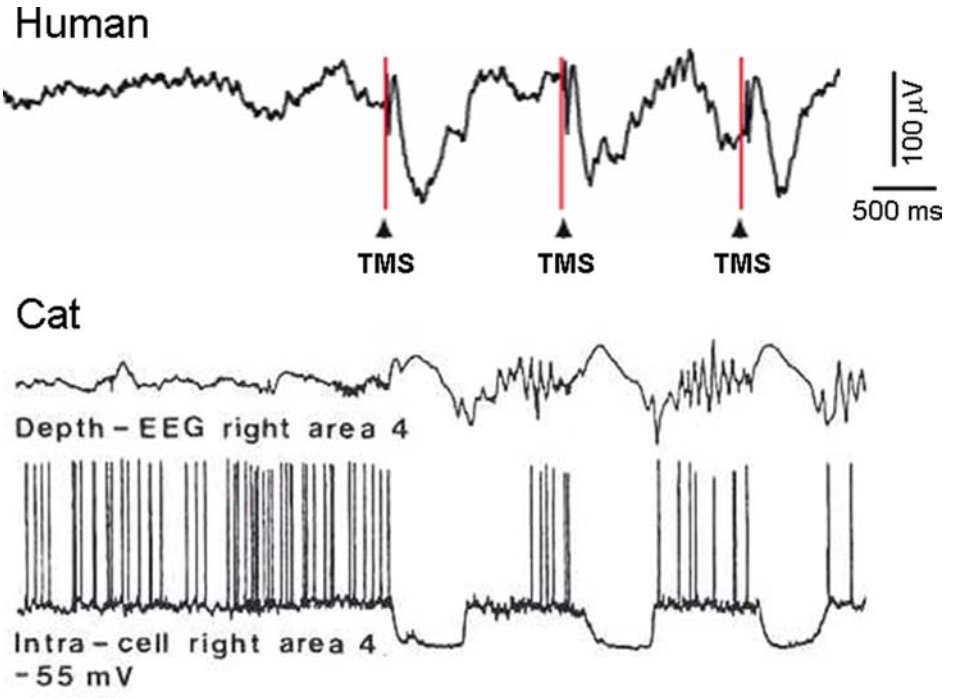

FIGURE 3. TMS triggering of full-fledged sleep slow oscillations. The upper trace is from one channel located under the stimulator. Each TMS pulse, delivered at high intensity (70% of maximum output), triggers a typical slow oscillation—a negative wave (more negative than −80 mV) followed by a positive rebound crowned by spindles. As highlighted by the comparison with recordings in cats (modified from Ref. 8), the surface negative wave recorded in human (depth positive in cats) corresponds to the intracellular downstate of the slow oscillation.

currents spread like an oil spot from the stimulated site to the rest of the brain. A possible explanation for this effect has to do with the strategic location of the median parietal cortex, which is heavily and directly connected to a large number of other cortical areas, as well as to the intralaminar thalamic nuclei, which are themselves diffusely connected to vast regions of the cortex. In this way, TMS might produce a strong diffuse depolarization that would be followed by a global downstate.

Altogether, these TMS–EEG measurements suggest that the sleeping brain, despite being active and reactive, loses its ability of entering states that are both integrated and differentiated; it either breaks down in causally independent modules or it bursts into an explosive and aspecific response. In no case, during NREM sleep, did TMS result in a long-range differentiated pattern of activation. The TMS–EEG perturbational approach also suggests that intrinsic bistability in thalamocortical networks may represent not only the key mechanism responsible for the occurrence of the spontaneous slow oscillations of sleep but also the reason why information integration is impaired in early NREM sleep. In this perspective, the inescapable occurrence of a downstate after a second or so in an upstate, be it a local or global phenomenon, suggests that as a result of bistability, the availability of a large repertoire of neural activation patterns is much reduced. Moreover, this bistability and consequent restriction of neural repertoire may be present but covert, even during periods of stable ongoing EEG with no slow oscillations. To uncover the network's bistability, it may be necessary to perturb it directly. And indeed, in our experiments, TMS applied to the cortex during NREM sleep invariably resulted in a stereotypical downstate, even during periods characterized by low-amplitude EEG.

Acknowledgments

The work was supported by the National Institute of Health Director Pioneer Award to G.T. This paper is based on previous publications by the authors.

Conflicts of Interest

The authors declare no conflicts of interest.

References

1. Bullock, T.H., M.C. McClune, J.Z. Achimowicz, *et al*. 1995. Temporal fluctuations in coherence of brain waves. Proc. Natl. Acad. Sci. USA **92:** 11568–11572.
2. Cantero, J.L., M. Atienza, J.R. Madsen & R. Stickgold. 2005. Precise timing accounts for posttraining sleep-dependent enhancements of the auditory mismatch negativity. Neuroimage **26:** 628–634.
3. Duckrow, R.B. & H.P. Zaveri. 2005. Coherence of the electroencephalogram during the first sleep cycle. Clin. Neurophysiol. **116:** 1088–1095.
4. Hill, S. & G. Tononi. 2005. Modeling sleep and wakefulness in the thalamocortical system. J. Neurophysiol. **93:** 1671–1698.
5. Kakigi, R., D. Naka, T. Okusa, *et al*. 2003. Sensory perception during sleep in humans: A magnetoencephalograhic study. Sleep Med. **4:** 493–507.
6. Massimini, M., F. Ferrarelli, R. Huber, *et al*. 2005. Breakdown of cortical effective connectivity during sleep. Science **309:** 2228–2232.
7. Sanchez-Vives, M.V. & D.A. McCormick. 2000. Cellular and network mechanisms of rhythmic recurrent activity in neocortex. Nat. Neurosci. **3:** 1027–1034.
8. Steriade, M. 2000. Corticothalamic resonance, states of vigilance and mentation. Neuroscience **101:** 243–276.
9. Steriade, M., I. Timofeev & F. Grenier. 2001. Natural waking and sleep states: A view from inside neocortical neurons. J. Neurophysiol. **85:** 1969–1985.
10. Tononi, G. 2004. An information integration theory of consciousness. BMC Neurosci. **5:** 42.
11. Vanderwolf, C.H. 2000. Are neocortical gamma waves related to consciousness? Brain Res. **855:** 217–224.
12. Coenen, A.M. 1998. Neuronal phenomena associated with vigilance and consciousness: from cellular mechanisms to electroencephalographic patterns. Conscious Cogn. **7:** 42–53.
13. Steriade, M., D. Contreras, F. Amzica, I. Timofeev. 1996. Synchronization of fast (30–40 Hz) spontaneous oscillations in intrathalamic and thalamocortical networks. J. Neurosci. **15:** 2788–2808.
14. Massimini, M., F. Ferrarelli, S.K. Esser, *et al.* 2007. Triggering sleep slow waves by transcranial magnetic stimulation. Proc. Natl. Acc. Sci. USA **104:** 8496–8501.