A Complete Biological Theory of Sleep

Ari Rappoport

The Hebrew University of Jerusalem, Israel
ari.rappoport@mail.huji.ac.il www.cs.huji.ac.il/~arir
April 28, 2019

Sleep is still considered a mystery, despite intense scientific investigation. Here we present the first complete biological theory of sleep. The role of sleep is to restore the optimal homeostatic state, which is essential for tissue performance and health. Non-rapid eye movement sleep (NREMS) restores cortical and most other brain neurons, via relaxed global activity managed by thalamocortical circuits. The role of REM sleep is to restore acetylcholine (ACh) neurons, which support focused responses and hence cannot participate in global oscillations. Sleep enhances learning and memory via state restoration and ACh-affected paths. NREMS induces a lack of consciousness because global synchronous activity prevents focused responses, which are essential for consciousness. Dreams result from focused neural firing during sharp-wave ripples and REMS, and have a sense of reality because they involve the same neurons representing focused perceptual responses during wake. Anesthetics utilize a variety of mechanisms that prevent focused responses.

Keywords: Sleep, anesthesia, rapid eye movement (REM) sleep, NREM sleep, sharp-wave ripples (SWRs), dreaming, consciousness.

Introduction

Sleep and dreaming have fascinated humanity for thousands of years, and are the subject of intense scientific study [Walker, 2017]. While advances have been made in understanding many aspects of sleep [Scammell et al., 2017], it is at present still largely a mystery [Joiner, 2016, Siegel, 2011].

A theory of sleep must answer at least the following questions. First, sleep is evidently crucial for survival. It disconnects the animal from its environment in a way that can put it in danger, yet it exists in some form in virtually all animals [Miyazaki et al., 2017]. What is the important role of sleep that justifies this? Second, sleep has two main stages, rapid
eye movement sleep (REMS) and non-REM sleep (NREMS), the latter exhibiting several EEG phenomena, including spindles, slow wave activity, and sharp wave ripples [Brown et al., 2012]. What are their roles and mechanisms? Third, sleep is known to be important for memory and learning [Rasch and Born, 2013]. What is the exact nature of the connection? Fourth, anesthesia can induce a sleep-like state. How do anesthetics work, and in what ways are they similar to and different from sleep? Finally, sleep involves unique consciousness states, both a sense of reality during dreams and a complete lack of consciousness otherwise. How can this be explained? Specifically, what are the roles and mechanisms of dreaming?

Many theories of the role of sleep have been proposed, including memory consolidation, synaptic homeostasis, energy replenishment and saving, and intracellular and extracellular clearance [Krueger et al., 2016, Hill et al., ]. However, none of these theories addresses all or even most of the questions above. Here, we present a comprehensive theory of sleep, having three main contributions. First, it is the first theory that addresses all of these questions, presenting a coherent story of what sleep is, how it relates to brain function, and how it is implemented. Second, it is the first mechanistic theory of dreaming. Third, it presents a novel account of the plasticity changes occurring in sleep, thereby reconciling ongoing debates.

### Theory Overview

During wake, the organism generates responses to sensory inputs conveying internal and external needs and threats. The brain manages the muscle contraction aspect of this process. To support the response process, the brain utilizes **arousal** agents, the main ones being orexin, norepinephrine, histamine, dopamine, serotonin and **acetylcholine (ACh)**, each having a specific response process role. The role of ACh is to support **focused responses**, generated via competition between response candidates after which neuronal winners are enhanced and executed while losers are suppressed. Executed focused responses are **learned** (automated) via long-term **plasticity** processes. The term ‘memory’ refers to a specific type of learned content.

Body tissues depend on an **optimal homeostatic state** for health and performance. Wake activity induces deviations from the optimal state, and these can damage cells and impair tissue function. Reactive species resulting from metabolism, excessive intracellular calcium ions (Ca\(^{2+}\)), and extracellular debris can all be toxic. Tissues and individual cells possess state restoration mechanisms, but these are not fully effective as long as activity continues. Sleep is a brain-managed state that enforces the cessation of activity and promotes efficient restoration of the optimal state. As such, it enhances all aspects of brain function, including both responses and plasticity, and is essential for tissue health. The role of NREMS is to efficiently restore large neuronal populations via sustained relaxed firing. All arousal neurons except ACh neurons are restored via tonic firing during wake. The role of REMS is to restore ACh neurons via firing.

**NREM sleep** is induced by circadian circuits and activity- and immune-induced sleep pressure. First, circadian areas such as the **suprachiasmatic nucleus** can suppress arousal neurons. Second, energy metabolism and the immune system (via prostaglandin D2) induce the accumulation of an agent conveying sleep pressure, **adenosine (ADN)**, in the extracellular space.
Via the A1 receptor, ADN activates potassium channels to suppress cells, including arousal neurons. Via the A2A receptor and A1R-mediated disinhibition, ADN excites neuronal groups in the preoptic area that mutually inhibit arousal neurons. The suppression of arousal neurons and the decrease of sensory inputs due to reduced activity relieve the drive to form focused responses. Together with an increased neural excitability due to degradation of the optimal state, this enables the emergence of fast corticothalamic oscillations (spindles) followed by slow wave activity (SWA), slow oscillations of large neuronal populations mediated by the synchronizing drive capacity of the thalamus. Neural firing during SWA restores the intracellular space, and prompts restoration of the extracellular space via glia cells, blood vessels, the cerebrospinal fluid (CSF) and the glymphatic system.

SWA alleviates the block on firing induced by state degradation. At some point, neurons can escape thalamic control to start firing individually. The first neurons to do so are those most strongly affected by prior wake activity, hippocampal and (later) ACh neurons. The former induce a sequential activation shown as EEG sharp wave ripples (SWRs) that spreads to cortical areas. The firing of ACh neurons yields REM sleep. REMS terminates SWA, because ACh opposes coordinated activation of large neuronal populations due to its support of focused responses. The plasticity effect of ACh is to enhance focused paths and suppress losing competitors and noise. The enhanced paths are those involved in focused responses during wake, marked by intracellular cation accumulation. This explains why sleep is important for memory (learning).

The REMS state is assisted by pontine sublaterodorsal nucleus (SLD) neurons, which counter the wake effects of ACh by suppressing movement, and by brainstem and hypothalamic neurons that suppress other arousal neurons.

The intense neural activity induced by ACh degrades the optimal state, yielding additional cycles of the same process until the suppression of arousal neurons cannot be maintained. In the healthy brain, this occurs when brain neurons have been sufficiently restored.

The cortical and hippocampal activity during SWRs and REMS is focused, and comprises neurons that represent real percepts and events during wake. Those are the neurons that represent the content of dreams. Hence, dreams involve the general themes present during wake. However, they generally do not recapitulate wake events, since these depend on the coordinated activity of neurons in many brain areas and on precise temporal sequencing, while dreams are driven by deviations from the optimal state. Once started, neural activations during dreams can take a life of their own. For this reason, dreams can include illogical or impossible events.

Our sense of reality during wake depends upon the activation of neurons that represent specific perceptual content, and consciousness involves neurons that support the reporting of perceptual content. Both content perception and reporting are focused responses, hence are neutralized during SWA and allowed during REMS and SWR (and possibly spindles). This explains why these sleep stages involve a lack of consciousness and a sense of reality, respectively.

Most anesthetics utilize the same brain mechanisms used by sleep, and induce a sleep-like state. Some anesthetics utilize different mechanisms, and induce a trance-like state.

The novel aspects of the proposed theory all involve the notion of neuronal paths imple-
menting focused execution via cations. There is currently no proposed term for this notion. We propose the term quax, and refer to the present theory as the Quax Theory of Sleep (QTS).

Wake activity

Sleep cannot be understood without a model of wake activity. During wake, the brain receives internal and external sensory inputs, and finds and executes movement responses. Four cell classes are involved in this process. Excitatory neurons represent content, inhibitory neurons support response coordination, agent-releasing neurons manage the different process stages, and glia cells provide energy, protection and plasticity.

Response content. Excitatory (mostly pyramidal) neurons represent perceptual, action, and valence (approach/avoid) content. In general, neural representations are specific, in that every particular neuron is activated in a relatively small number of events. Sensory inputs of different types reach highly specific areas of the brain, and muscle contraction responses are also generated by specific neurons. Although response generation may initially involve widespread neural excitation, response execution is eventually done by a relatively small number of neurons. Thus, a major aspect of brain function is to discover and enhance a neural subset that connects inputs to outputs, and to suppress other neurons (Figure 1). We say that the selected neurons form a focused response path. The selected neurons are distributed, not linearly connected as may be implied by ‘path’. We use this term to emphasize the fact that the selected subset is relatively small and focused.

Focused responses are generated via a response process (figure 2). Sensory inputs can trigger innate responses. If this does not happen, they flow into cortex, which is the main area of adaptive representations. In cortex, sensory inputs generate widespread alerting bottom-up
flow, mainly in layers 2-4, which is answered by top-down (TD) flow in layers 2/3/5. The TD flow activates a (possibly large) number of response candidates, which respond to features of the current input learned due to the execution of past responses. If an urgent response is required, a coarse response is rapidly executed. Otherwise, the response candidates compete, and the winners constitute a focused response (‘center’) that is executed by neurons in layers 5/6. Layer 5b contains the cortical neurons projecting to motor centers that actually drive responses. The losing candidates are suppressed. The focused response is sustained by the basal ganglia and thalamus, which can provide strong synchronizing drive throughout cortex.

This description integrates several prominent existing theories that rely on a large body of evidence, including biased competition [Desimone and Duncan, 1995], the action perception cycle [Fuster, 2015], predictive coding [Adams et al., 2013], and the reverse hierarchy theory [Ahissar and Hochstein, 2004]. To cite just one important piece of evidence, consider EEG event-related potential (ERP) components [Kotchoubey, 2006]. The activation of candidates in the superficial layers is what generates negative ERPs such as the N170 associated with face perception and the N400 associated with semantics (especially in language). Surprises yield increased negative ERP amplitude, because they induce the activation of more candidates. Competition and its resolution, which involve neurons in the deep layers, are associated with positive ERP components. For example, the language P600 ERP is larger when the ending of a sentence requires reinterpretation.

**Coordination.** The selection of the winning subset and the execution of responses by it are assisted by inhibitory interneurons (IINs) releasing GABA or glycine. These agents open ion
channels that let Cl⁻ enter the cell and polarize it, suppressing its firing. Although this effect is technically inhibitory, it is misleading to say that the function of IINs is inhibition. When an IIN fires just before its targets are about to fire, it indeed inhibits them. However, when an IIN fires just after its targets have fired, it only resets their charging states. If these targets accumulate charge at approximately the same rate, the effect is to synchronize their firing, making it more efficient and powerful. Thus, IINs implement a ‘join-or-stop’ selection mechanism whereby some neurons (winners) are enhanced while others (losers) are suppressed. Moreover, after the winners are selected, IINs continue synchronizing winner firing to support efficient sustained execution. Thus, a more accurate term for the role of local GABAergic and glycinergic neurons is response coordination.

Mechanistically, coordination is done as follows. When flow reaches an area, it activates some excitatory neurons. These excite IINs located around them, which may be viewed as sending coordination requests. IINs target the whole local neighborhood indiscriminately, establishing a join-or-stop positive feedback loop. This mechanism is fundamental in creating the well-known phenomenon of surround suppression [Chettih and Harvey, 2019].

There are several IIN classes, each supporting a different aspect of this process [Tremblay et al., 2016]. Soma-targeting IINs expressing parvalbumin (PV) are fast spiking due to high expression of calcium-permeable (CP) AMPA receptors [Wang and Gao, 2010], inducing a rapid powerful effect over the whole cell. PV binds Ca²⁺ with high affinity, allowing rapid repeated excitation at gamma frequency (30-80Hz) [Henzl, 2013]. Cholecystokinin (CCK)-expressing IINs also target cell somas and express calcium binding proteins (calbindin and PV). Thus, these IINs are suitable for sustaining intense synchronous activity. Non-fast spiking IINs expressing somatostatin (SOM) target apical dendrites of pyramidal neurons, to suppress specific losing synapses during path selection. IINs expressing vasoactive intestinal polypeptide (VIP) suppress the SOMs to disinhibit (enhance) the winning centers [Letzkus et al., 2015, Karnani et al., 2016]. Neurogliaform cells (NGFCs) complement VIPs by inducing broad suppression of surrounding neurons [Jiang et al., 2013]. Finally, Chandelier IINs targeting the axon initial segment induce rapid response suppression.

Management agents. The brain’s response process is assisted by various agents, each supporting a specific process stage. The main arousal agents and their main roles are as follows. Orexin (hypocretin) conveys general needs, e.g., for food, and excite the dopamine system [Chen et al., 2015]. Vasopressin indicates the need for water and various social needs [Stoop, 2012]. Histamine conveys the need to activate the immune system and to remove pathogens by itch [Haas et al., 2008]. Norepinephrine (NE) is triggered by surprises to induce global alert and energy recruitment (increased heart rate, vasoconstriction, glycogenolysis). Dopamine (DA) is triggered by strong need cues to support goal-setting [Salamone and Correa, 2012] and immediate coarse responses [Sharples et al., 2014]. Serotonin (SER) restrains rapid responses to promote slower, patient, flexible planning [Dalley and Roiser, 2012].

All of these agents promote the formation of responses, so oppose the coordinated activity
of large neuronal populations. Hence they need to be suppressed to allow sleep.

**ACh.** ACh has a unique role in sleep and is hence discussed at greater depth here. Its role in
the response process is to support the selection and execution of focused responses. ACh acts
via two receptor families, nicotinic (nAChR) and muscarinic (mAChR) receptors. nAChRs are
ionotropic and excitatory. They have two common forms, alpha7, which is CP, and alpha4beta2,
which is not permeable to calcium. The latter have a higher agonist affinity and desensitize
rapidly, supporting efficient automated execution. mAChRs are G protein-coupled receptors
with two sub-types, M1/3/5, which bind to Gq and suppress potassium channels to promote
sustained activation, and M2/4, which bind to Gi/o to suppress the cell via potassium channels
and reduced energy production.

Brain ACh is released by nuclei groups in the basal forebrain (BF, midbrain) and the brain-
stem [Mesulam, 2013]. The former send widespread innervation throughout cortex and hip-
pocampus, while the latter innervate subcortical areas, including the thalamus, basal ganglia,
agent nuclei and motor areas. The BF group is heavily innervated by cortex, most notably pre-
frontal cortex (PFC). ACh is released in a precise spatiotemporal manner [Muñoz and Rudy,
2014], targeting both excitatory neurons and IINs. It increases glutamate release by excita-
tory neurons via alpha7 nAChRs [Yakel, 2014], affects SOM IINs [Fanselow et al., 2008], PV
IINs via presynaptic alpha4beta2 nAChRs [Albuquerque et al., 2009], VIP IINs via nAChRs
[Fu et al., 2014], and NGFC IINs via mAChRs [Brombas et al., 2014]. Thus, ACh promotes
the emergence and sustenance of focused responses by its strong effect on IINs and through a
positive feedback loop with executing content neurons.

The strong relationship between ACh and focused execution is beyond doubt. It is the agent
used by the motor unit to induce muscle contraction, and by the parasympathetic autonomic
nervous system (ANS) to end the alert mode and support execution. It is known to promote
attended execution and its ensuing plasticity [Sarter et al., 2016].

ACh has a strong effect on plasticity. In fact, its effect should be stronger than that of other
agents, because the goal of plasticity is to automate the responses that were executed (this is
the overarching principle of Hebbian learning), and ACh is the agent that directly promotes the
winning, executing subset. ACh promotes intracellular calcium, which is the main trigger of
plasticity processes. ACh promotes direct Ca\(^{2+}\) entry into its target cells via alpha7 nAChRs, re-
cruitment of Ca\(^{2+}\) from intracellular stores via M1/3/5 mAChRs, and indirect entry via M1/3 and
all nAChRs by activating voltage-gated calcium channels (VGCCs) [Muñoz and Rudy, 2014].
PV IINs, which are excited by alpha4beta2 nAChRs, enjoy calcium entry via highly expressed
CP-AMPARs. ACh promotes the generation and release of nitric oxide (NO), which produces
vasodilation to increase blood flow and also suppresses the Na\(^{+}/K^{+}\)-ATPase pump in choroid
plexus cells, countering pump-dependent calcium extrusion [Ellis et al., 2000]. Indeed, alpha7
nAChR is essential for ontogenic development [Fernandes et al., 2014], and neurogenesis, an
extreme form of brain plasticity, is enhanced by M1 mAChR [Van Kampen and Eckman, 2010].

We note that not all focused responses depend on ACh. Acute responses do, but automated
(learned) responses utilize accurate contextual predictions, which minimize ACh-dependent
competition. Acute responses involve Ca\(^{2+}\), so utilize NMDA glutamate receptors, CP-AMPARs, and alpha7 nAChR, while automated responses utilize CI receptors.

In summary, brain operation consists of focused responses, selected and maintained by IINs with the essential support of ACh in case of acute (non-automated) responses.

**Homeostasis and its Restoration**

**The optimal state.** Body tissues have an optimal homeostatic state, defined by extracellular and intracellular (cytoplasmic and intra-organelle) levels of various agents (ions, proteins, molecular debris, reactive oxygen species (ROS), neurotransmitters, neuromodulators etc.). The optimal state is important for both tissue health and performance.

Deviation from the optimal state can damage cells. Metabolism yields by-products that can damage cells via oxidative stress, and cells can accumulate agents (e.g., Ca\(^{2+}\)) that are toxic under many conditions [Connolly and Prehn, 2015]. Activity triggers genomic processes that can take hours and are easily impaired by activity-induced agents. Remnants of normal and pathogenic agents can aggregate and impair processing, and activate the immune system to induce inflammation.

Deviation from the optimal state is not only harmful in the long-term, but also impairs short-term performance. Brain function depends on neural responsiveness, which requires specific levels of ion concentration gradients and protein expression.

**Ca\(^{2+}\) clearance is essential.** During activity, Ca\(^{2+}\) enters brain cells in large amounts, due to the large concentration gradient across the cell’s membrane and the presence of calcium-permeable channels. In addition, calcium is recruited from internal stores into the cytoplasm. Due to the internal gradients induced by proton pumping, Ca\(^{2+}\) enters mitochondria, where it signals the need for increased ATP production [Gleichmann and Mattson, 2011]. Ca\(^{2+}\) can rapidly accumulate in mitochondria, because the sodium-calcium exchanger that extrudes Ca\(^{2+}\) is slower than its uniporter. Excessive Ca\(^{2+}\) enhances the generation of ROS, and triggers Ca\(^{2+}\) extrusion via the mitochondria permeability transition pore, which can induce cytochrome c release and apoptosis [Brookes et al., 2004]. Excessive accumulation of Ca\(^{2+}\) in internal endoplasmic reticulum (ER) stores can also be harmful.

Wake activity induces a rapid rise in brain extracellular K\(^{+}\) and a decrease in extracellular Ca\(^{2+}\) and extracellular volume [Ding et al., 2016], supporting a wake-induced neural accumulation of Ca\(^{2+}\). Direct evidence for such accumulation was given in [Liu et al., 2016], where prolonged wake induced a sustained increase in intracellular Ca\(^{2+}\) and synaptic plasticity markers in sleep-inducing Drosophila R2 neurons.

**Restoration mechanisms.** Cells possess intricate mechanisms that restore the optimal state, including antioxidant enzymes, ion pumps, DNA repair processes, degradation and clearance pathways (e.g., autophagy), and anti-inflammatory cytokines. Tissues possess restoration mechanisms beyond those of individual cells. The lymphatic system cleans the interstitial fluid of debris. In the brain, astrocytes maintain the perisynaptic environment, and microglia and as-
tocytes monitor inflammatory agents. Agents can diffuse into the extrasynaptic space, where they are generally out of the reach of astrocytes and can escape microglia but can affect cells via extrasynaptic receptors. The extrasynaptic space is continuously cleared via the recently discovered glymphatic system, which utilizes the CSF and blood vessels [Jessen et al., 2015].

However, this continuous restoration is insufficient, because as long as activity continues, these mechanisms cannot be fully utilized and are thus not fully effective. This holds for basically all tissue. We and others (see below) propose that the solution that evolution has found to this problem is to enforce a period of minimal activity, sleep. Indeed, sleep deprivation has been shown to be harmful to the cardiovascular, immune and endocrine systems [Liu and Chen, 2019, Orzeł-Gryglewska, 2010], to muscle [Jówko et al., 2018], and to the liver, lung, and small intestine [Everson et al., 2014].

**State deviation, the brain, and ACh.** The brain may need restoration more than other tissue. Its function depends on large ion concentration gradients, yielding high energy requirements and ROS production, yet its antioxidant mechanisms are relatively small compared to other tissue [Cobley et al., 2018]. This issue pertains most strongly to ACh-supported paths, since these support prolonged focused execution that does not allow parallel effective maintenance. ACh neurons themselves are especially vulnerable. The number of ACh neurons is much smaller than the number of neurons in supported paths, so the load on each ACh neuron is high. Indeed, ACh neurons are impaired in almost all neuron degeneration disorders, as well as in normal aging [Schliebs and Arendt, 2011].

**State deviation increases excitability.** In the optimal state, neurons are kept polarized, with pumps maintaining the large ion concentration gradients across membranes. As the optimal state degrades, neural excitability increases. The wake-induced decrease in brain extracellular Mg$^{2+}$, Ca$^{2+}$ and H$^+$ [Ding et al., 2016] implies an increased permeability of NMDA receptors to Ca$^{2+}$, increased accumulation of cations in cells, and therefore increased cell depolarization and excitability.

**Neuronal firing facilitates restoration.** The exit of Ca$^{2+}$ from the ER is triggered by increased cytoplasmic Ca$^{2+}$ levels. Ca$^{2+}$ clearance from mitochondria is primarily done by the sodium-calcium exchange pump (NCX). Since neural firing involves the cytoplasmic entry of Ca$^{2+}$ via VGCCs and of sodium (Na$^+$) via VGSCs and HCN channels (Ih), firing quickly improves the conditions enabling the clearance of Ca$^{2+}$ from internal stores. Firing also recruits astrocytes to clear the synaptic cleft and to promote neurovascular communication. Unlike high-frequency firing, relaxed firing does not have high energy requirements and does not induce Ca$^{2+}$ accumulation in mitochondria. Thus, relaxed firing supports restoration more effectively than pump operation alone.

Note that the increased neural excitability during wake constitutes a restoration mechanism, by promoting spontaneous neural firing from time to time. Some neurons are indeed restored by tonic firing (e.g., DA neurons, see below). Apparently, this mechanism is not sufficient for restoring neurons supporting sustained focused responses or the extracellular space.
Additional evidence. There is substantial evidence that sleep deprivation increases oxidative stress. [Jówko et al., 2018] showed that deprivation with 36 hour survival training impairs antioxidant enzymes, induces muscle damage, and increases lipid peroxidation. Sleep deprivation increased oxidative stress in PV IIINs [Harkness et al., 2018]. A single night of sleep deprivation in humans significantly reduced antioxidants, ATP, and morning cortisol [Trivedi et al., 2017]. Short sleep drosophila suffer from increased oxidative stress and shorter survival [Hill et al., 2018]. Some experiments found no oxidative damage following sleep deprivation [Gopalakrishnan et al., 2004], but the weight of the evidence points otherwise. A recent review concluded that the molecular correlates of sleep deprivation are related to ER stress, protein misfolding, and oxidative stress [Naidoo, 2019].

There is also evidence showing a direct benefit of sleep on homeostatic measures. [Zada et al., 2019] showed that sleep increases chromosome dynamics, reducing accumulating neuronal DNA damage. A comprehensive study found that the genes upregulated during sleep are mostly related to anabolic processes (cholesterol and protein synthesis, lipid transport, vesicle pools), antioxidation, and energy regulation [Mackiewicz et al., 2007].

NREM Sleep

NREMS mechanisms are a relatively well-understood aspect of sleep [Neske, 2016, Brown et al., 2012], so here we only summarize the important points. Sleep is induced by suppressing arousal neurons, and involves reduced sensory input. NE, OX and histamine amplify bottom-up sensory flow conveying needs, so their suppression is essential for sleep. DA and SER support top-down responses to sensory inputs, and can be allowed to continue their tonic firing. Indeed, DA neurons do not change their mean firing rate during sleep (and can fire in bursts during REM) [Torterolo et al., 2016], and the effect of SER on sleep is equivocal, generally promoting wake and suppressing REMS but also enhancing sleep in some situations [Monti, 2011].

When bottom-up flow is suppressed, the brain is left without triggers that drive the formation of focused responses. Thalamic neurons possess an inherent oscillation capacity due to ion channels and the interaction between thalamic excitatory neurons and inhibitory thalamic reticular nucleus (TRN) neurons [Huguenard and McCormick, 2007], and exert a strong drive over all cortical areas [Clascá et al., 2012].

At the initial stage of sleep (N2), local corticothalamic interactions induce fast (10-15Hz) EEG oscillations called spindles. An additional EEG phenomenon at this stage is the K-complex, sharp negative EEG deflections commonly having an external sensory (usually auditory) origin. As explained above, IIINs tend to synchronize the neurons around them that receive dominant flow while suppressing others. When the flow conveyed by the thalamus is generated by inherent capacity (due to reduced sensory inputs), its cortical targets receive approximately the same flow, and the system gets into a global synchronous mode. This is slow-wave sleep (SWS). SWS is the exact opposite of wake and REMS, in that it involves the synchronous activity of large neuronal populations instead of small focused paths.

Adenosine (ADN), a by-product of energy production, is the main agent conveying sleep
pressure [Porkka-Heiskanen and Kalinchuk, 2011]. As more energy is produced, higher amounts of ADN are released into the extracellular space. ADN acts in two ways. First, the A1 receptor (A1R) suppresses ATP production via Gi/o and activates potassium channels (KCs), most notably G protein-coupled inwardly-rectifying KCs (GIRK channels) [Rainnie et al., 1994]. KCs are normally activated after neural firing as part of the return to the cell’s optimal state, and are important for SWA. ACh inhibits the M-current, a potassium current, via M1 mAChRs, so its suppression allows KC activation. In turn, activation of KCs enables burst firing due to T-type VGCCs causing a low-threshold spike, on top of which a series of high frequency action potentials can ride [Cheong and Shin, 2013]. Indeed, SWA is characterized by cycles comprised of a relatively long down state followed by a bursty up state [Brown et al., 2012]. SWA spindles are bursts and are probably induced by the TRN [Lüthi, 2014].

Importantly, A1R has a particularly strong effect in the basal forebrain, where ACh and sleep-inducing preoptic neurons are located. Following 3 hours of sleep deprivation, A1R-induced NFkB nuclear translocation occurs almost only in the ACh neurons of the basal forebrain [Ramesh et al., 2007]. A meta-analysis and an experiment showed that ADN accumulation is specifically high in the BF [Leenaars et al., 2018].

A second effect of ADN is that the A2A receptor (A2AR), which is excitatory, excites sleep-active neurons in selected areas. Important areas include the basal forebrain ventral lateral preoptic area (VLPO) and the median preoptic area (MnPN) [Gallopin et al., 2005], which mutually inhibit ACh, SER, and NE neurons, and inhibit histamine neurons. VLPO neurons are also disinhibited by the effect of A1R on GABAergic neurons.

The projection neurons of the striatum nucleus of the basal ganglia (BG) oscillate between up and down states during sleep [Mahon et al., 2006]. The BG contains two paths, direct and indirect, and A2A excites those belonging to the indirect path, ensuring their restoration during sleep (direct path neurons require DA, so cannot get included).

Lateral hypothalamic melanin concentrating hormone (MCH) neurons can assist sleep by suppressing orexin, histamine, SER, NE and wake-only ACh neurons [Monti et al., 2018]. MCH neurons mutually suppress NE (alerts, stress) during wake to promote exploration and eating [González et al., 2016], so their firing during sleep (which is stronger during REMS and may be specific to it) may be a side-effect of NE suppression and not due to a specific sleep promoting role.

According with slow-wave sleep being driven by a factor that gradually diminishes, it is most prominent in the first sleep cycle, decreasing in duration with further cycles.

In addition to restoring neurons, the relaxed SWA promotes the restoration of the extracellular space by facilitating blood flow via astrocytes, which improves clearing via the glymphatic system [Mestre et al., 2018].

Additional evidence. The effect of ADN on sleep pressure is supported by a large amount of evidence [Huang et al., 2014]. For example, mice lacking the primary metabolizing enzyme of ADN show enhanced sleep drive, and A1 knockout significantly decreases sleep drive [Bjorness et al., 2016]. ADN inhibits all ACh neurons [Rainnie et al., 1994].
With respect to KCs, mutations in the Shaker KC and its promoters Hyperkinetic and sleepless reduce sleep and increase excitability in drosophila [Allada et al., 2017, Koh et al., 2008]. The sleepless gene antagonizes nAChRs, and lynx1, a mammalian homolog, can partially restore normal sleep to mutants [Wu et al., 2014]. [Yoshida et al., 2018] generated 14 types of leak KC knockout mice, and in all of them sleep duration decreased. Moreover, they implemented a computational model showing that leak KCs can regulate SW A-like EEG along with Ca\textsuperscript{2+} channels. The same group also showed that impaired Ca\textsuperscript{2+}-dependent KCs and VGCCs decrease sleep duration, while impaired plasma membrane Ca\textsuperscript{2+} ATPase increases sleep duration [Tatsuki et al., 2016]. Finally, KCs are activated by most anesthetics [Li et al., 2018].

An elegant demonstration of the role of specific wake activity in inducing sleep is given by so-called local sleep [Huber et al., 2006]. When an arm was immobilized during the day, during subsequent sleep, SWA over the corresponding cortical area was markedly reduced. Complementing this result, a learning task involving a specific brain region produced a local increase in sleep SWA.

The important role of slow wave activity in clearing the extracellular space was shown in [Ju et al., 2017], where a specific disruption of SWA increased CSF amyloid beta.

**REM Sleep**

A recent review described the adaptive role of REMS as still being a complete mystery [Siegel, 2011]. In QTS, the main role of REMS is to restore ACh neurons. As explained above, ACh neurons are especially sensitive to deviations from the optimal state. However, they cannot be restored by SWA, because their inclusion would introduce a strong drive for competition and focused paths, preventing the global synchronous oscillations of SWA. For the same reason, they cannot be restored by tonic firing during wake. REM sleep follows SWS and restores ACh neurons through firing.

Mechanistically, REMS is triggered by SWS-induced restoration of the extracellular space, which clears it from ADN. Since ADN most strongly affects the basal forebrain, its clearance disinhibits BF ACh neurons (and also brainstem ACh neurons). ACh neurons accumulate cations during wake, and are also possibly excited by cortical projections during SWS. As a result, when they are disinhibited, they start firing. This firing drives the activation of focused sequences, which terminates SWS and induces dreaming (see below).

The hallmark of REMS are rapid eye movements. These occur because ACh supports attended execution, which involves saccades to attended spatial locations. The claustrum, which takes part in attending to spatial locations [Goll et al., 2015], is activated during REMS and excites cortex [Luppi et al., 2017a], contributing to ACh-induced activations.

ACh-driven focused responses can induce strong muscle contraction. To prevent this, glutamatergic REMS SLD neurons innervate glycinergic neurons in the raphe magnus (RMg), the ventral and alpha gigantocellular nuclei (GiV and GiA) and the lateral paragigantocellular nucleus (LPGi), which suppress spinal, facial and trigeminal motor neurons [Luppi et al., 2017b]. This induces muscle atonia. The circuits activating these SLD neurons during REMS are not
fully clear. They are suppressed during wake, possibly via tonic GABAergic inputs arriving from the brainstem ventrolateral periaqueductal grey (vlPAG) and the dorsal (or deep) mesencephalic reticular nucleus (dMeRN).

The suppression of arousal neurons must continue during REMS, otherwise they would induce wake. A population of GABAergic/galaninergic neurons in the extended VLPO suppresses the DRN (SER), LC (NE) and vlPAG/dMeRN [Scammell et al., 2017]. Neurons in the dorsal paragigantocellular reticular (DPGi) and lateral paragigantocellular (LPGi) nuclei also suppress LC NE neurons during REMS. MCH neurons, which suppress arousal neurons, are particularly active during REMS [Vetrivelan et al., 2016].

**SWRs.** Sharp-wave ripples (reviewed in [Buzsáki, 2015]) occur during both wake and sleep. In sleep, they occur mostly during NREM, but are described here because they bear a greater resemblance to REMS activity than to SWA. They occur at the later stages of SWS, when the brain has been sufficiently restored to allow some neurons to escape from the synchronous oscillations imposed by the thalamus.

As in REMS, the first neurons to fire individually are those that have accumulated more cations during previous wake activity. Some evidence for this is provided by the fact that the source of SWRs are the principal neurons in the CA3 field of the hippocampus. These neurons are unique in the brain in that during acute responses they are excited not only by Ca$^{2+}$ but also by Zn$^{2+}$ [Ji et al., 2019]. Zinc is a strong promoter of proliferation and plasticity, so zinc-innervated neurons are expected to be among the first to fire. Their activation ripples through the hippocampus CA1 field and into cortical areas.

SWRs have been shown to be capable of recapitulating wake sequences. However, ‘replay’ is just one pattern among many exhibited by SWRs [Gupta et al., 2010]$^3$. Both data points support our account, because firing order during sleep is determined by cation accumulation and does not necessarily have any content-related logic.

**Consciousness**

Possibly the most perplexing features of sleep are the complete lack of consciousness during SWA and the sense of reality felt during dreaming. The scientific examination of consciousness is a deep and contentious field of study [Koch, 2004]. Because consciousness is a subjective feeling, it is currently accessible to science only via reporting (by words or actions) [Dehaene, 2014]. Fortunately, to explain the relationship between consciousness and sleep, we only need an aspect of consciousness that we hope would be relatively uncontested, the fact that it involves focused responses.

The key observation is that consciousness involves **specific content**. We can be conscious of the external world, in which case we perceive specific objects and events. We can be conscious of internal brain activations (silent speech, music, static or dynamic imagery) without external sensory inputs, and these are specific as well (silent speech involves a sequence of well-defined

$^3$At least some wake SWRs are probably involved in planning future actions [Roumis and Frank, 2015].
words in a particular order, imagery has well-defined visual components, etc.)\(^4\). Finally, it can be argued that consciousness of being alive is the core consciousness feeling and that it is different from thinking. The feeling of being alive depends on attending to interoceptive sensory inputs, which arrive at highly specific cortical locations [Craig, 2014]. This is supported by Cotard’s syndrome, in which patients deny being alive and which is associated with right frontal (especially insular) lesions [Sahoo and Josephs, 2017]. Note that the right insula receives the strongest heart inputs.

Since the content of consciousness is specific, it involves the activation of a small and focused set of neurons representing this content. Regardless of whether reporting is only a scientific way to investigate consciousness or is its crucial defining aspect, reporting clearly involves focused responses. In other words, both the content of consciousness and its reporting require a focused response.

Using this observation, the lack of consciousness during SWA is easily explained by the fact that the synchronous activity of large neuronal populations prevents focused responses. Note that a similar situation occurs during epileptic seizures.

**Evidence.** ACh, the agent supporting focused responses, has been described as the ‘neurotransmitter correlate of consciousness’, based on its effects on attention, disorders, and anesthesia [Perry et al., 1999]. A recent review concluded that the minimal set of neurons required for a conscious percepct includes a temporo-parieto-occipital area associated with perceptual experiences, and a frontal area associated with thought-like (i.e., reporting) experiences [Koch et al., 2016]. During dreaming, posterior areas exhibit reduced low frequency EEG activity, with high frequency activity being associated with specific content [Koch et al., 2016]. A meta-analysis showed that dreaming involves increased activation in posterior parietal cortex, mPFC (heavily innervated by ACh, the focus agent), and hippocampus, and reduced activation in dIPFC, inferior frontal gyrus, and OFC (all DA-dominated areas) [Fox et al., 2013]. In addition, lack of consciousness (albeit via propofol, which blocks NE) was characterized by disintegration of inter-areal synchrony, which is a property of focused responses [Lewis et al., 2012].

A serious theory of consciousness posits that it requires a global broadcast of information in cortex [Dehaene, 2014], which may seem to oppose focused paths. However, the broadcast of this theory exactly corresponds to the QTS activation of candidates competing for inclusion in a focused response. Thus, the two accounts are complementary\(^5\).

**Dreaming.** In REMS, the main sleep stage associated with dreaming, focused paths are strongly encouraged by ACh. Focused paths can emerge during NREM without ACh to yield SWRs, since firing excitatory neurons send coordination requests to local IINs. When the synchronizing power of the thalamus diminishes, this can induce competition with focused winners. As during wake and REMS, competition resolution quickly induces forward activations, explain-

---

\(^4\)This argument holds regardless of how exactly imagery and thought are implemented in the brain.

\(^5\)Note that we do not equate focused responses with consciousness. For example, highly automated and/or unattended responses are focused but involve very little competition, and it is indeed common not to remember (be aware of) having performed such actions.
ing the association between SWRs at the final NREM stage and dreaming [Siclari and Tononi, 2016].

The sense of consciousness felt during dreaming arises because the neurons activated during dreaming include content neurons (e.g., at the hippocampus and temporal cortex) that are activated in focused paths during wake.

A fake sense of reality (i.e., one contradicting external reality) is not unique to sleep. It can occur during wake, and can be easily induced in healthy people by various hallucinogens [Nichols, 2016]. Since a sense of reality requires focused paths, disrupting the signaling of all agents that support these, including ACh, SER and DA, can induce hallucinations. Conversely, the nAChR agonist nicotine generally sharpens attention (reality).

Thus, the question unique to sleep (the lack or presence of consciousness of reality) is readily answered: sleep suppresses or activates the same neurons inducing the sense of reality during wake. The questions pertaining to both wake and sleep (which neurons are these, what mechanisms are used) is not answered here. In other words, what may be mysterious is not the fact that sleep can obliterate or induce a sense of reality, but the sense of reality itself.

Dreams rarely recapitulate the events experienced during wake [Schredl, 2018]. This may seem at odds with our theory, because the drive of neuronal firing during dreams is cation accumulated during wake events. However, real events are represented by a highly distributed network of neurons firing at precise temporal sequences, and the various neurons involved do not accumulate cations at the same rate. Thus, the content of dreams is strongly inspired by the components of wake events, but does not recapitulate their order or logic.

Dream stories are commonly not logical or are plainly impossible (e.g., flying), and develop without a sense of control [Siclari and Tononi, 2016]. This is what we would expect, since the brain can clearly conjure up illogical imagery as long as it can represent its individual components. Dreams are also commonly highly emotional, which can be explained by the strong activity of emotion circuits during wake, reflected by a high activation of the amygdala during REMS [Schredl, 2018]. Lucid dreamers are people who are aware that they are dreaming while they do so, and some can exert some control over dream content. Awareness of one’s situation is a report, and the focused responses that allow reports are in principle possible during sleep (similarly to dream content). The capacity of exerting control over dream content points to a higher than normal activation of control areas (e.g., dlPFC) during dreaming, possibly due to increased innervation by ACh or DA. DA is the agent supporting goal-setting, which is a major component of the sense of control. DA neurons continue firing during sleep, and can fire bursts during REMS [Torterolo et al., 2016]. Being aware of dreaming may be based on short term memory of getting ready to sleep and the plausibility of the dream scenario.

Dreaming is a completely different brain mode from visual imagery during wake. Dreaming involves the activation of the same neurons activated when experiencing reality, while imagery most likely does not. Indeed, images in dreams are described as more vivid than in imagery [Siclari and Tononi, 2016].

Dreams can be useful for finding creative solutions to problems that occupy the dreamer during wake [Walker, 2017], because they search the solution space in a way that is hard to
invoke during wake, when the brain is more restrained by set goals (preconceptions, concrete plans).

In summary, dreams are a side-effect of optimal state restoration, mainly of hippocampal and ACh neurons. They have no particular role, but can be useful for creativity. Their content reflects the dreamer’s wake activities and thoughts, and is experienced as real because the neurons involved in wake perception and dreaming are the same neurons.

Memory and Plasticity

Sleep is clearly important for learning [Rasch and Born, 2013]. Sleep deprivation impairs, and sleep enhances, learning. Several theories view memory consolidation or plasticity as the main role of sleep [Walker, 2017, Krueger et al., 2016].

In QTS, improved learning is a natural result of restoration of the optimal homeostatic state, which improves all brain processes. Specifically, restoration of the intracellular space is essential for the genomic processes triggered by wake activity. In addition, the significant ACh flow into cortex during REMS serves as a powerful plasticity catalyst.

The synaptic homeostasis hypothesis (SHY) [Tononi and Cirelli, 2014] argues that wake activity strengthens synapses and that this process cannot go on indefinitely due to increased energy requirements, decreased performance (signal to noise), and saturated learning. In SHY, the role of sleep is to normalize plasticity by synapse weakening. Indeed, SHY cites evidence that synapse sizes and numbers decrease only if flies are allowed to sleep, and that sleep/wakefulness yield net spine loss/gain, respectively. Conversely, much evidence counters SHY. For example, sleep involves both synaptic strengthening and weakening [Frank and Cantera, 2014], most cortical synapses are upregulated during SWS [Timofeev and Chauvette, 2017], and there is LTP during SWS [Chauvette et al., 2012]. Behaviorally, sleep induces both memory consolidation and forgetting [Feld and Born, 2017].

The notion of focused responses, emphasized by the present theory, bridges these two accounts. Recall that focused responses result from competition between a possibly large number of candidates. Thus, the number of losing synapses is much larger than the number of winning ones. Moreover, acute responses induce intracellular Ca$^{2+}$, which stimulates the formation of new spines, and many of those are associated with losers. Post-execution plasticity weakens or removes the losers and enhances the winners. Our account predicts that synapse weakening should statistically dominate, but that some synapses would be strengthened or retained.

This is exactly what the evidence shows. In [Diering et al., 2017], sleep induced a GluA1 (an AMPAR subunit) increase in some spines, but a reduction in a larger number of spines. In [De Vivo et al., 2017], the axon-dendritic spine contact area decreased by about 20% during sleep, but this held only for the smaller 80% of the spines. In other words, stronger synapses were not downscaled. According with the strengthening of some synapses, NREMS involved the removal of CP-AMPARs, which make synapses more labile [Lanté et al., 2011]. Sleep augmented the connectivity and firing rates of stimulus-selective neurons, and decreased those of less selective (i.e., competition losing) neurons [Clawson et al., 2018]. In [Li et al.,
REMS pruned new spines created after motor learning, facilitated subsequently generated spines, and strengthened a fraction of new spines. Reactivation of task-specific neurons during NREMS was involved in forming new synapses after learning [Yang et al., 2014]. Firing rate responses to stimuli are potentiated during sleep, not wake, and are blocked by sleep deprivation [Durkin and Aton, 2016]. Finally, LTP yields increased spine size, but also an overall decreased spine density, via local competition between stimulated spines [Stein and Zito, 2019].

Mechanistically, whether a synapse or spine is weakened or enhanced can be based on Ca$^{2+}$ amplitude, duration, or influx location [Evans and Blackwell, 2015], and on the effect of neurotrophic agents. BDNF enhances synapses via the TrkB receptor, while proBDNF weakens them via the p75 receptor [Park and Poo, 2013].

Another piece of evidence that supports our account is that waking rest also contributes to memory consolidation [Wamsley, 2019]. Waking rest allows restoration, which has a positive effect on plasticity.

It should be mentioned that glucocorticoids (cortisol in humans) are released during sleep, and so is testosterone in males. Both of these agents support plasticity.

**Anesthesia**

Anesthetics are exogenous drugs and discussing them at depth is out of the scope of this paper. Nonetheless, their general mechanisms of action, and their effects on consciousness, are readily explained by QTS.

Like sleep, anesthesia is considered to be mysterious [Franks, 2008]. It is perplexing that a relatively large and structurally diverse group of molecules can induce a similar sleep-like state. QTS explains anesthesia by noting that each family of anesthetics makes use of a mechanism used by sleep as well. Most general anesthetics target potassium channels (KCs) and other ion channels [Li et al., 2018, Franks, 2008]. For example, isoflurane and sevoflurane activate KCs, while propofol suppresses VGSCs. Propofol also suppresses NE, the main arousal agent. Similarly, most anesthetics potentiate GABA$A$-induced currents and directly activate GABA$A$ receptors at higher concentrations. These are the same mechanisms that promote sleep. KCs are used by ADN, and wide activation of GABA$A$Rs triggers their suppressive action and does not allow focused responses. Note that GABAB$R$s activate KCs via Gi/o.

Because anesthetics are commonly applied when normal sleep pressure is not strong enough, they must be continuously provided in order to sustain anesthesia.

Ketamine, xenon and nitrous oxide are commonly viewed as special cases. They (especially the first two) inhibit NMDARs, possibly by competing with the co-agonist glycine. Ketamine blocks calcium-permeable (CP) glutamate receptors, but promotes calcium-impermeable AMPARs [Abdallah et al., 2018]. Thus, it suppresses acute responses, but can increase spontaneous neural firing. This counters normal consciousness and thinking, but the effect is not sleep-like. Indeed, ketamine yields a trance-like state, while other anesthetics induce a sleep-like state. Nitrous oxide and xenon inhibit alpha7nAChR [Suzuki et al., 2003], the CP ACh receptor that most strongly promotes focused responses.
Discussion

This paper presented the first complete biological theory of sleep. Our main thesis is that sleep serves to restore the optimal homeostatic state, which degrades during wake. All major aspects of sleep can be explained by understanding the central role of focused responses in wake activity and of the cation accumulation that they induce. ADN and circadian circuits apply sleep pressure by suppressing arousal neurons, and this lets the thalamus and cortex self-organize in an oscillatory SWA pattern that restores the optimal state. In general, REMS restores ACh neurons, while NREMS restores other neurons and the extracellular space. Optimal state restoration facilitates all brain processes, including memory consolidation and overall plasticity. Slow global oscillations do not allow consciousness during SWS, while the sense of reality associated with dreaming results from the activation of neurons implementing focused responses during both wake and dreams.

Existing theories. QTS agrees with the benefits of sleep proposed by previous theories cited above. Sleep restores intracellular energy stores, whose usage is triggered by NE, which is suppressed during sleep. Sleep uses the glymphatic system for clearing the extracellular space. By restoring the optimal state, it counters wake-induced performance degradation, improves cell health, reduces oxidative stress, and improves plasticity. The specific aspects that are novel here are the emphasis on focused responses and Ca\textsuperscript{2+} accumulation and their mechanistic description, the role of ACh, the account of consciousness and dreaming, and the realization that these are sufficient to present a complete story.

[Walker, 2017] proposes that sleeps promotes creativity and downregulates the emotional aspects of events, thereby supporting calm planning and reducing anxiety and false alerts. The former indeed holds (see above), because the focused responses activated during sleep are likely to not occur during wake. The latter requires clarification. Sleep enhances the responses executed during wake. Normally, these are focused responses that suppress their emotional triggers via the parasympathetic ANS (promoted by ACh). In these cases, sleep would indeed enhance TD control over learned emotional reactions. However, sometimes the nature of alerts is such that they trigger innate responses or coarse adaptive responses. In this case, plasticity processes, and thus sleep as well, would enhance these responses, yielding emotional hyper-sensitivity (as occurs e.g. in PTSD). Recollection of emotional events involves a far smaller activation of the sympathetic ANS than when the event occurs, but this is not due to sleep but to the lack of surprising alerts and emotion-inducing sensory inputs.

Other tissue. In this paper we emphasized brain processes, because the sleep state is managed by the brain. Nonetheless, sleep is important for restoring other body tissues as well. For example, sleep promotes immune memory consolidation and restores immune cells [Westermann et al., 2015] (see above for additional tissues).

Note that the function of the immune system is close to that of sleep. Sleep restores degradation resulting from normal activity, while the immune system generally addresses degradation resulting from abnormal situations (direct threats, damage to blood vessels, etc.). However, the
dividing line is diffuse, because the presence of excessive levels of normal agents, even if resulting from normal activity, can activate the immune system. In this case, sleep and immunity cooperate towards the same goal.

The gut may use unique mechanisms to trigger sleep after eating. However, the fact that animals get tired after eating can also be explained by noting that OX neurons, which provide a major part of arousal input, are suppressed by glucose. When OX neurons are silent, DA and NE drive diminishes, and MCH drive increases, providing further suppression of NE.

**Restoration of DA, NE, SER, and OX neurons**. The arousal neurons releasing OX, DA, NE and SER serve acute responses during wake and are thus in need of restoration. However, NE and OX neurons are suppressed during all sleep stages, and the data with respect to DA and SER is equivocal, which seems to contradict our main thesis. This issue is resolved by noting that all of these neurons are tonically active during wake, and at least DA neurons are tonically active during sleep as well. We propose that their restoration is done by this relaxed tonic firing. ACh neurons are different because they are under heavier load, are more important for plasticity, and if they fired tonically they would risk the erroneous formation of focused responses. Most of these neurons express suppressive autoreceptors (alpha2-NER, D2R, SER1a, M2/4 mAChR), providing another protection mechanism.

A note regarding dopamine and plasticity. Some cortical areas strongly innervated by DA (e.g., dlPFC) are normally not heavily innervated by ACh. These areas undergo plasticity as the rest of the brain (possibly more, since development of frontal areas is disproportional during adolescence, a critical period in brain plasticity). DA neurons fire bursts during REMS (see above), which may indicate that NREM is not sufficient for plasticity. Indeed, although specific REMS suppression does not always induce a strong behavioral effect [Siegel, 2011], several experiments showed that it induces neuronal apoptosis and mitochondria damage [Somarajan et al., 2016]. Thus, REMS may indeed be crucial for memory.

**Restoration vs. plasticity**. The role assigned to sleep here is state restoration, while plasticity (memory consolidation) was presented as a side-effect. Note that we could have said that the root cause of sleep is that the body is an adaptive (learning) system based on Ca\(^{2+}\). Since Ca\(^{2+}\) signaling is metabolically demanding, increases oxidative stress, and can trigger apoptotic processes, it has also promoted the evolution of a special state to undo these undesired side-effects. Thus, memory and restoration are tightly interconnected.

**Sleep disorders**. Sleep involves many mechanisms, and an impairment in each of them can impair it. For example, impairments in the OX or neurovascular systems can interfere with the normal sleep-wake cycle. Discussing sleep disorders at any reasonable depth is out of the scope of this paper.

**Future work**. Most aspects of QTS are supported by solid experimental evidence. The existence and nature of focused responses, their promotion by ACh, the degradation of homeostasis induced by sleep, the damages associated with it, and the positive effect of sleep on homeostasis and plasticity are beyond doubt. Many of the circuits and agents involved in transition between
sleep stages have been mapped.

Future research should provide more evidence for the following aspects. First, direct evidence that what triggers neuronal firing during SWRs and REMS is cation accumulation is needed. Second, it is important to discover the exact location of the representations giving rise to the sense of reality, during both wake and sleep. In addition to pure science, such a discovery could be of enormous practical use for treating patients in comatose situations. Third, the mechanistic circuits promoting NREMS and REMS have not been fully mapped. For example, it is not clear what activates REMS SLD neurons. Fourth, ADN may not be the only agent conveying sleep pressure; other agents (e.g., ROS) could complement its action. Fifth, zinc is probably important for sleep [Cherasse and Uraden, 2017], but its role is not fully clear. Finally, if QTS is correct, it should have implications for the treatment of sleep disorders.

Conclusion

In summary, we have presented the first complete theory of sleep. By focusing on a small number of key notions, the important of which being the notion of focused responses, it allows the integration of existing knowledge to provide a concrete biological account of all major aspects of sleep. Sleep is not a mystery but a biological state with a clear biological logic and well-defined mechanisms.

Author’s Note

The theory of sleep presented here is based on a theory I developed that explains brain function [Rappoport, 2018b, Rappoport, 2018a]. To make reading the paper easier, it is completely self-contained. Readers who are interested in the brain’s response process, the implementation of focused responses and consciousness, and the roles of various agents (e.g., dopamine, serotonin, MCH, orexin) should consult these two references and my homepage.

References


