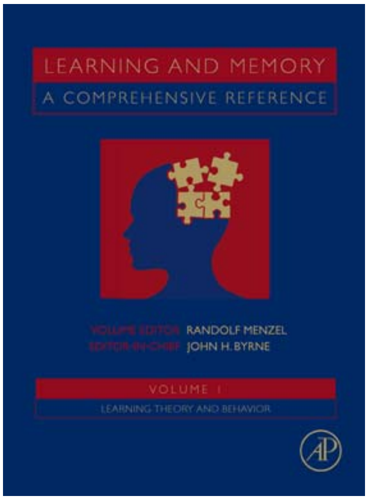
Provided for non-commercial research and educational use.
Not for reproduction, distribution or commercial use.

This article was originally published in the *Learning and Memory: A Comprehensive Reference*, Volumes 1-4 published by Elsevier, and the attached copy is provided by Elsevier for the author's benefit and for the benefit of the author's institution, for non-commercial research and educational use including without limitation use in instruction at your institution, sending it to specific colleagues who you know, and providing a copy to your institution's administrator.



All other uses, reproduction and distribution, including without limitation commercial reprints, selling or licensing copies or access, or posting on open internet sites, your personal or institution's website or repository, are prohibited. For exceptions, permission may be sought for such use through Elsevier's permissions site at:

http://www.elsevier.com/locate/permissionusematerial

J-P Changeux. The Neuronal Workspace Model: Conscious Processing and Learning.
In R. Menzel (Ed.), Learning Theory and Behavior. Vol. [1] of Learning and Memory:
A Comprehensive Reference, 4 vols.
(J.Byrne Editor), pp. [729-758] Oxford: Elsevier.

Author's personal copy

1.38 The Neuronal Workspace Model: Conscious Processing and Learning

J.-P. Changeux, URACNRS 2182, Collège de France and Institut Pasteur, Paris, France

S. Dehaene, Collège de France, Paris, France; and INSERM-CEA Cognitive Neuroimaging Unit, Neurospin Center, Gif-sur-Yvette, France

© 2008 Elsevier Ltd. All rights reserved.

1.38.1 Introduction

Researchers in the field of computational biology have mostly focused their attention in the recent years on the sequences of eukaryotic genomes, on their annotation, and on the understanding of how these linear sources of information give rise to the three-dimensional organization of the body and, in

Author's personal copy

particular, of the brain. Ultimately, from the DNA sequences already stored *in silico* of the human genome, one should be able to compute the main features of the species-specific functional organization of our brain. Yet, the brain strikingly differs from the other organs of the body in several respects. First, the complexity of its cellular and supracellular organization is orders of magnitudes higher, and second, it is able to learn and store information from the outside world. Finally, its structure is under constant reorganization as a function of its internal physiological states of activity, either endogenously generated or evoked by signals from the outside world. These are a few of the many reasons why computational neuroscience has recently developed as a rather autonomous and fast-moving discipline. Its principal project is to understand the multiple modes of signal processing by the brain ultimately resulting in behavior and/or tacit mental events. It is also to build up formal models, expressed in terms of neuronal networks, that link the molecular, neuronal, physiological, and behavioral/mental data in a coherent, noncontradictory though minimal form (Changeux and Dehaene, 1989). Ultimately implemented as 'formal organisms' (Changeux et al., 1973), these neurocomputational models should altogether account for the available data and produce experimentally testable predictions at all those levels. Being minimal, they are not anticipated to give an exhaustive description of reality, but nevertheless to validate or invalidate theories and, if necessary, give rise to new ones, thus contributing to our understanding of how the human brain works.

If multiple attempts have been successfully done to model the relationships between the states of activity of neuronal networks and overt behaviors in simple systems (Grillner and Graybiel, 2006), a most fascinating intellectual challenge of today's neuroscience remains to understand the explicit processing of the mental events which invest our brain, in other words to establish a comprehensive theory of consciousness on the basis of the presently available scientific knowledge. As stated by Delacour (1997),

> Multiple explanations have already been suggested like such-and-such a sophisticated algorithm, the oscillation of the extracellular electrical field in the cortex, the probabilistic character of synaptic transmission, or some still mysterious property of 'quantum gravity.' These theories represent little advance over the pineal gland theory of René Descartes. (Delacour, 1997: 127)

The case of consciousness is indeed exceptional. First of all, a fundamental ambiguity exists in the use of the term which, depending on context and author, can equally refer to waking state, personal experience, mental processing, or the multifaceted concept of self. Second, its comprehensive description crosses multiple disciplines ranging from basic neuroscience to molecular biology, psychology, and philosophy which, in the present situation of the academic community, show considerable difficulties to interact and cooperate. Last, any plausible theory on consciousness has to refer to a phenomenal subjective experience reported through introspection that a long tradition, from positivism to behaviorism and, most of all, the current practices of bench work research in neuroscience or pharmacology, has banned from any serious form of scientific enquiry.

Nevertheless, in the past decades, the situation has significantly changed because of a 'renaissance' (Zeman, 2005) of empirical research on consciousness and the proposal of constructive and plausible mechanistic theories, which aim at accounting for the subjective experience of a unified or global 'space or scene,' where some kind of *synthesis* between past, present, and future takes place, where multimodal perceptions, emotions, and feelings (present), and evoked memories of prior experiences (past), together with anticipations of actions (future), become subjectively integrated in a continuously changing and dynamic flow of consciousness, "altogether one and multiple at any of its moments" (Fessard, 1954; Bogen, 1995; Edelman and Tononi, 2000; Crick and Koch, 2003; Dehaene and Changeux, 2004). Among these theories, Zeman (2005) distinguishes the following.

1.38.1.1 Information Processing Theories

Following the lead of William James in associating consciousness with selective attention and memory, Baars (1989) has proposed a psychological model which postulates that the content of consciousness is broadcasted to the whole brain through a 'global workspace' which recruits the operation of multiple unconscious and automatic processors. Yet, Baars proposed as the essential neural basis of his global workspace the ascending reticular formation, the nonspecific nuclei of the thalamus, and only casually mentioned that "it is possible that corticocortical connections should also be included."

Author's personal copy

Shallice (1988) has also suggested an integrative role of consciousness through a 'supervisory attentive system' which would control the activities of lower level psychological systems mediated by some kind of 'contention scheduling' system and has placed emphasis on parieto-prefrontal networks in relation with the supervisory system (see also Frith et al., 1999).

1.38.1.2 Neurobiological Theories

In their pioneering efforts to specify the neural correlates of consciousness, Crick and Koch (Crick, 1994; Crick and Koch, 1995, 2004) have successively emphasized the importance of gamma-band oscillations around 40 Hz as a correlate of conscious processing; then, successively, the role of connections to and from prefrontal cortex in conscious perception (though more recently they have defended the opposite view that prefrontal cortex works as an 'unconscious homunculus'); and last, the possible role of the claustrum in the integration of conscious percepts (Crick and Koch, 2005).

Edelman and Tononi (2000) and Tononi and Edelman (1998) have emphasized the role of information integration and of reentrant connections in establishing a shifting assembly or 'dynamic core' linking distributed cortical and thalamic neurons. Its representation content, at the same time diversified and unitary, could not be localized to single parts of the brain and would vary significantly among individuals, but yet would correspond to the content of phenomenal consciousness. Lumer et al. (1997a,b) have developed a formalism in their simulations focused on early visual processing, with reentrant connections but without establishing a link with the notion of consciousness and specifically with the dynamic core hypothesis.

The hypothesis of a conscious 'neuronal workspace' (Dehaene et al., 1998; Dehaene and Changeux, 2000; Dehaene and Naccache, 2001; Dehaene et al., 2003b) emphasizes the role of distributed neurons with long-distance connections, particularly dense in prefrontal, cingulate, and parietal regions, interconnecting multiple specialized processors and broadcasting signals at the brain scale in a spontaneous and sudden manner, forming a conscious 'global neuronal workspace.' This model is extensively presented and discussed in this chapter.

1.38.1.3 Social Theories

The philosopher Strawson (1974) has argued that the concept of one's own mind presupposes the concept of other minds. This, together with the notion that language is critical for human consciousness, has led to the notion that conscious experience would be more a social construction than a physiological or psychological phenomenon (Rose, 1999). Without contesting the importance of social relationships and in particular of language (see Edelman, 1989) in human consciousness, one may wonder, however, whether this is the adequate level of explanation for a comprehensive theory of consciousness, which should ultimately capture such basic phenomena as sleep and anesthesia, masking, or attentional blink (see following).

In this chapter we successively analyze: (1) the theoretical premises of the neuronal workspace hypothesis; (2) the formal representation of the neuronal workspace model; (3) simulations with the workspace model of states of consciousness and access to consciousness in cognitive tasks; and (4) the neuronal workspace model and the evolution of consciousness.

1.38.2 The Neuronal Workspace Hypothesis: Premises and Theoretical Statements

The views presented in this chapter developed from three sets of complementary data: anatomical, computational, and psychophysical.

1.38.2.1 Anatomical Data

Early observations by Cajal (1892) (see DeFelipe and Jones, 1988) underlined the 'special morphology' of the pyramidal cells from the cerebral cortex and suggested that they might be "the substratum of the highest nervous activities," calling them '*the psychic cells*.' Cajal mentioned their very numerous and complex dendritic cells and also noted that "the pyramidal cells from cortical layers II and III possess long axons with multiple collaterals." Cajal further distinguished in the white matter: projection fibers which enter the cerebral peduncle, callosal fibers which associate the two hemispheres, and fibers of association that "bring into relation . . . different territories and different lobes of the same hemisphere" (Cajal, 1892). He also noted that these fibers of association increase in number in humans and large

Author's personal copy

mammals, where they form "the main mass of the white matter." Recent investigations have confirmed the view that the corticocortical and callosal fibers primarily (though not exclusively) arise from layer II-III pyramids (Jones, 1984) (**Figure 1(d)**).

Furthermore, von Economo (1929), a follower of Brodman, distinguished in his "Cytoarchitectonics of the human cerebral cortex" five "fundamental types of cortical structure" (**Figures 1(a)** and **1(b)**) and among them: the "frontal type 2 (which) . . . possesses large, well-formed, and well-arranged pyramidal cells in layers III and V" while in the "parietal type 3 (these cells) are smaller, more slender, and numerous . . . ." Von Economo also noted that the type 2 is "spread over the anterior two-thirds of the frontal lobe, over the superior parietal lobule . . ." as well as over the cingulate cortex, among other cortical areas and concludes that "type 2 and 3 isocortex . . . are the chief station for the commemorative and higher psychic functions."

Interestingly, recent quantitative analysis of the dendritic field morphology of layer III pyramidal neurons in the occipitofrontal cortical 'stream' revealed a continuous increase of complexity up to the prefrontal cortex within a given species (Elston and Rosa, 1997, 1998; DeFelipe and Farinas, 1992) (**Figure 1(c)**) and from lower species (owl monkey, marmoset) to humans (Elston, 2003). A correlative increase of the relative surface of the prefrontal cortex accompanies this increased complexity (see Changeux, 2004). Moreover, mapping of long-range connections in the monkey cerebral cortex revealed long-range connections linking, among others, the prefrontal cortex (area 46), the superior temporal sulcus, parietal area 7a, and the hippocampus together with the contralateral anterior and posterior cingulum, area 19, and the parahippocampal gyrus (Goldman-Rakic, 1988). These circuits were suggested to contribute to working memory (Goldman-Rakic, 1994) and

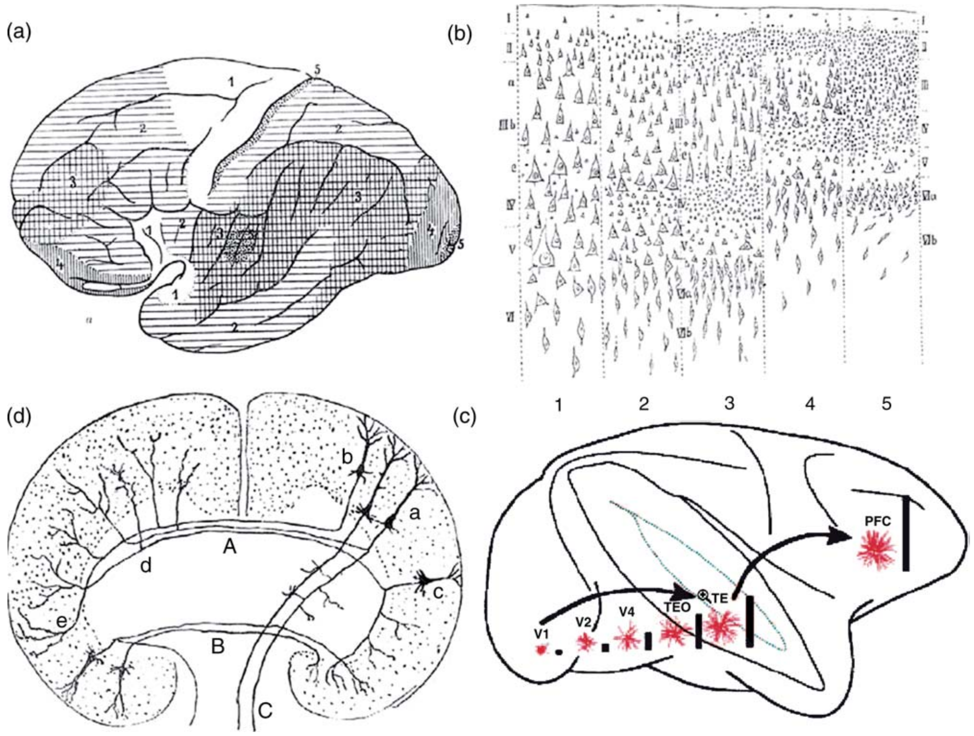


Figure 1 The anatomical basis of the neuronal workspace hypothesis. The pyramidal neurons with long corticocortical axons from layers 2–3 of the cerebral cortex (d) display increased complexity of the basal dendrites (and thus increased connectivity) from primary visual areas (V1) to prefrontal cortex (PFC) (here in the monkey (c)) and from primitive vertebrates to humans. These neurons are particularly abundant in what von Economo referred to as type 2 cortex (b), which primarily occupies the frontal, parietotemporal, and cingulate areas (a). Parts (a) and (b) from von Economo C (1929) *The Cytoarchitectonics of the Human Cerebral Cortex.* London: Oxford University Press; used with permission. (c) from Elston GN (2003) Cortex, cognition and the cell: New insights into the pyramidal neuron and prefrontal function. *Cereb. Cortex* 13: 1124–1138; used with permission from Oxford University Press. (d) from Cajal S (1892) El nuevo concepto de la histologia de los centros nerviosos. *Rev. Ciencias Med.*18: 457–476; used with permission.

Author's personal copy

proposed to contribute to the anatomical basis of the neuronal workspace model (Dehaene et al., 1998).

1.38.2.2 Computational Data

The present neurocomputational approach to conscious versus nonconscious processing originates from the design of neural network models that aimed at specifying the contribution of prefrontal cortex to increasingly higher cognitive tasks (Dehaene and Changeux, 1989, 1991, 1997; Dehaene et al., 1998). Successively, these models considered the issues of how a network could retain an active memory across the long delay of a delayed-response task (Dehaene and Changeux, 1989), how it could encode abstract rules that might be selected from external or internal rewards (Dehaene and Changeux, 1991), and finally how networks based on those principles could pass complex planning tasks such as the Tower of London test or the Stroop test (Dehaene and Changeux, 1997; Dehaene et al., 1998). The 'conscious *neuronal* workspace' model (Dehaene et al., 1998; Dehaene and Naccache, 2001; Dehaene et al., 2003b; Dehaene and Changeux, 2005) is the last development of these models which emphasize the role of distributed neurons with long-distance connections, particularly dense in prefrontal, cingulate, and parietal regions, which are capable of interconnecting multiple specialized processors and can broadcast signals at the brain scale in a spontaneous and sudden manner.

1.38.2.3 Cognitive Psychology, Neuropsychology, and Human Neuroimaging

A long psychophysical and neuropsychological tradition, dating back to Hughlings Jackson and perpetuated among others by Baddeley, Shallice, Mesulam, or Posner, has emphasized the hierarchical organization of the brain and separates lower automatized systems from increasingly higher and more autonomous supervisory executive systems. It has also been influenced by Fodor's distinction between the vertical 'modular faculties' and a distinct 'isotropic central and horizontal system' capable of sharing information across modules. Empirically, finally, it has taken advantage of a variety of experimental techniques, starting with behavioral analysis, neuropsychological observation in brain-lesioned patients, and most recently, human neuroimaging with functional magnetic resonance imaging (fMRI) and causal interference with transcranial magnetic stimulation (TMS).

It is beyond the scope of this chapter to discuss the variety of experimental contributions that are relevant to the ongoing science of consciousness (see, e.g., Laureys, 2005), but they can be briefly sketched as belonging to two main lines of research. The first line, starting with the pioneering research of Weiskrantz on blindsight and Marcel on subliminal priming, has investigated the extent of nonconscious processing in humans. In combination with fMRI, this research has uncovered that not only subcortical, but in fact a variety of specialized cortical systems were capable of activating in the absence of any conscious report of stimulus presence (for review, see Naccache and Dehaene, 2001; Naccache et al., 2005; Dehaene et al., 2006). The second line, exemplified by Posner's or Shallice's work, has studied the properties of a central executive or executive attention system, whose activity seems to index conscious top-down attention and control (Norman and Shallice, 1986; Amati and Shallice, 2007). Through the use of Baars's contrastive method, which consists of opposing two minimally different conditions, one of which is conscious and the other is not, it has been observed that executive control is deployed only following consciously perceived trials (e.g., Kunde, 2004) and is consistently associated with a sharp increase in dorsal and midline prefrontal as well as anterior cingulate and, in many cases, inferior or posterior parietal activation (see Gusnard and Raichle, 2001; Dosenbach et al., 2006).

1.38.2.4 The Model

Inspired by these observations, our theoretical work attempted to capture them within a minimal theoretical model. In the following, the headlines of the theoretical premises will be presented following the initial presentation of Dehaene et al. (1998), yet updated in a few of its formulations in the subsequent papers of Dehaene et al. (2003a) and Dehaene and Changeux (2005).

1.38.2.4.1 Two computational spaces

The neuronal workspace hypothesis distinguishes, in a first approach, two main computational spaces within the brain (**Figure 2**), each characterized by a distinct pattern of connectivity:

a. A processing network, composed of a set of parallel, distributed, and functionally specialized

Author's personal copy

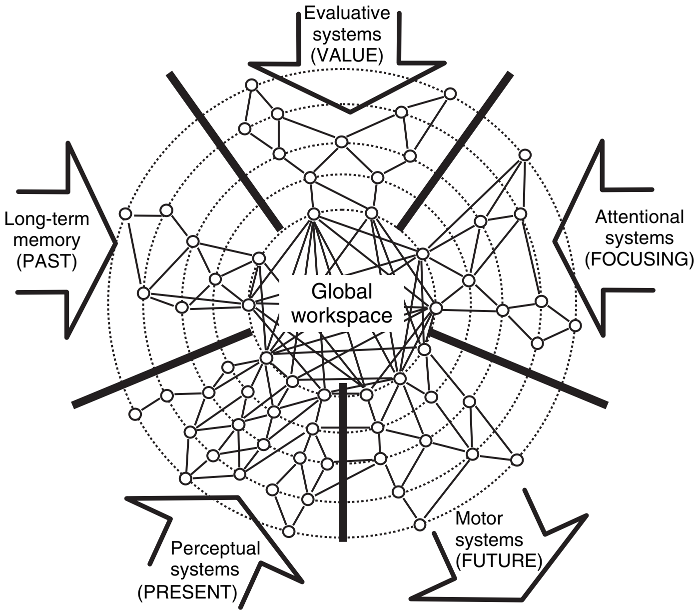


Figure 2 Schematic representation of the neuronal workspace hypothesis as initially proposed by Dehaene et al. (1998). The model distinguishes two computational spaces: (1) specialized processors, which are modular, encapsulated, and automatic, labeled here as perceptual systems, long-term memory (including autobiographic memory and self), and attentional and evaluative systems and (2) the global workspace, with long-range axon neurons broadcasting signals to multiple areas yielding subjective experience and reportability. From Dehaene S, Kerszberg M, and Changeux JP (1998) A neuronal model of a global workspace in effortful cognitive tasks. *Proc. Natl. Acad. Sci. USA* 95: 14529–14534; used with permission from the National Academy of Sciences.

processors (Baars, 1989) or modular subsystems (Shallice, 1988) subsumed by topologically distinct cortical domains with highly specific local or medium-range connections that encapsulate information relevant to its function. This specialized network processes information in a bottom-up manner (see Miyashita and Hayashi, 2000).

b. A global workspace, consisting of a distributed set of cortical neurons characterized by their ability to receive from and send back to homologous neurons in other cortical areas, horizontal projections through long-range excitatory axons (which may as well impinge on excitatory or inhibitory neurons). Such long-range corticocortical tangential connections include callosal connections and mostly originate from the pyramidal cells of layers 2 and 3. We therefore propose that the extent to which a given brain area contributes to the global workspace would be simply related to the fraction of its pyramidal neurons contributing to layers 2 and 3, which is particularly elevated in von Economo's type 2 (dorsolateral prefrontal) and type 3 (inferior parietal) cortical structures (von Economo, 1929). The pyramidal neurons from layers 2 and 3 establish, in addition, vertical and reciprocal connections with layer 5 neurons and thus corresponding thalamic nuclei. These connections contribute to both the stability and the dynamics of workspace activity, via, for instance, self-sustained circuits, but also mediate the direct access to and from the processing networks (Brecht et al., 2003). The global network neurons typically process information in a top-down manner.

1.38.2.4.2 Content of the global workspace

In the original formulation of the neuronal workspace hypothesis (Dehaene et al., 1998), five major categories of processors were distinguished which could be dynamically mobilized and multiply reconfigured (**Figure 2**).

a. *Perceptual circuits* give the workspace access to the present state of the external world. Empirically, perceptual circuits may include the primary and secondary sensory areas together with the object-oriented ventral and lateral areas of the temporal lobes in both visual (Mishkin and Ungerleider, 1982; Goodale et al., 1991; Fang and He, 2005) and auditory (Rauschecker and Tian, 2000) modalities as well as

Author's personal copy

the temporal and inferior parietal areas involved in language comprehension (including Wernicke's area) (Mesulam, 1998). Accordingly, the content of any external stimulus, attended object, or linguistic input can access the global workspace. Such access may take place stepwise through hierarchical stages of processing through primary and secondary sensory areas such as V1 and FEF (Lamme and Roelfsema, 2000) and then higher association areas of temporal, frontal, and cingulate cortex. In Dehaene et al. (2003a) and Dehaene and Changeux (2005) formulations, each area was further assumed to establish with the neighboring area bottom-up feedforward connections and top-down feedback projections, the top-down connections being slower, more numerous, and more diffuse (Felleman and van Essen, 1991; Salin and Bullier, 1995). Moreover, bottom-up connections were thought to impinge on glutamate α-amino-3-hydroxy-5-methyl-4-isoxazole propionic acid (AMPA) receptors, whereas the top-down ones would primarily mobilize glutamate *N*-methyl-D-aspartate (NMDA) receptors (see Lumer et al., 1997).

b. *Motor programming circuits* allow the content of the workspace to be used to guide future motor behaviors and actions. A hierarchy of nested circuits implements motor intentions, from the highest level of abstract plans to individual actions, themselves composed of moves and gestures (Jeannerod and Jacob, 2005).

Empirically in humans, these circuits include premotor cortex, posterior parietal cortex, supplementary motor area, basal ganglia (notably the caudate nucleus), and cerebellum, as well as the high-level speech production circuits of the left inferior frontal lobe, including Broca's area. Connections of the workspace to motor and language circuits at the higher levels of this hierarchy endow any active representation in the workspace with the property of reportability (Weiskrantz, 1997; Dehaene et al., 2006), namely, the fact that it can be described or commented on using words or gestures.

c. *Long-term memory circuits* provide the workspace with an access to stored past percepts and events.

These long-term memory stores are likely distributed throughout the cortex according to their original content and modality; hippocampal and parahippocampal areas through reciprocal links with workspace neurons may play a special role in mediating the storage in and retrieval from these long-term stores.

d. *Evaluation circuits* (Dehaene and Changeux, 1989, 1991; Friston et al., 1994; Schultz et al., 1997) allow representations in the workspace to be selected according to a positive or negative value.

Empirically, the main anatomical systems in this respect include mesocortical noradrenergic, dopaminergic, serotoninergic, and cholinergic pathways together with the orbitofrontal cortex, anterior cingulate, hypothalamus, amygdala, and ventral striatum (see Everitt and Robbins, 2005). Autoevaluation systems develop from reciprocal projections allowing evaluation circuits to be internally activated by the current workspace content (Dehaene and Changeux, 1991) and, conversely, to selectively maintain or change workspace activity according to whether its value is predicted to be positive or negative (Dehaene and Changeux, 1991, 1995, 1997). These evaluation systems are the targets of drugs as instrumental reinforcers eventually resulting in drug self-administration or drug taking and mobilizing in particular medial prefrontal cortex and nucleus accumbens core (Everitt and Robbins, 2005; Christakou et al., 2004).

e. *Attention circuits* allow the workspace to mobilize its own circuits independently from the external world. Changes in workspace contents need not necessarily lead to changes in overt behavior, but may result in covert attention switches to selectively amplify or attenuate the signals from a subset of processor neurons.

Although all descending projections from workspace neurons to peripheral modular processors are important in this selective amplification process, a particular role is played by areas of the parietal lobe in visuospatial attention (Posner, 1994; Posner and Dehaene, 1994).

1.38.2.4.3 Global modulation of workspace activation

The state of activation of workspace neurons is assumed to be under the control of global vigilance signals from the ascending reticular activating system. Empirically they may include cholinergic nuclei in the upper brainstem and basal forebrain, noradrenergic nuclei (e.g., from the locus coeruleus), a histaminergic projection from the posterior hypothalamus, and dopaminergic and serotoninergic pathways arising from the brainstem (McCarley, 1999) together with recently identified orexin neurons from lateral hypothalamus (de Lecea et al., 1998; Harris et al., 2005). Much if not all of the influence exerted by these pathways is mediated by the thalamus and characterized by an increase in the excitability of the corticothalamic neurons.

Author's personal copy

Slow-wave sleep, on the other hand, coincides with a reduction of activity in the cholinergic, noradrenergic, and histaminergic nuclei, the anterior hypothalamus and basal forebrain being candidates for a critical role in sleep induction (Zeman, 2001). These signals are powerful enough to control major transitions between the awake state (workspace active) and slow-wave sleep (workspace inactive) states. Others provide graded inputs that modulate the amplitude of workspace activation, which is enhanced whenever novel, unpredicted, or emotionally relevant signals occur and, conversely, drops when the organism is involved in routine activity.

In the waking state the corticothalamic neurons are tonically depolarized by a blocking of hyperpolarizing potassium conductance (Steriade, 1999) (e.g., by acetylcholine acting on muscarinic receptors and norepinephrine on α_1-adrenergic receptors) switching them out of the slow-bursting mode and into fast gamma-band oscillations (Steriade et al., 1993; Llinas and Steriade, 2006). Similar effects can be obtained by electrical stimulation of the brainstem or by direct application of acetylcholine (McCormick and Bal, 1997). Moreover, mutations in the $\alpha4$ and $\beta2$-subunit genes of the nicotinic acetylcholine receptor cause autosomal dominant frontal lobe epilepsy (Steinlein, 2004), and deletion of the $\beta2$-subunit is accompanied by a decrease of micro-arousals, which take place during slow-wave sleep (Léna et al., 2004).

As a consequence, the corticothalamic neurons show increased excitability: their signal-to-noise ratio is increased, and the response to sensory stimuli is facilitated. Their spontaneous activity is high and characterized by stochastic independence of time intervals between successive action potentials (Llinas and Paré, 1991; Steriade et al., 1993; Llinas and Steriade, 2006). By contrast, during slow-wave sleep, where consciousness is absent, signal-to-noise ratios to sensory responses are decreased, and most neurons tend to discharge in bursts synchronized over large populations, thus introducing distortions or blocking information transmission (Livingstone and Hubel, 1981).

In the resting awake state, the brain is the seat of an important baseline (Gusnard and Raichle, 2001; Raichle and Gusnard, 2005) or ongoing metabolic activity; a very large fraction of it (about 80%) being correlated with glutamate cycling and, hence, active synaptic signaling processes (Shulman and Rothman, 1998; Hyder et al., 2002; Raichle and Gusnard, 2002; Shulman et al., 2004). During slow-wave sleep, anesthesia, or coma, global cerebral glucose metabolism falls by about 20% (Heiss et al., 1985; Buchsbaum et al., 1989; Shulman et al., 2004), particularly in frontal and parietal cortices (Laureys, 2005; Laureys et al., 2006). Interestingly, optical imaging of visual cortex in anesthetized animals revealed structural states of activity which have a similar global organization as activity patterns evoked by external stimuli (Tsodyks et al., 1999; Kenet et al., 2003).

1.38.2.4.4 Spatiotemporal dynamics of workspace activity

From a theoretical point of view, the global workspace is considered the seat of a particular kind of brain-scale activity state characterized by the spontaneous activation, in a sudden, coherent, and exclusive manner, referred to as ignition (Dehaene and Changeux, 2005), of a subset of workspace neurons, the rest of workspace neurons being inhibited. The transition to this state of highly correlated activity is fast and characterized by an amplification of local neural activation and the subsequent activation of multiple distant areas. The entire workspace is globally interconnected in such a way that only one such workspace representation can be active at any given time (see Sigman and Dehaene, 2005, 2006). This all-or-none invasive property distinguishes it from peripheral processors in which, due to local patterns of connections, several representations with different formats may coexist.

A representation which has invaded the workspace may remain active in an autonomous manner and resist changes in peripheral activity (see Dehaene and Changeux, 1989, 1991). If it is negatively evaluated, or if attention fails, it may, however, be spontaneously and randomly replaced by another discrete combination of workspace neurons. Functionally, this neural property implements an active generator of diversity which constantly projects and tests hypotheses (or prerepresentations) on the outside world (Dehaene and Changeux, 1989, 1991, 1997). The dynamics of workspace neuron activity is thus characterized be a constant flow of individual coherent episodes of variable duration and their selection.

Although a variety of processor areas project to the interconnected set of neurons composing the global workspace, at any given time only a subset of inputs effectively accesses it. We postulate that this gating is implemented by descending modulatory projections from workspace neurons to more peripheral processor neurons. These projections may selectively amplify or extinguish the ascending inputs from processing neurons, thus mobilizing, at a given time, a specific set of

Author's personal copy

processors in the workspace while suppressing the contribution of others. In other words, the pattern of mobilized processor neurons defines the actual subjective content of conscious perception.

1.38.3 Formal Representation of the Neuronal Workspace Model

Since the initial formulation of Dehaene et al. (1998), the general architecture and dynamics of the neural network representing the global workspace and the relevant processors have been further specified in Dehaene et al. (2003a) and Dehaene and Changeux (2005). It is well understood that in any instance these computer simulations are partial and incomplete. Yet they are expected to point to the importance of particular components or features of these minimal architectures, thus leading to critical experimental tests. In our work, we found it useful to develop two quite distinct types of computer simulations. Some of them, referred to here as 'Type 1' models, were intended to describe an entire task-related cognitive architecture and thus focused more on global connectivity than on fine physiological details (e.g., Dehaene et al., 1998). Others, referred to as 'Type 2' models, were intended to capture some fine-grained physiological characteristics of neuronal firing trains and event-related potentials during conscious and subliminal perception. These simulations therefore necessarily incorporated considerably more physiological details of receptor types and cortical layers, but were not extended in a brain-scale architecture solving a precise task (e.g., Dehaene et al., 2003a; Dehaene and Changeux, 2005). Here, we describe their principles in turn.

1.38.3.1 Detailed Physiological Simulations of Access to the Conscious Workspace

In those Type 2 simulations, we intended to describe only part of the workspace (Dehaene and Changeux, 2005), but to do so with physiological details. The goal was to simulate the bottom-up/top-down interactions occurring between four hierarchically organized areas, the lowest of which was in contact with the external world while the highest was assumed to contact other workspace areas (not simulated).

1.38.3.1.1 Single-neuron model

The model (**Figure 3(a)**), adapted from Lumer et al. (1997a,b), was simulated at the level of single-compartment integrate-and-fire neurons whose membrane potential evolved according to semirealistic differential equations taking into account realistic temporal delays and AMPA, NMDA, and gamma-aminobutyric acid (GABA) currents. Neurons also received a diffuse neuromodulator input summarizing the known depolarizing effects of ascending activating systems, such as those from cholinergic, noradrenergic, and serotoninergic nuclei in the brainstem, basal forebrain, and hypothalamus (Steriade et al., 1993; Llinas and Steriade, 2006). This parameter was used to control the level of wakefulness (see following discussion).

1.38.3.1.2 Columnar structure

The neurons were organized into simulated thalamocortical columns comprising 80 excitatory and 40 inhibitory neurons and organized in a three-layered structure, schematizing supragranular, infragranular, and layer IV cortical neurons and a corresponding thalamic sector (**Figure 3(b)**). A fairly realistic scheme of connections was implemented, whereby thalamic excitatory neurons projected to layer IV (AMPA, 3 ms delay) and, with lesser strength, to infragranular neurons (AMPA, 3 ms). Layer IV excitatory neurons projected to supragranular neurons (AMPA, 2 ms). Supragranular excitatory neurons projected to infragranular neurons (AMPA, 2 ms). Finally, infragranular excitatory neurons projected to layer 4 (AMPA, 7 ms), to supragranular neurons (AMPA, 7 ms), and to the thalamus (AMPA, 8 ms). Those principles and parameter values capture the major properties of translaminar connections (Lumer et al., 1997a,b; Douglas and Martin, 2004), though they do not attempt to capture the possible functional roles of the different layers (see, e.g., Raizada and Grossberg, 2003).

1.38.3.1.3 Long-range connections

For corticocortical projections, supragranular excitatory neurons of each area projected to layer IV of the next area (AMPA, 3 ms). In agreement with physiological observations (Felleman and Van Essen, 1991; Salin and Bullier, 1995), top-down connections were slower, more numerous, and more diffuse. They connected the supra- and infragranular excitatory neurons of a given column to the supra- and infragranular layers of all areas of a lower hierarchical level. Strong top-down connections linked columns coding for the same stimulus, whereas weaker top-down connections projected to all columns of a lower area. Both were NMDA mediated, and transmission delays increased with cortical distance (delay $= 5 + 3\delta$ ms,

Author's personal copy

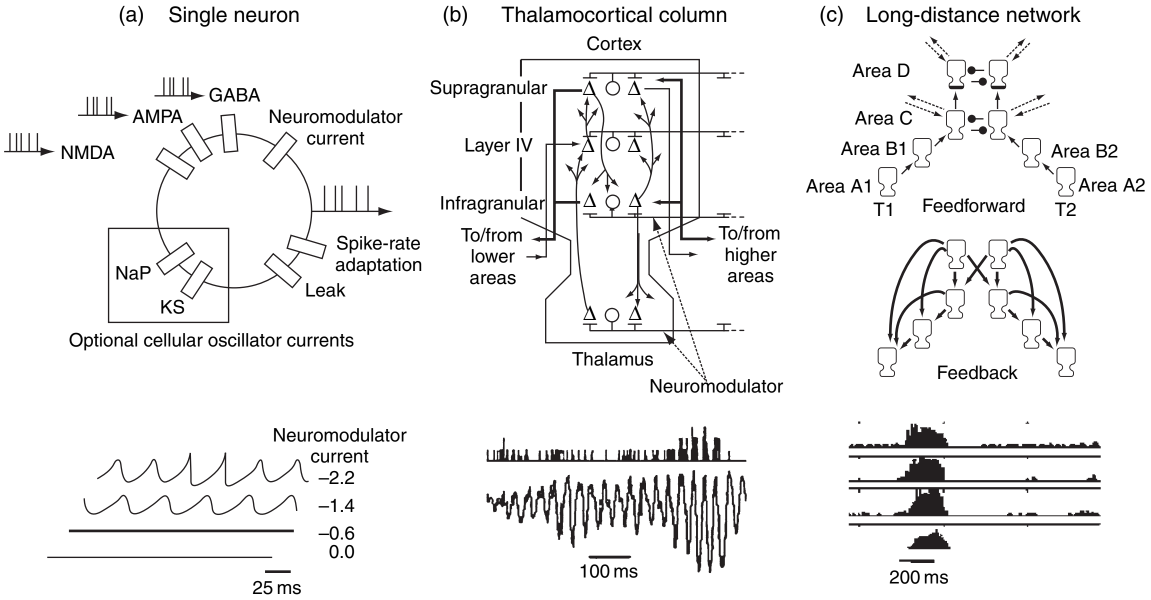


Figure 3 Detailed implementation, at the cellular and molecular levels, of the neuronal workspace model in the case of access to consciousness as in Dehaene et al. (2003b) and Dehaene and Changeux (2005). Single neurons may generate sustained oscillations of membrane potential at high enough level neuromodulator current (a) but only the thalamocortical column (b), and global network levels do generate complex waxing and waning EEG-like oscillations (b) and metastable global states of sustained firing (or 'ignition') (c). From Dehaene S, Sergent C, and Changeux JP (2003b) A neuronal network model linking subjective reports and objective physiological data during conscious perception. *Proc. Natl. Acad. Sci. USA* 100: 8520–8525, with permission from the National Academy of Sciences; and Dehaene S and Changeux JP (2005) Ongoing spontaneous activity controls access to consciousness: A neuronal model for inattentional blindness. *PLoS Biol.* 3: 910–927, with permission from the Public Library of Science.

with $\delta = 1$ for consecutive areas, 2 for areas two levels apart in the hierarchy, etc.).

1.38.3.1.4 Spontaneous activity

Although we studied stimulus-evoked activity (see later discussion), a most important goal of these stimulations was to also capture spontaneous corticothalamic activity and its modulation with states of vigilance. We studied two possible sources for spontaneous activity, both of which were meant as theoretical idealizations on a continuum of possibilities. The first case, hereafter the 'cellular oscillator model,' was a purely deterministic model in which neurons follow simple differential equations incorporating persistent sodium and slowly inactivating potassium currents whose interplay generates intrinsic gamma-band oscillations of membrane potential (Wang, 1993), comparable to those recorded experimentally (Llinas et al., 1998). We also described another simulation, hereafter called the 'random spikes' model, in which stochastic spontaneous activity arose from fast random fluctuations in membrane potential, capturing the joint effects of synaptic and postsynaptic noise on spike initiation. Both cellular sources turned out to have a similar effect on global spontaneous corticothalamic states (for details, see Dehaene and Changeux, 2005).

1.38.3.2 Minimal Models of Cognitive Architectures for Effortful Tasks

Another line of models, Type 1 simulations, attempted to capture behavioral and neuroimaging observations on higher-level 'executive' tasks that depend on prefrontal cortex (Dehaene et al., 1998), such as the Delayed-Response, Stroop, Wisconsin, or Tower of London tests (Dehaene and Changeux, 1989, 1991, 1997). We have not found it possible to design a pertinent model of such tasks while working at the level of detailed spiking neurons and columns. Thus, those models were of a more abstract nature and incorporated neural 'units' formally similar to single neurons or clusters whose average firing rate was simulated by McCulloch-Pitts units. No attempt was made to capture intracolumnar or thalamic dynamics, but the network incorporated a series of assemblies assumed to represent relevant cortical activity at several hierarchical levels.

Author's personal copy

1.38.3.2.1 Learning by reward

An interesting advantage of these coarser cognitive models is that they allowed for the simulation of longer periods of time and, therefore, of selective learning by reward, which empirically can be accompanied by drastic changes in vigilance and conscious access to task-relevant features. The network received a reward signal (R) provided after each network response ($R = +1$, correct; $R = -1$, incorrect). This reward led to two types of internal changes: classical synaptic weight changes of the Hebbian type, and a more original hypothesis of direct workspace activity modulation.

1.38.3.2.2 Synaptic weight changes

In the initial formulation of the theory (Dehaene et al., 1998) and for simplicity, only the synaptic weights between two excitatory units were assumed to be modifiable according to a reward-modulated Hebbian rule $\Delta w^{\text{post,pre}} = \varepsilon\, R\, S^{\text{pre}}\, (2\, S^{\text{post}} - 1)$, where R is the reward signal, pre is the presynaptic unit, and post is the postsynaptic unit (Dehaene and Changeux, 1989).

1.38.3.2.3 Workspace activity changes

Starting with our earliest modeling approaches (Dehaene and Changeux, 1989), we have assumed that reward entry can have either a stabilizing or destabilizing effect on prefrontal neuron activity. In Dehaene et al. (1998), we assumed that workspace neuron activity is under the influence of both vigilance and reward signals. The vigilance signal V is treated as having a modulatory influence on all workspace neurons. It is updated after each response: if $R > 0$, then $\Delta V = -0.1\, V$, otherwise $\Delta V = 0.5\ (1 - V)$. This has the effect of a slowly decreasing vigilance with sharp increases on error trials. The reward signal R influences the stability of workspace activity through a short-term depression or potentiation of synaptic weights (Dehaene and Changeux, 1989, 1991, 1997). A plausible molecular implementation of this rule has been proposed in terms of allosteric receptors (Dehaene and Changeux, 1989, 1991). It postulates that the time coincidence of a diffuse reward signal and of a postsynaptic marker of recent neuronal activity transiently shifts the allosteric equilibrium either toward, or from, a desensitized refractory conformation (Heidmann and Changeux, 1982, see also Changeux and Edelstein, 2005). Through this chemical Hebb rule, negative reward destabilizes the self-sustaining excitatory connections between currently active workspace neurons, thus initiating a change in workspace activity.

1.38.4 States of Vigilance as Spontaneous Thalamocortical Rhythms and Their Brain Imaging

The simulations of the vigilance states by Dehaene and Changeux (2005) type 2 modeling incorporate only minimal physiological mechanisms such as changes in a single current $I_{\text{neuromodul}}$, the depolarizing influence of ascending neuromodulation systems onto thalamic and cortical neurons (**Figure 3(a)**). They nevertheless suffice to generate a dynamical phase transition whose properties bear interesting similarity with actual empirical observations. We observed a robust threshold value of ascending neuromodulatory signaling beyond which structured neuronal activity emerged in the form of spontaneous thalamocortical oscillations in the gamma band (20–100 Hz, with a peak of the power spectrum around 40 Hz) (see also Bush and Sejnowski, 1996; Fuentes et al., 1996). The simulated waxing and waning synchronous bursts of oscillations bear similarity with empirical observations of transient periods of thalamocortical resonance, detected as bouts of gamma-band oscillations using electrophysiological recordings, for instance in the cat thalamus and cortex (Steriade et al., 1993; Steriade et al., 1996), or in humans using electro- and magneto-encephalography (Llinas and Ribary, 1993) (**Figure 3(b)**). They are proposed to represent the state of consciousness referred to as vigilance.

An original feature of the simulations is to characterize precisely the change in state in terms of a dynamical phase transition, referred to as a Hopf bifurcation. The Hopf bifurcation is continuous in the amplitude of spontaneous activity, which increases steadily from zero as vigilance increases. Thus it implements a true continuum of consciousness states, from high vigilance to drowsiness, and the various states of sleep anesthesia, or coma (Gajraj et al., 1999; Bonhomme et al., 2000; Sleigh et al., 2001). However, the Hopf bifurcation is also discontinuous in frequency space as the ascending neuromodulation increases. This may capture the observation that, during awakening or returning from anesthesia, there is a definite threshold for regaining of consciousness, which coincides with the threshold for emergence of high-frequency spontaneous thalamocortical oscillations.

Author's personal copy

Anatomically, the model predicts that in the awake state, spontaneous activity is present in all areas, but exhibits a higher degree of organization in higher cortical association areas, whose neurons are tightly interconnected by long distance into a global neuronal workspace and mobilize other low-level areas in a top-down manner. Thus, the model predicts that brain territories particularly rich in 'workspace neurons' with long-distance connections (i.e., prefrontal, parietal, superior temporal, and cingulate cortices) show the most intense and consistent spontaneous activity in the awake state. This prediction fits with the observation that the 'baseline' activity of the awake human brain at rest points to a network linking dorsal and ventral medial prefrontal, lateral parietotemporal, and posterior cingulate cortices (Gusnard and Raichle, 2001; Mazoyer et al., 2001; Raichle et al., 2001) which constantly fluctuates in synchrony with changes in electroencephalographic spectral content (Laufs et al., 2003) and shows the greatest drop in metabolism during anesthesia, sleep, coma, or the vegetative state (Maquet and Phillips, 1998; Fiset et al., 1999; Laureys et al., 2000; Paus, 2000; Balkin et al., 2002; Heinke and Schwarzbauer, 2002; Shulman et al., 2003). In striking agreement with the workspace model, volatile anesthetics have been recently shown to disrupt frontoparietal recurrent information transfer at gamma frequencies in the rat (Imas et al., 2005).

1.38.5 Interactions between External Stimuli and Ongoing Spontaneous Activity: Facilitation versus Competition

A most original aspect of Dehaene and Changeux's (2005) type 2 simulations on ongoing spontaneous activity concerns its interactions with external stimuli. These interactions can be facilitatory (higher spontaneous activity facilitates the detection of weak stimuli) or inhibitory (very high spontaneous activity preventing access to other external stimuli).

First, spontaneous activity may affect activation caused by external stimuli. The model predicts that ascending neuromodulatory current and the external input current combine in a smooth and largely additive fashion. The threshold for conscious access (ignition) is not fixed, but decreases as vigilance increases. At one extreme, very low levels of vigilance completely prevent the possibility of ignition, even by long and intense stimuli. Such stimuli only lead to a short pulse of activation through the thalamus and the early sensory areas of the model. Thus, we expect that early sensory signal can be processed, while higher cortical ones are attenuated, during altered states of consciousness. This prediction is consistent with empirical observations of auditory processing during sleep (Portas et al., 2000) or the vegetative state (Laureys et al., 2000), where stimuli activate the thalamus and auditory cortex, but fail to generate the distributed state of correlated prefrontal, parietal, and cingulate activity observed in awake normal subjects. Similar observations have been made with tactile or pain stimuli, suggesting that the lack of prefrontal–parietal–cingulate ignition is quite characteristic of those states (Laureys et al., 2002; Laureys et al., 2004).

Pharmacological agents such as nicotine, which can mimic and potentiate ascending cholinergic systems, might also have an influence on the perceptual threshold in visual masking or other psychophysical tests, which should be measurable both psychophysically and with brain imaging measures of ignition (e.g., prefrontal–cingulate activity in fMRI, P300 in event-related potentials). In schizophrenic patients, subliminal processing is intact, but the threshold for conscious perception of masked visual stimuli is increased, possibly relating to an impairment of top-down prefrontal–cingulate connectivity (Dehaene et al., 2003b; Del Cul et al., 2006). In such patients, we predict that nicotine might partially bring the conscious access threshold back toward its normal value.

Second, the model predicts that very high levels of spontaneous activity can prevent ignition by external stimuli (**Figure 4**). As mentioned, optical imaging of visual cortex in anesthetized animals has revealed structured states of spontaneous ongoing activity, which have the same global organization as activity patterns evoked by external stimulation (Tsodyks et al., 1999; Kenet et al., 2003). Moreover, high levels of spontaneous activity inhibit the sensory responses evoked by external stimuli, for instance, by whisker deflection in somatosensory cortex (Petersen et al., 2003). Such interactions with ongoing activity can provide an explanation for the large variability in spike trains evoked by repeated identical sensory stimuli (Arieli et al., 1996; Petersen et al., 2003).

The complete blocking of some incoming stimuli that occurs in the simulations of the model offers a plausible explanation for the psychological phenomenon of inattentional blindness (Newby and Rock, 1998). In this phenomenon, human observers engaged into an intense mental activity (such as

Author's personal copy

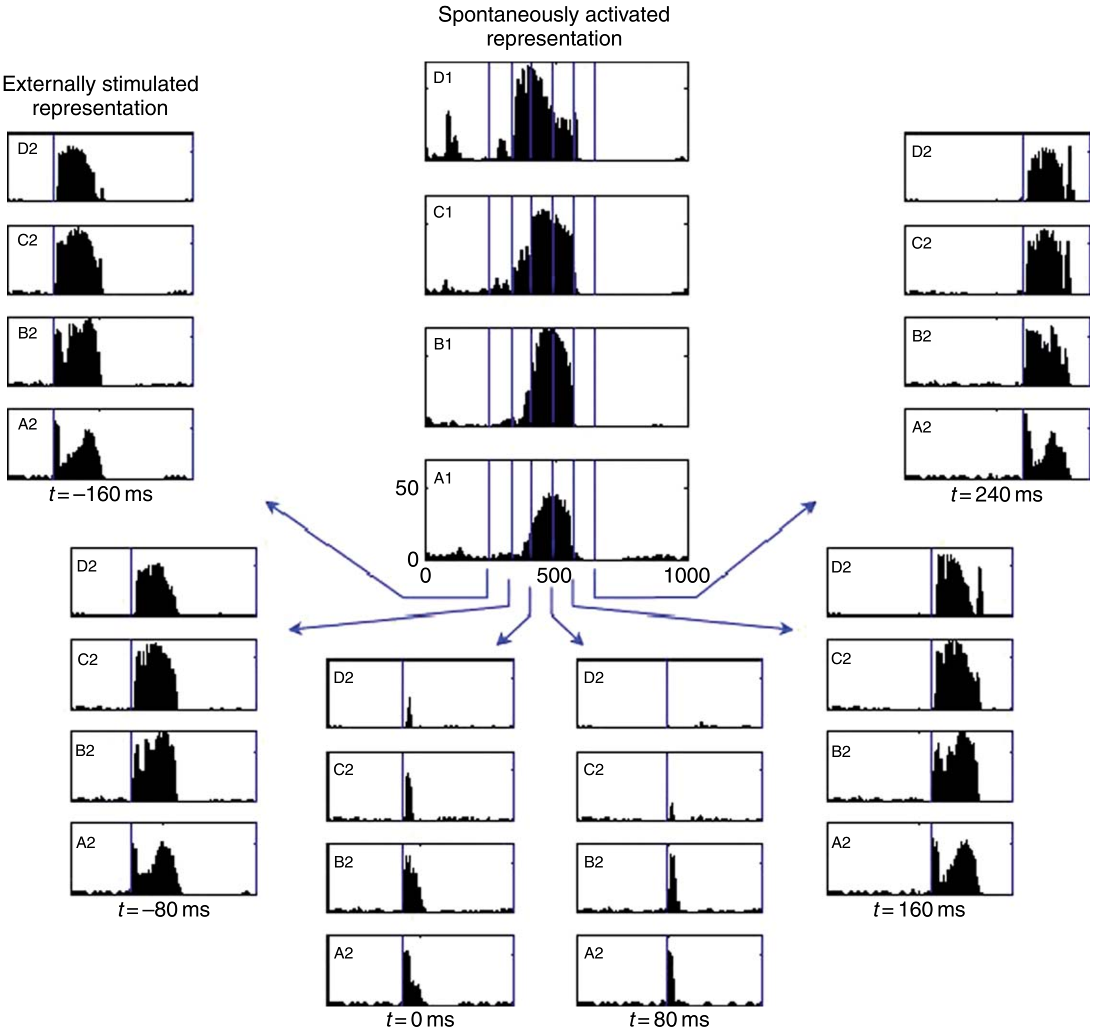


Figure 4 Competition between spontaneous workspace activity and external sensory simulation—a plausible model of the inattentional blindness state (Dehaene and Changeux (2005)). In the state of inattentional blindness, the subject fails to consciously detect external stimuli during periods of spontaneous thought. This can be reproduced by the simulations. In the boxes on the left of the figure, *before* the period of ongoing spontaneous activity, the external stimulus propagates in a bottom-up manner from the lowest (A) to the highest (B) areas of the sensory system yielding activation of the workspace (see **Figure 3(c)**). In the right boxes the same occurs *after* the period of spontaneous activity. On the other hand, when the external stimulus coincides with the period of spontaneous activity (center), its access to the workspace is inhibited. From Dehaene S and Changeux JP (2005) Ongoing spontaneous activity controls access to consciousness: A neuronal model for inattentional blindness. *PLoS Biol.* 3: 910–927; used with permission from the Public Library of Science.

detecting or counting stimuli of a certain type) become totally oblivious to other irrelevant stimuli, even when they occur within the fovea for a long duration (Simons and Chabris, 1999; Chun and Marois, 2002). Although inattentional blindness is typically studied in the laboratory by placing subjects in a predefined task, the simulations suggest that spontaneous trains of thought, unrelated to external stimuli and instructions, may also exert a temporary blocking. The model predicts that this state should be characterized by (1) an intense prefrontal–parietal–cingulate activation by the distracting thought or object prior to the presentation of the target stimulus; and (2) a proportional reduction of the target-induced activation to a brief bottom-up activation in specialized processors. fMRI studies provide direct support for those predictions (Rees et al., 1999; Weismann et al., 2005). Future research should extend those paradigms using time-resolved neuroimaging methods such as event-related potentials to

Author's personal copy

test the prediction that early bottom-up activation is preserved, but top-down recurrent reverberations are suppressed in an all-or-none manner.

1.38.6 Competition between Sensory Stimuli for Access to Consciousness: Looking for Objective Records of Subjective Perception

The model was also studied under conditions simulating a classical perceptual phenomenon referred to as the attentional blink where two sensory stimuli compete for access to reportable conscious perception (Raymond et al., 1992). In a typical experiment, participants are asked to process two successive targets, T1 and T2. When T2 is presented between 100 and 500 s after T1, the ability to report it drops, as if the participants' attention had 'blinked.' In other words, rather paradoxically, perception of a first visual stimulus may prevent the subjective perception of a second one. The paradigm is sufficiently simple and explicit to study why some patterns of brain activity have access to subjective experience and thus to establish, in a causal, mechanistic manner, a link between subjective reports and objective physiological recordings.

The minimal network proposed (Dehaene et al., 2003a; **Figure 3**) is composed of four hierarchical stages of processing where stimuli T1 and T2 evoke neuronal assemblies. At the lower level, A and B correspond to primary and secondary visual areas and C and D to higher association areas, including temporal and frontal cortex. At the lower level, the assemblies do not inhibit each other (see Arnell and Jolicoeur, 1997), but further on, T1 and T2 reach higher association areas C and D, where they compete for global access via reciprocal inhibitory interactions.

With the network placed in a regime of spontaneous thalamocortical oscillations, corresponding to a state of wakefulness (see earlier discussion), one distinguishes two principal modes of signal processing by the network. First, let us consider the simple case where T1, in the absence of competing stimuli, is consciously perceived. T1 evokes a short burst of phasic physiological activity that propagates across the A to D corticothalamic hierarchy. Reaching the highest cortical levels, the sensory input generates top-down amplification signals which, about 80 ms later, cause sustained firing in areas A and B. In a larger-scale simulation, such a long-lasting dynamic state would generate brain-scale propagation of stimulus information into the entire workspace network. It is proposed that this global broadcasting constitutes the physiological basis of conscious reportability.

The network also simulates the conditions of the 'blink' when T1 and T2 are presented in close succession. Experimentally it is known that when T2 is either presented simultaneously with T1, or long after, it is, in both cases, subjectively perceived. The network simulation indeed shows that in both conditions, sustained firing supported by joint bottom-up activation and top-down amplification takes place. Yet, under conditions of close temporal succession, a T2 stimulus presented during T1-elicited global firing elicits bottom-up activation restricted to levels A and B, but fails to propagate to higher cortical levels. As a result, the second phase of top-down amplification does not occur. The T2 stimulus is blinked from conscious perception (**Figure 5**).

The simulation predicts that a temporary drop in firing rate of pyramidal cells coding for T2 in areas C or D is associated with a loss in performance typical of the attentional blink. The model also shows a global drop of power emitted in the gamma band and of cross-correlations between distant T2-coding neurons. Thus, several indexes of firing and synchrony all point to a drop in global activity during the blink, particularly evident in the higher areas C and D.

An original property of the model is the distinction of two modes of signal processing – a nearest-neighbor bottom-up propagation of sensory stimulation across the hierarchy of areas and a long-distance top-down network that sends amplification signals back to all levels below it. In particular, it predicts a dynamic all-or-none bifurcation between the two modes associated with different subjective perception of the stimulus. Indeed, objective physiological data indicate that during this blink, T2 fails to evoke a P300 potential, but still elicits event-related potentials associated with visual and semantic processing (P1, N1, and N400) (Vogel et al., 1998).

The prediction of an all-or-none loss of conscious perception and of T2-induced higher-level brain-scale activation during the attentional blink was tested experimentally (Sergent and Dehaene, 2004; Sergent et al., 2005). We used a modified attentional blink paradigm in which human subjects merely had to report to what extent they had seen a word (T2) within a rapid letter stream that contained another target letter string (T1). To obtain a continuous measure of subjective perception, subjects were asked to move a

Author's personal copy

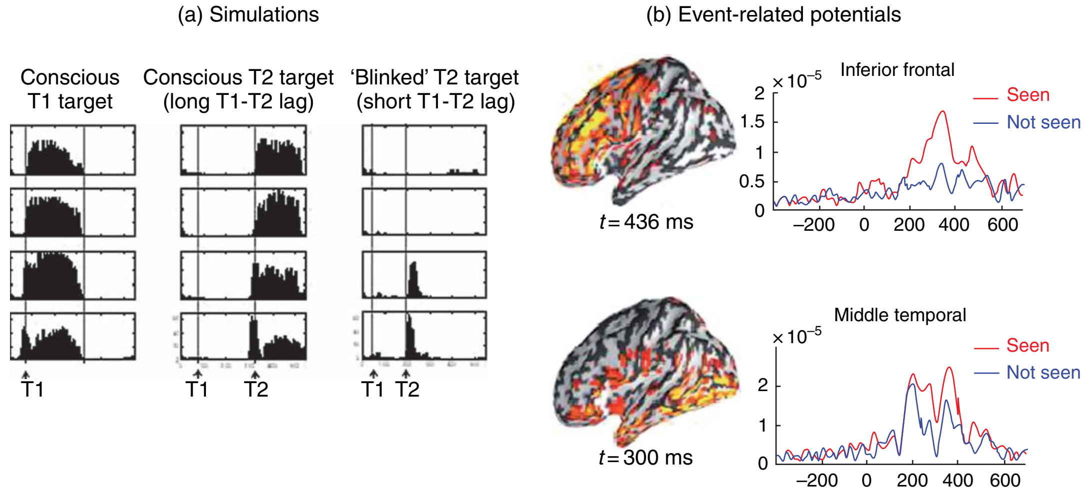


Figure 5 Comparison of the predictions of the neuronal workspace model (from Dehaene et al., 2003b) with the temporal dynamics of cortical electrical activity evoked by seen and unseen stimuli (from Sergent et al., 2005) during the attentional blink task. (a) Simulation of three trials of the attentional blink task. In each column with eight boxes (reproducing the superposed areas of **Figure 3(c)**) is shown the evolution of the computed firing rate in excitatory neurons. *Left column:* the sensory stimulus T1 accesses the workspace; *central column*: the lag between T1 and the second stimulus T2 is such that the sustained activity elicited by T1 has decayed in such a way that T2 accesses the conscious workspace; *right column*: the short lag between T2 and T1 is such that the sustained activity in T1 interferes with the access of T2 to the workspace. (b) Event-related potentials recorded during the attentional blink task at the level of the inferior frontal cortex where differences are noted between subjectively seen and not seen stimuli. Such difference is not observed in the initial bottom-up activation of the temporal cortex (200 ms) but converges later on during a global reverberation phase (300–400 ms). From Dehaene S, Sergent C, and Changeux JP (2003b) A neuronal network model linking subjective reports and objective physiological data during conscious perception. *Proc. Natl. Acad. Sci. USA* 100: 8520–8525, with permission from the National Academy of Sciences; and Sergent C, Baillet S, and Dehaene S (2005) Timing of the brain events underlying access to consciousness during the attentional blink. *Nat. Neurosci.* 8: 1391–1400, with permission from Nature Publishing Group.

cursor on a continuous scale, from 'not seen' on the left to 'maximal visibility' on the right. The results indicate that the reported subjective perception during the blink is indeed all-or-none (Sergent and Dehaene, 2004; Sergent et al., 2005) and relates to the loss of activation in a distributed, synchronous network prominently involving inferior and lateral prefrontal cortices as well as anterior cingulate (Marois et al., 2004; Gross et al., 2004; Sergent et al., 2005; Hommel et al., 2006).

The model in its simple formulation is coherent with previous proposals of a role of top-down recurrent (Lamme and Roelfsema, 2000), reentrant (Edelman, 1993), or resonant (Llinas et al., 1998) connections in the integrative processing of consciously perceived signals. It is also supported by recent experimental data. Because the blink is attributed to competition for workspace access, the proportion of T2 targets that are blinked, as a function of time, roughly traces the inverse shape of the neural activity evoked by T1 in higher-level areas C and D. In actual experiments, similarly, there is an inverse relation between the P300 waveform evoked by T1 and the size of the blink (Sergent et al., 2005). Furthermore, the fMRI activation elicited by T1 in parietal, frontal, and cingulate areas predicts the size of the blink (Marois et al., 2000). Last, the timing of the brain events during the attentional blink using letter strings was recently resolved by event-related potentials. The data show that early potentials (P1 and N1) were equally evoked by seen and not seen words, indicating that these early brain events do not fit with conscious perception. However, a rapid divergence was observed around 270 ms, with late brain events solely evoked by seen words (Sergent et al., 2005) (**Figure 5**). The data are thus fully consistent with the proposal of the model (Dehaene et al., 2003a) that top-down amplification signals and sustained firing into the workspace network constitutes the physiological basis of conscious reportability.

Although there have been no single-neuron recordings during the attentional blink, the simulated profiles of single-neuron activity to seen and blinked T2 targets can be compared to electrophysiological recordings obtained in other paradigms of conscious and unconscious processing. In perceptual areas A and B of the

Author's personal copy

model, neurons fire phasically in tight synchrony with the stimulus, then show a broader period of late amplification only in seen trials, not in blinked trials. This parallels experimental recordings in areas V1 and IT under conditions of inattention, reduced contrast, masking, or anesthesia, where late amplification occurs only for reportable stimuli (Lamme and Roelfsema, 2000; Super et al., 2001; Lamme et al., 2002). The presently available physiological data are thus consistent with the proposal of the neuronal workspace model that conscious access of reportable signals is a sudden self-amplifying bifurcation leading to a global brain-scale pattern of activity in the workspace network.

1.38.7 Preconscious States of Activity

Despite considerable progress in the empirical research on the brain imaging of conscious perception, debates have arisen about the coherence of these data. For instance, some researchers emphasize a correlation of conscious visual perception with early occipital events (Zéki, 2003), others with late parietofrontal activity (Sergent et al., 2005). Also, following Weiskrantz (1997), we insisted on the notion that subjective reports are the basic criterion that can establish whether a percept is conscious or not. Yet, the philosopher Ned Block (2005) has suggested that in reality, we may experience conscious experiences that are richer in content than what we can report. For instance, when an array of letters is flashed, viewers claim to see the whole array, although they can later report only one subsequently cued row or column. The initial processing of the array might already be considered as 'phenomenally' conscious though not 'seen' in a fully conscious manner (Lamme, 2003; Block, 2005).

We have expressed our disagreement with the phenomenal/access distinction, whose empirical testability is debatable, and have argued instead that within nonconscious processing, one must introduce a transient preconscious state of activity in which information is potentially accessible, yet not accessed (Dehaene et al., 2006). This led to the formal distinction (Dehaene et al., 2006) (**Figure 6**) within nonconscious information processing of:

(1) ***Subliminal processing*** of input signals that may occur when bottom-up activation is insufficient to trigger a large-scale reverberating state. The described simulations of a minimal thalamocortical network (Dehaene and Changeux, 2005) show that in a global network of neurons with long-range axons exhibiting nonlinear self-amplifying properties, a well-defined dynamic threshold exists beyond which activity quickly grows until a full-scale ignition is seen, while a slightly weaker activation quickly dies out.

(2) ***Preconscious processing***, a term coined to design a neural process that potentially carries enough activation for conscious access, but is temporarily buffered in a nonconscious store. Such a buffering might result

Subliminal

- Feedforward activation
- Activation decreases with depth
- Depth of processing depends on attention and task set
- Activation can reach semantic level
- Short-lived priming
- No durable frontoparietal activity
- No reportability

Preconscious

- Intense activation, yet confined to sensorimotor processors
- Occipitotemporal loops and local synchrony
- Priming at multiple levels
- No reportability while attention is occupied elsewhere

Conscious

- Orientation of top-down attention
- Amplification of sensorimotor activity
- Intense activation spreading to parietofrontal network
- Long-distance loops and global synchrony
- Durable activation, maintained at will
- Reportability

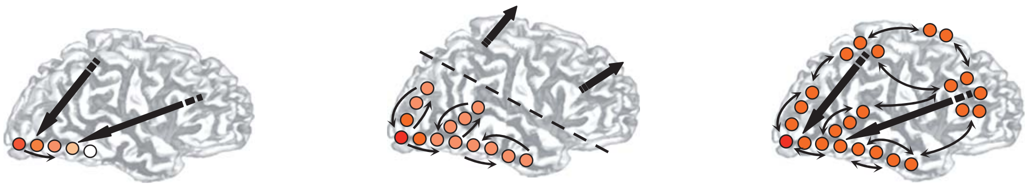


Figure 6 Schematic representation of the subliminal, preconscious, and conscious states of processing of visual stimuli. From Dehaene S, Changeux JP, Naccache L, Sackur J, and Sergent C. (2006) Conscious, preconscious, and subliminal processing: A testable taxonomy. *Trends Cogn. Sci.* 10: 204–211; used with permission from Elsevier.

Author's personal copy

from a lack of top-down attentional amplification, for example, owing to transient occupancy of the central workspace system (see preceding discussion and Dehaene and Changeux, 2005). The formal analysis of the attentional blink and inattentional blindness paradigms, indeed, has shown that even strong visual stimuli may remain temporarily preconscious. They are potentially accessible (they could quickly gain access to conscious report if they were attended), but they are not consciously accessed at the moment. At the neurocomputational level, preconscious processing is proposed to involve resonant loops within medium-range connections which maintain the representation of the stimulus temporarily active in a sensory buffer for a few hundred milliseconds.

In a fair attempt to establish objective recordings of subjectively reported conscious perception, these various conditions have to be experimentally and theoretically examined. In particular, other independent methods for decoding conscious states based, for instance, on trained pattern classifiers of fMRI or alternative physiological signals might be used to 'objectively' track signal processing in the course of conscious perception (Haynes and Rees, 2005, 2006), thus offering closer tests of the theoretical models.

1.38.8 Performance of an Effortful Deduction Task: The Stroop Task

Our modeling approach was developed along two distinct lines: *type 2* models were intended to capture some fine-grained physiological characteristics of neuronal firing trains and event-related potentials during conscious and subliminal perception. These simulations therefore necessarily incorporated considerably more physiological details of receptor types and cortical layers, but were not extended in a brain-scale architecture solving a precise task (e.g., Dehaene et al., 2003a; Dehaene and Changeux, 2005). On the other hand, the *type 1* models were intended to describe an entire task-related cognitive architecture and thus focused more on global connectivity than on fine physiological details (e.g., Dehaene et al., 1998). The Stroop task (McLeod, 1991) was selected as a simple experimental paradigm where a subject has to make a decision about the meaning of a written word under conditions where interference may occur. For instance, the subject is asked to give the color of the ink with which a color word is printed, and the meaning of the word may differ from the actual color of the printed word. Under this last condition, the subject has to make a conscious effort to give the correct response. The task was originally simulated with the standard neuronal workspace model (Dehaene et al., 1998), in which four input units were dedicated to encoding four color words, four other input units encoded the color of the ink used to print the word, and four internal units corresponded to the four naming responses. Routine color naming and word naming are implemented by direct one-to-one connections between these units and the corresponding output naming units. Workspace activation is not needed for any of these tasks. However, the effortful task (color naming with word interference) consists of providing conflicting word and color inputs and rewarding the network for turning on the naming unit appropriate to the ink color. When the naive network is switched to the effortful condition, an initial series of errors takes place as the network steadily applies the routine naming response. Yet, the delivery of negative reward leads to an increase in vigilance and to the sudden activation of variable patterns among workspace units resulting in a search phase for the next ∼30 trials. Workspace activation varies in a partly random manner as various response rules are explored, but the workspace activation patterns that lead to activating the incorrect response unit are negatively rewarded and tend to be eliminated in subsequent trials (**Figure 7**). Eventually, the network settles into a stable activation pattern, with a fringe of variability that slowly disappears in subsequent trials. This stable pattern, which leads to correct performance, is characterized by the differential amplification of the relevant word units relative to color units and by strong excitatory connections among active workspace units maintaining the pattern active in the intertrial interval.

Following the search phase, the network goes through a phase of effortful task execution in which workspace activation remains indispensable to correct performance. During this phase, workspace activity remains high, even on occasional trials in which the word and ink color information do not conflict. When performance is correct for a series of consecutive trials, vigilance tends to drop. However, any lapse in workspace activation is immediately sanctioned by an error. Each error is immediately followed by an intense reactivation of the workspace. Progressively, though, the task becomes routinized as the Hebbian rule applied to processor units tends to increase the color-to-name connections and to decrease the

Author's personal copy

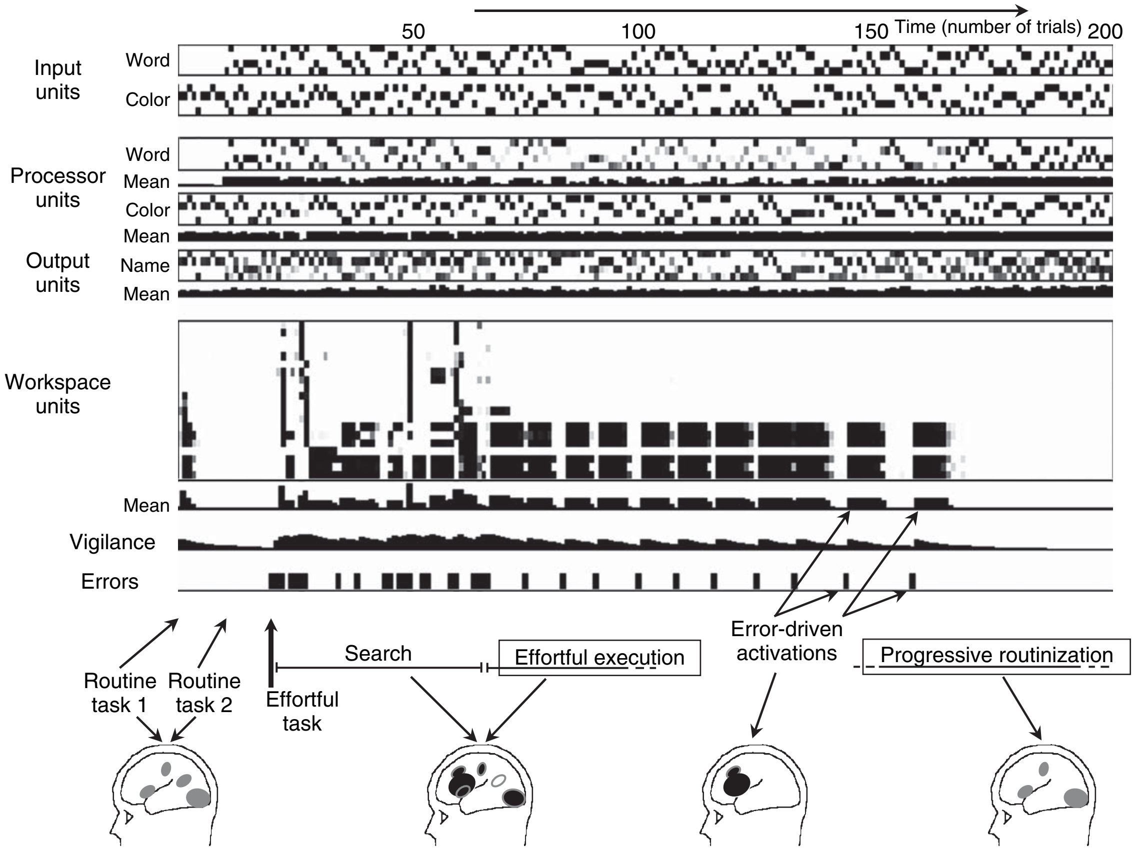


Figure 7 Simulation of the temporal dynamics of the Stroop task based on the original simulation of the neuronal workspace model of Dehaeneet al. (1998). The Stroop task was introduced without warning after routine trial no. 20. Note the selective activation of workspace units with a simultaneous amplification of color processors and a suppression of word processors. Workspace unit activation is seen in the initial phase of searching the appropriate response rule during the effortful execution of the task and following each erroneous response. Lower line: putative brain imaging correlates of workspace activation and routine are shown. From Dehaene S, Kerszberg M, and Changeux JP (1998) A neuronal model of a global workspace in effortful cognitive tasks. *Proc. Natl. Acad. Sci. USA* 95: 14529–14534; used with permission from the National Academy of Sciences.

word-to-name connections. Routinization is characterized by increasingly longer periods of correct performance in the absence of workspace activation.

The key empirical prediction of our hypothesis in the domain of brain imaging is the existence of a strong correlation between cortical areas that are found active in conscious effortful tasks and areas that possess a strong long-distance corticocortical connectivity (**Figure 7**). The global activation of neurons dispersed in multiple cortical areas is expected to be visualized as a temporary increase in the long-distance coherence of brain activity in electro- and magnetoencephalography or in studies of functional connectivity with fMRI.

The model also predicts that areas rich in workspace neurons will appear as suddenly activated when a novel, nonroutine task is introduced while being absent during routine tasks and will vary semirandomly during the initial learning of a novel task. The level of activation should be high and stable during execution of a known but not yet routinized effortful task and should decrease during routinization, but should resume sharply following an error.

Brain-imaging experiments indicate that the workspace network which includes dorsolateral prefrontal cortex (dlPFC) and anterior cingulate (AC) is active in effortful cognitive tasks, including the Stroop test, with a graded level of activation as a function of task difficulty (Pardo et al., 1990; Cohen et al., 1997; Paus et al., 1998). With automatization, activation decreases in dlPFC and AC, but it immediately recovers if a novel, nonroutine situation occurs (Raichle, 1994). AC activates in tight synchrony with subjects' errors (Dehaene and

Author's personal copy

Cohen, 1994; Carter et al., 1998). Data consistent with the model have been recently obtained with a rather elaborate reward-based logical deduction task referred to as 'master-brain' task (Landmann et al., 2006). In the course of this trial-and-error learning process, the subjects have to infer the identity of an unknown four-key code on the basis of successive feedback signals but also through internal autoevaluation deductions. This search period is followed by a routine period during which subjects merely repeatedly execute the identified sequence. fMRI measurements reveal a sudden activation during search of the expected workspace circuits which include bilateral frontoparietal (particularly the lateral orbitofrontal and dorsolateral prefrontal cortex) and anterior cingulated cortex, striatum, and midbrain together with the cerebellum (**Figure 8**). This activation collapses during ensuing periods of routine sequence repetition. Furthermore, examination of brain activation during logical and chance discovery showed an early collapse when the correct logical sequence could be deduced without waiting to receive an actual external reward. In agreement with the simulations of the neuronal workspace model in the case of the Stroop task (Dehaene et al., 1998), the data reveal large-scale changes in interconnected distant cortical areas, which may take place tacitly according to an autoevaluation process. They add an important aspect, the differential activation of the striatum: The activation of the putamen and a part of the right caudate body indexes the sign of the reward prediction error (referred to as a ventral 'critic'), and the head of the right caudate indexes the available information that can be extracted from it (referred to as a dorsal 'actor') (see Sutton and Barto, 1981; Schultz et al., 1997; O'Doherty, 2004). This relationship between the workspace network and the underlying subcortical reward networks,

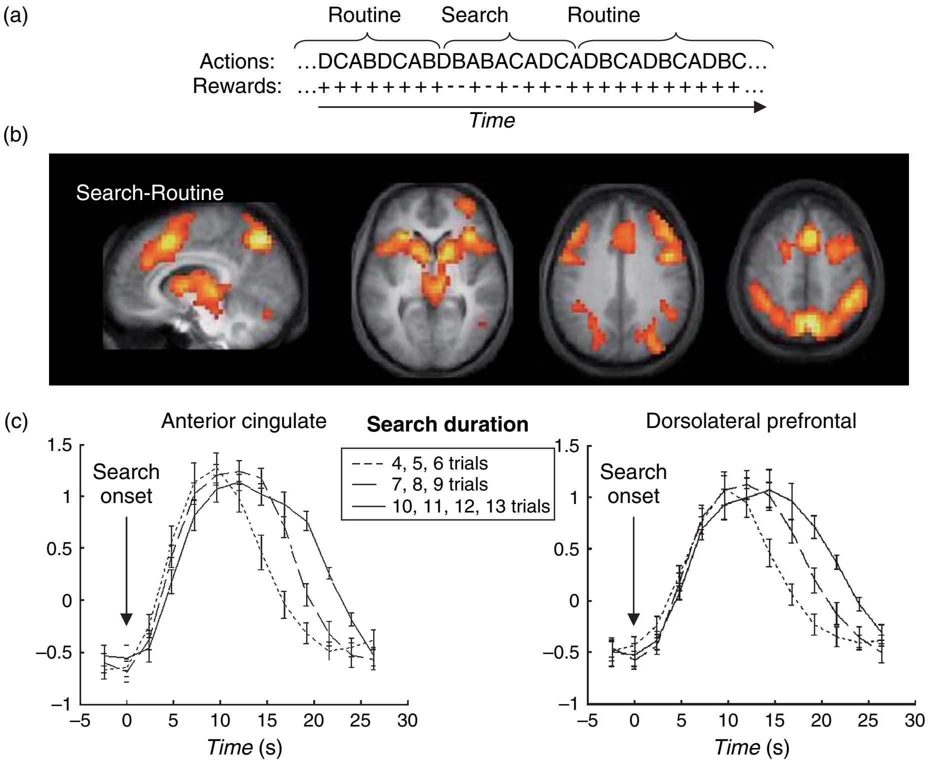


Figure 8 Brain imaging fMRI data of the dynamics of prefrontal and cingulate activity during a reward-based logical deduction task (Landmann et al., 2006). (a) Subjects were engaged in a motor trial-and-error learning task, in which they had to infer the identity of a hidden four-key press by trial-and-error (here ADBC). The design of the task allowed subjects to base their inferences not only on the feedback they received but also on internal deductions and evaluations. (b) fMRI imaging revealed, in agreement with the neuronal workspace model (**Figure 7**), a large bilateral activation of parietal, prefrontal, cingulate, and striatal networks during the search period that collapsed during ensuing routine execution. From Landmann C, Dehaene S, Pappata S, et al. (2006) Dynamics of prefrontal and cingulate activity during a reward-based logical deduction task. *Cereb. Cortex* 17: 749–759; used with permission from Oxford University Press.

Author's personal copy

which was not taken into consideration in the detailed model of access to consciousness, becomes accessible on both theoretical and experimental grounds.

1.38.9 The Evolution of Consciousness

1.38.9.1 Animal and Human Consciousness

Since the provocative statements of Thomas Huxley that "we are conscious automata," that "brutes" share consciousness with humans (1874) and that "all states of consciousness in us, as in them, are immediately caused by molecular changes of brain substance," nonhuman species and especially laboratory animals have served as experimental models for the scientific investigation of behavior but also of animal consciousness (Thorndike, 1898; Yerkes, 1916; Barresi and Moore, 1996; Jasper, 1998; Koch, 2004; Changeux, 2006). In the framework of the present discussions about neurocomputational models of consciousness, a first challenging issue is thus to what extent the neuronal workspace model may be usefully exploited to define and evaluate consciousness in animals. A second related question is whether or not functionally homologous (rather than analogous) neural structures might be at work in these nonhuman species. Answers to both questions are of importance, in particular if one wishes to use laboratory animals like the mouse as experimental models to investigate the neural bases of consciousness (see Changeux, 2006).

As early as 1921, the Italian neurologist Luigi Bianchi stated that "among the phenomenal factors of the activities of living organisms" arises a "bond of coherence" which "progresses with the development and complexity of living organisms and of their nervous system" that he referred to as consciousness. He further specified that "the dawn of higher consciousness coincides with the apparition of the frontal lobes in the evolution of the brain" and also noted their "inhibitory power" and their capacity for "intellectual syntheses." Bianchi in many respects anticipated several of the presently suggested ideas and theories about consciousness (Changeux, 1983; Shallice, 1988; Edelman, 1989; Crick, 1994; Crick and Koch, 2003; Dehaene and Changeux, 2004; Changeux and Edelstein, 2005; Naccache, 2005; and Naccache et al., 2005, among many others). Following these views, one may then argue that lower species like mice, or even birds, which have a reduced or even absent prefrontal cortex – a critical component of the neuronal workspace circuits – have little if any consciousness. Would experimental and theoretical investigations with the laboratory mouse be simply irrelevant to a scientific investigation of consciousness (see Block, 2005)?

It is of interest to reevaluate these issues in the framework of the neuronal workspace model with nonhumans as well as with humans in the course of both ontogenetic and phylogenetic evolutions. A first point to note is that looking at the developing human infants, Preyer already stated in 1894, "There is not the least reason for assuming in advance that every human being comes into the world endowed with complete consciousness of self" and "there are several grades of consciousness." Developmental studies with humans from the newborn to the adult (Zelazo, 1996; Lagercrantz et al., 2002; Zelazo, 2004; Johnson, 2005; Lagercrantz, 2005; Bartocci et al., 2006) reveal without ambiguity that, beyond its diverse definitions, consciousness cannot be viewed as an irreducible and unique global entity. A similar conclusion emerges from ethological and experimental studies with evolutionarily distant species, such as mice, monkeys, and apes (see Boakes, 1984; Trivers, 1985; Barresi and Moore, 1996; Changeux, 2002). Comparative analysis of these systems suggests a breakdown of consciousness into multiple nested hierarchical levels. Here, we tentatively propose a first classification into four levels, all the while realizing that it remains necessarily arbitrary and simplistic (see Changeux, 2006). Note that these represent landmark points on a continuum rather than sharp distinctions, because our goal is to emphasize a continuity of phylogenetic and ontogenetic stages to full-blown adult human consciousness. The proposed levels comprise:

a. A lower level of minimal consciousness for simple organisms, like rats or mice, which undergo cycles of sleep and wakefulness, possess the capacity to display spontaneous motor activity and to create representations, for instance from visual and auditory experience, to store them into long-term memory and use them for approach and avoidance behavior; these organisms exhibit what is referred to as exploratory behavior (see Thinus-Blanc et al., 1996; Granon et al., 2003); they are amenable, as humans, to trace conditioning (which in humans requires awareness) and to delay conditioning (which does not) (Han et al., 2003). These organisms do not make reference to an *explicit* sense of self and display

Author's personal copy

minimal social interactions.

In humans, the 25- to 30-week preterm fetus processes tactile and painful stimuli in the sensory cortex (see brain imaging studies by Bartocci et al., 2006) and might thus perceive pain when awake; he/she might have reached a stage of brain maturation analogous (though not identical) to that of a newborn rat/mouse (see Lagercrantz, 2005).

b. Basic consciousness, present, for instance, in vervet monkeys (possibly also in some birds), manifests itself by functional use of objects, protodeclarative pointing, and searching for hidden objects; organisms at this level may display elaborate social interactions, imitations, social referencing, and joint attention; they possess the capacity to hold several mental representations in memory simultaneously and are able to evaluate relations of others to self.

In humans, newborn infants exhibit, in addition to sensory awareness, the ability to process memorized mental representations (e.g., of a pacifier), to express emotions, and to show signs of shared feelings or empathy (Singer, 2006). Even newborns differentiate between self and nonself touch (Rochat, 2003) and imitate the tongue protrusion of an adult (Meltzoff, 1990). At a few months of age, responses to novelty are present and include a late negativity wave which has been tentatively assigned to prefrontal cortex (Dehaene-Lambertz and Dehaene, 1994; Reynolds and Richards, 2005). Prefrontal cortex is active in response to speech, but only in awake infants, not when asleep (Dehaene-Lambertz et al., 2002). Thus access of sensory information to consciousness may already be present, though capacities for internal manipulations in working memory are reduced if not absent.

c. Explicit self-consciousness develops in infants at the end of the second year, together with working and episodic memory and language; it is characterized by self-recognition in mirror tests and by the use of single arbitrary rules with knowledge of one's own behavioral potential and self–other distinction; to some extent chimpanzees might reach this level (see Boakes, 1984).

d. Reflective consciousness with theory of mind and full conscious experience, first-person ontology, and explicit report, is unique to humans and develops after 3–5 years in children.

Examination of this first preliminary classification leads to a simple and unambiguous conclusion. First, adult human consciousness develops progressively, starting from rather rudimentary dispositions in the newborn. Second, mice and rats do not go far beyond the level of minimal consciousness.

1.38.9.2 Minimal Consciousness in Mice and Rats

Careful examination of mouse and rat behavior (see Brown and Bowman, 2002; Granon et al., 2003; Han et al., 2003) in the context of the neuronal workspace and as possible models of human psychiatric disorders (see Granon and Changeux, 2006) leads to the following conclusions.

a. Multiple states of consciousness (such as wakefulness, sleep, coma, general anesthesia, epileptic seizures) and the regulation of their reversible transitions occur in the mouse, as in all mammals, including humans. The circadian sleep–waking cycle mobilizes rather universal mechanisms (Llinas and Steriade, 2006) and is controlled throughout vertebrate species by brainstem reticular formation and intralaminar nuclei of the thalamus (Bogen, 1995; Jones, 1998) with complex patterned releases of neuromodulatory substances. For instance, in the framework of the current studies on nicotinic acetylcholine receptors (nAChR) (Changeux and Edelstein 2005), gene inactivation studies demonstrate the positive contribution of the $\beta 2$ subunit of the nAChR to the phasic expression of arousal promoting mechanisms by endogenous acetylcholine (Cohen et al., 2002; Léna et al., 2004), and similar phenotypes as those noticed in the knockout mice are observed after chronic in utero exposure of the fetus to nicotine (Cohen et al., 2005). These mice thus offer a plausible animal model of sudden infant death syndrome, whose prevalence is known to increase in smoking pregnant women (Cohen et al., 2005).

The use of the neuronal workspace formalism (see Dehaene and Changeux, 2005) might, in this respect, be of some help. One may, for example, view the various graded states of consciousness (from deep anesthesia and coma to full awareness) as directly related to the spontaneous activity of recurrent thalamocortical loops and reticular thalamic nuclei (see Llinas and Paré, 1991; Steriade et al., 1993) as described in the model. Conclusively, the description of the neural mechanisms involved in the transition to wakefulness appears relevant to the case of the mouse as well as its application to models of human pathologies.

Author's personal copy

b. Delayed-response tasks, exploratory behavior, and flexible goal-directed behaviors. Since the early 1990s (Kolb, 1990), a rich body of behavioral and pharmacological observations has revealed that rats display working memory, attentional processes, and flexible goal-directed behavior which rely on the contribution of the prefrontal cortex (Kolb, 1990). For example, rats may perform an effortful, counterinstinctive, delayed-response task referred to as delayed matching-to-sample that lesions of the prefrontal cortex selectively impair, at variance with a spontaneous delayed nonmatching-to-sample task (Granon et al., 1994). Moreover, in these pioneering studies, it was already shown that cholinergic pathways (Nordberg and Winblad, 1986; Levin, 1992) and specifically their nicotinic component selectively control these cognitive processes (injection of neuronal bungarotoxin into the prelimbic area of the prefrontal cortex selectively impairs the task) (Granon et al., 1995).

In the mouse, exploratory activity is a spontaneous behavior (Thinus-Blanc et al., 1996; Poucet and Herrmann, 2001) that serves to gather and store spatial information which allows allocentric coding of space, itself necessary for flexible navigational processes. Quantitative analyses (Faure et al., 2003) with mice lacking the high-affinity nAChR further reveal that the balance between navigation and exploration shifts in favor of navigation to the detriment of more precise exploration of the environment. Additional comparative studies including several objects presentations in an open-field arena and elementary social behavior (between a test resident mouse and a social intruder) revealed that mice deleted for the nAChR $\beta 2$ subunit are more rigid and exhibited less behavioral flexibility than wild-type mice (Granon et al., 2003; Maskos et al., 2005). The prefrontal cortex is particularly reduced in size in the mouse, yet its lesions (Granon, unpublished) cause evident deficit in the aforementioned conflict-resolution situations and, moreover, create a behavioral phenotype which displays several features in common with the loss of $\beta 2$-nAChR.

These studies unambiguously demonstrate: (1) the specialization of neural circuits engaged in such executive functions which most often (though not always) mobilize the prefrontal cortex; (2) the gating of these functions by nAChRs activated through endogenously released acetylcholine; and last, (3) the intimate relationships between reward and cognition evidenced by the joint recovery of exploratory behavior and reward functions by the targeted reexpression of the $\beta 2$ subunit in the ventral tegmental area dopaminergic nucleus (Maskos et al., 2005).

In conclusion, mice under these experimental conditions do far more than to simply react to sensory information. They engage in complex extended behaviors geared toward far-removed goals and sensitive to rewards. Using the word of Denton (2005), they would display some kind of curiosity being able to orchestrate locomotor behaviors according to what might tentatively be named – at our own risks – conscious intentions. Moreover most, if not all, of all these behaviors need the integrity of the prefrontal cortex. Even if in the case of rats and mice the exact homologies with primates and humans prefrontal cortex areas are still debated (Brown and Bowman, 2005), one may – still hypothetically – propose that the flexible goal-directed behaviors examined with the mouse fall into the category of conscious processes described by the neuronal workspace model.

The model was primarily designed to account for access to consciousness and reportability (see Dehaene et al., 1998). Reporting responses have been demonstrated with some animal species, for instance, using a commentary key (Weiskrantz, 1991; Cowey and Stoerig, 1995) in the case of macaque blind-sight. Yet, a still unanswered question is whether or not reportability can be demonstrated in a species like the mouse. In any case, the invention of a reliable assay with this species is urgently needed. In conclusion, the neuronal workspace model obviously deals with important features of *minimal consciousness*. It accounts, for instance, for the active maintenance of abstract rules through top-down amplification and the flexible control of tasks that require a novel interconnection of existing processors as it typically occurs in the Stroop task (see earlier discussion) as well as with the aforementioned mice behaviors. It deals, *a fortiori*, with the active maintenance of information during a delay period (see Han et al., 2003; Koch, 2004). Even though the relevant simulations have not been carried out, the neuronal workspace architecture should adequately fit exploratory behavior and offer an appropriate mechanism for the ultimate stage of spatial processing introduced by Poucet's model (1993) where, in a workspace homologue, unified location-independent representations with one unique reference direction are being built.

Author's personal copy

1.38.9.3 Social Relationships and Consciousness

The hierarchical scale of levels of consciousness mentioned at the beginning of this section underlines the importance of social interactions to the extent that empathy was viewed as a characteristic feature of newborn consciousness. Social organisms, including humans, represent intentional relations of themselves and other agents, yet at different levels. They unambiguously distinguish their own intentional relations (or first-person information) from the qualitatively different information available about other agents' intentional relations (or third-person information). In this respect, one should remember that the analysis of the exploratory behavior leads to the distinction between allocentric and egocentric motor behavior in the mouse (Rondi-Reig et al., 2006), pointing to the still highly speculative occurrence of a self. In the mouse, such self is primarily oriented toward the outer physical world, though empathy to pain has been recently reported in the mouse (Langford et al., 2006). No evidence was found at this stage with the mouse for comparability between the actions of self and others, as is found in higher species, and no sign of imitating goal-directed activity or of understanding the viewpoints of others was observed (Barresi and Moore, 1996). In other words, the presently available evidence does not support the occurrence of authentic social relationships in the adult mouse.

The human infant at birth is already at a stage more advanced than the adult mouse; in this respect, as mentioned earlier, he or she may already distinguish between his own and others' movements, in particular by touch, and the newborn displays rudiments of imitations (Meltzoff and Gopnik, 1993; Barresi and Moore, 1996; Lagercrantz, 2005). Moreover, human neonates display emotional contagion by responding more with crying when hearing another newborn crying than when hearing white noise or their own cry (see Decety and Jackson, 2004). Thus, the mouse cannot be a good animal model to investigate intentional relations and social understanding whose highest level is reached exclusively in humans. On the other hand, it may serve as *baseline* to define the elementary neural circuits mobilized by these social relationships in higher mammals and humans.

Extension of the neuronal workspace model to these issues may help in the definition of the minimal components of neural networks able to simulate what may be referred to as social consciousness. One has first to realize that several successive hierarchical levels have to nest the basic states of consciousness of the newborn to reach the full reflective consciousness of the human adult. They include imitation, social referencing, and joint attention, but also what is referred to as the standard theory of mind (see Premack and Woodruff, 1978; Baron-Cohen et al., 2000; Frith and Frith, 2003; Gusnard, 2005). This disposition to represent other people's intentions and beliefs, commonly referred to as propositional attitudes, mobilizes circuits distinct from empathy (Singer, 2006). It develops relatively late in the child (4–5 years), long after empathy. Moreover, both empathy and mentalizing are the objects of a severe maturation before the child reaches the stage of reflective consciousness; it includes in particular the general use of symbols (linguistic or not). Interestingly, this evolution through childhood and adolescence is accompanied by a nonlinear loss of gray matter in the cerebral cortex linked to the selective stabilization (pruning) of synapses (Giedd et al., 1999a,b; Singer, 2006), which takes place during postnatal development (see Changeux, 1983, 2004).

Considerable work has to be developed to establish a useful match between an extended neuronal workspace model and reflective consciousness.

1.38.9.4 The Neuronal Workspace and Human Pathologies

On the clinical side, the neuronal workspace model offers simple interpretations of a variety of human pathologies which cannot be reviewed here. For instance, the neuronal workspace model may account for characteristic deficits caused by frontal lobe lesions in performing delayed-response tasks like the Wisconsin card sorting task (Dehaene and Changeux, 1991) or the Stroop task (see the first simulations of the model; Dehaene et al., 1998) and/or in working memory or declarative memory tasks (Squire, 1987–1988; Ungerleider, 1995; Naccache, 2005; Zeman, 2005).

Frontal lobe pathology is also associated with senescence and dementia (Parkin and Walter, 1992). Particularly relevant to the theory is the case of frontal lobe dementia. This degenerative disease is characterized by apathy, unconcern, disinhibition, distractibility, loss of social awareness, and loss of emotional empathy (Brun, 1987; Neary et al., 1988; Baker et al., 2006). Interestingly, it can be caused by either a mutation in the microtubule-associated protein tau at Chr17q21 (characterized by cytoplasmic neurofibrillary inclusions) or a null mutation in the gene of a growth factor, progranulin, at the Chr17q21

Author's personal copy

31 locus (characterized by ubiquitin–immunoreactive neuronal inclusions) (Baker et al., 2006; Mackenzie et al., 2006). It manifests itself by a selective loss of layers 2 and 3 pyramidal cells of the prefrontal cortex, the long axon neurons which were postulated as the basic anatomical components of the workspace circuits (Dehaene et al., 1998). In other words, this disease offers a striking example of a genetic dissection of the neuronal workspace in the adult human brain.

Impairments at the level of workspace neurons might also shed some light on the cognitive deficits underlying psychiatric diseases such as schizophrenia. Indeed, cognitive deficits in schizophrenia often affect a broad variety of cognitive tests, and thus may fit better within the present perspective than within the classical neuropsychological perspective, whereby an individual patient's deficits are explained by a local impairment within a modular architecture of specialized subsystems. Many neuroimaging studies suggest decreased frontal and anterior cingulate activation in schizophrenia, as well as decreased long-distance connectivity (Andreasen et al., 1997; Friston, 1998). Furthermore, a dissociation between preserved subliminal processing and impaired conscious access has been reported: the threshold for masking is systematically elevated in schizophrenia, and preserved visual, semantic, and even motor priming suggest that this deficit is due to a central integration impairment, not a basic sensory impairment (Dehaene et al., 2003; Del Cul et al., 2006). Interestingly, similar deficits of access to consciousness are also seen in patients with early multiple sclerosis (MS) and diffuse white matter damage (Reuter et al., 2007). The parallels between MS and schizophrenia, and the conceptualization of at least part of their cognitive deficits as affecting a global workspace for flexible conscious processing, offer interesting avenues for future research.

Another consequence of the workspace theory is that it leads to a plausible interpretation of drug (e.g., nicotine) addiction. Addiction may indeed be viewed as an escape from the voluntary control of drug taking behavior, for instance, as a consequence of the disconnection of a reciprocal-loop linking the neuronal workspace circuits, including prefrontal cortex, dopaminergic neurons, and striatum, thus uncovering the compulsive nonconscious aspect of drug addiction (for discussion see Gutkin et al., 2006).

Moreover, as noted earlier, the nicotinic receptor knockout mice which are compulsively navigating without pausing for exploration may offer an animal model for human attention-deficit/hyperactivity disorder behavior, for which hyperactivity symptoms are known to improve with nicotine treatment (Shytle et al., 2002; Granon and Changeux, 2006).

Last, the differential role of the ventral tegmental area in the recovery of some aspects of cognitive functions in the mouse by local reexpression of nicotinic receptors (Maskos et al., 2005) points to an analogy with the human disease called auto-activation deficit described by Laplane and Dubois (2001). Human patients display a characteristic inertia—they stay at the same place all day long without signs of spontaneous activity but may perform complex activities when stimulated. They show an empty mind for hours yet without cognitive impairment but with stereotyped activities and flattened affects. Their deficit is caused by striatopallidal lesions accompanied by frontal hypometabolism, suggesting, as in the case of the mouse model, a close link between reward and, here, the content of consciousness.

These are a few examples of human pathologies in which the Neuronal Workspace model offers simple and productive interpretations.

References

Andreasen NC, O'Leary DS, Flaum M, et al. (1997) Hypofrontality in schizophrenia: Distributed dysfunctional circuits in neuroleptic-naive patients. *Lancet* 349(9067): 1730–1734.

Amati D and Shallice T (2007) On the emergence of modern humans. *Cognition* 103(3): 358–385.

Arieli A, Sterkin A, Grinvald A, and Aertsen A (1996) Dynamics of ongoing activity: Explanation of the large variability in evoked cortical responses. *Science* 273: 1868–1871.

Arnell KM and Jolicoeur P (1997) Repetition blindness for pseudoobject pictures. *J. Exp. Psychol. Hum. Percept. Perform.* 23: 999–1013.

Baars B (1989) *A Cognitive Theory of Consciousness*. Cambridge, UK: Cambridge University Press.

Baker M, Mackenzie IR, Pickering-Brown SM, et al. (2006) Mutations in progranulin cause tau-negative frontotemporal dementia linked to chromosome 17. *Nature* 442: 916–919.

Barresi J and Moore C (1996) Intentional relation and social understanding. *Behav. Brain Sci.* 19: 107–154.

Balkin TJ, Braun AR, Wesensten NJ, et al. (2002) The process of awakening: A PET study of regional brain activity patterns mediating the re-establishment of alertness and consciousness. *Brain* 125: 2308–2319.

Baron-Cohen S, Wheelwright S, Cox A, et al. (2000) Early identification of autism by the Checklist for Autism in Toddlers (CHAT). *J. R. Soc. Med.* 93: 521–525.

Bartocci M, Bergqvist LL, Lagercrantz H, and Anand KJ (2006) Pain activates cortical areas in the preterm newborn brain. *Pain* 22: 109–117.

Bianchi L (1921) *La mécanique du cerveau et la fonction des lobes frontaux*. Paris: Louis Arnette.

Block N (2005) Two neural correlates of consciousness. *Trends Cogn. Sci.* 9: 46–52.

Author's personal copy

Boakes R (1984) From Darwin to behaviourism. In: *Psychology and the Minds of Animals*. Cambridge, UK: Cambridge University Press.

Bogen JE (1995) On the neurophysiology of consciousness: I. An overview. *Conscious Cogn.* 4: 52–62.

Bonhomme V, Plourde G, Meuret P, and Fiset P (2000) Auditory steady-state response and bispectral index for assessing level of consciousness during propofol sedation and hypnosis. *Anesth. Analg.* 91: 1398–403.

Brecht M, Roth A, and Sakmann B (2003) Dynamic receptive fields of reconstructed pyramidal cells in layers 3 and 2 of rat somatosensory barrel cortex. *J Physiol.* 553: 243–265.

Brown VJ and Bowman EM (2002) Rodent models of prefrontal cortical function. *Trends Neurosci.* 25: 340–343.

Brun A (1987) Frontal lobe degeneration of non-Alzheimer type I. Neuropathology. *Arch. Gerontol. Geriatr.* 6: 193–208.

Bush P and Sejnowski T (1996) Inhibition synchronizes sparsely connected cortical neurons within and between columns in realistic network models. *J. Comput. Neurosci.* 3: 91–110.

Buchsbaum MS, Gillin JC, Wu J, et al. (1989) Regional cerebral glucose metabolic rate in human sleep assessed by positron emission tomography. *Life Sci.* 45: 1349–1356.

Cajal S (1892) El nuevo concepto de la histologia de los centros nerviosos. *Rev. Ciencias Med.* 18: 457–476.

Carter JA, McNair LD, Corbin WR, and Black DH (1998) Effects of priming positive and negative outcomes on drinking responses. *Exp. Clin. Psychopharmacol.* 6: 399–405.

Changeux JP (1983) *L'homme neuronal*. Paris: Fayard.

Changeux JP (2002) Reflections on the origins of the human brain. In: Lagerkrantz H, Hanson M, Evrard P, and Rodeck C (eds.) *The Newborn Brain*, pp. 1–28. Cambridge, UK: Cambridge University Press.

Changeux JP (2004) *The Physiology of Truth*. Cambridge, MA: Harvard University Press.

Changeux JP (2006) The Ferrier Lecture 1998. The molecular biology of consciousness investigated with genetically modified mice. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 361: 2239–2259.

Changeux JP, Courrege P, and Danchin A (1973) A theory of the epigenesis of neuronal networks by selective stabilization of synapses. *Proc. Natl. Acad. Sci. USA* 70: 2974–2978.

Changeux JP and Dehaene S (1989) Neuronal models of cognitive functions. *Cognition* 33: 63–109.

Changeux JP and Edelstein SJ (2005) Allosteric mechanisms of signal transduction. *Science* 308: 1424–1428.

Christakou A, Robbins TW, and Everitt BJ (2004) Prefrontal cortical-ventral striatal interactions involved in affective modulation of attentional performance: Implications for corticostriatal circuit function. *J. Neurosci.* 28: 773–780.

Chun MM and Marois R (2002) The dark side of visual attention. *Curr. Opin. Neurobiol.* 12(2): 184–189.

Cohen G, Han ZY, Grailhe R, et al. (2002) Beta-2 nicotinic acetylcholine receptor subunit modulates protective responses to stress: A receptor basis for sleep-disordered breathing after nicotine exposure. *Proc. Natl. Acad. Sci. USA* 99: 13272–13277.

Cohen G, Malcolm G, and Henderson-Smart D (1997) A comparison of the ventilatory response of sleeping newborn lambs to step and progressive hypoxaemia. *J. Physiol.* 503: 203–213.

Cohen G, Roux JC, Grailhe R, Malcolm G, Changeux JP, and Lagercrantz H (2005) Perinatal exposure to nicotine causes deficits associated with a loss of nicotinic receptor function. *Proc. Natl. Acad. Sci. USA* 102: 3817–3821.

Cowey A and Stoerig P (1995) Blindsight in monkeys. *Nature* 373: 247–249.

Crick F (1994) On consciousness. *Nature* 369: 86.

Crick FC and Koch C (1995) Are we aware of neural activity in primary visual cortex? *Nature* 375: 121–123.

Crick FC and Koch C (2003) A framework for consciousness. *Nat. Neurosci.* 6: 119–126.

Crick FC and Koch C (2005) What is the function of the claustrum? *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360: 1271–1279.

Decety J and Jackson PL (2004) The functional architecture of human empathy. *Behav. Cogn. Neurosci. Rev.* 3: 71–100.

Dehaene S, Artiges E, Naccache L, et al. (2003a) Conscious and subliminal conflicts in normal subjects and patients with schizophrenia: The role of the anterior cingulate. *Proc. Natl. Acad. Sci. USA* 100: 13722–13727.

Dehaene S and Changeux JP (1989) A simple model of prefrontal cortex function in delayed-response tasks. *J. Cogn. Neurosci.* 1: 244–261.

Dehaene S and Changeux JP (1991) The Wisconsin Card Sorting Test: Theoretical analysis and modeling in a neuronal network. *Cereb. Cortex* 1: 62–79.

Dehaene S and Changeux JP (1995) Neuronal models of prefrontal cortical functions. *Ann. N.Y. Acad. Sci.* 769: 305–319.

Dehaene S and Changeux JP (1997) A hierarchical neuronal network for planning behavior. *Proc. Natl. Acad. Sci. USA* 94: 13293–13298.

Dehaene S and Changeux JP (2000) Reward-dependent learning in neuronal networks for planning and decision making. *Prog. Brain Res.* 126: 217–229.

Dehaene S and Changeux JP (2004) Neural mechanisms for access to consciousness. In: Gazzaniga MS (ed.) *The Cognitive Neuroscience of Consciousness*, pp. 1145–1158. Cambridge, MA: MIT Press.

Dehaene S and Changeux JP (2005) Ongoing spontaneous activity controls access to consciousness: A neuronal model for inattentional blindness. *PLoS Biol.* 3: 910–927.

Dehaene S, Changeux JP, Naccache L, Sackur J, and Sergent C (2006) Conscious, preconscious, and subliminal processing: A testable taxonomy. *Trends Cogn. Sci.* 10: 204–211.

Dehaene S and Cohen L (1994) Dissociable mechanisms of subitizing and counting: Neuropsychological evidence from simultanagnosic patients. *J. Exp. Psychol. Hum. Percept. Perform.* 20: 958–975.

Dehaene S, Kerszberg M, and Changeux JP (1998) A neuronal model of a global workspace in effortful cognitive tasks. *Proc. Natl. Acad. Sci. USA* 95: 14529–14534.

Dehaene S and Naccache L (2001) Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework. *Cognition* 792: 1–37.

Dehaene S, Sergent C, and Changeux JP (2003b) A neuronal network model linking subjective reports and objective physiological data during conscious perception. *Proc. Natl. Acad. Sci. USA* 100: 8520–8525.

Dehaene-Lambertz G and Dehaene S (1994) Speed and cerebral correlates of syllable discrimination in infants. *Nature* 370: 292–295.

Dehaene-Lambertz G, Dehaene S, and Hertz-Pannier L (2002) Functional neuroimaging of speech perception in infants. *Science* 298: 2013–2015.

DeFelipe J and Farinas I (1992) The pyramidal neuron of the cerebral cortex: Morphological and chemical characteristics of the synaptic inputs. *Prog. Neurobiol.* 39: 563–607.

DeFelipe J and Jones EG (1988) *Cajal on the Cerebral Cortex*. New York: Oxford University Press.

Delacour J (1997) Neurobiology of consciousness: An overview. *Behav. Brain Res.* 85: 127–141.

Del Cul A, Dehaene S, and Leboyer M (2006) Preserved subliminal processing and impaired conscious access in schizophrenia. *Arch. Gen. Psychiatry* 63(12): 1313–1323.

Author's personal copy

de Lecea L, Kilduff TS, Peyron C, et al. (1998) The hypocretins: Hypothalamus-specific peptides with neuroexcitatory activity. *Proc. Natl. Acad. Sci. USA* 6: 322–327.

Denton D (2005) *Les Emotion primordiales et l'éveil de la conscience.* Paris, France: Flammarion. [Translated 2005, *The Primordial Emotions: The Dawning of Consciousness*, Oxford, UK: Oxford University Press.]

Dosenbach NU, Visscher KM, Palmer ED, et al. (2006) A core system for the implementation of task sets. *Neuron* 50: 799–812.

Douglas RJ and Martin KA (2004) Neuronal circuits of the neocortex. *Annu. Rev. Neurosci.* 27: 419–451.

Edelman GM (1989) *The Remembered Present: A Biological Theory of Consciousness*. New York: Basic Books.

Edelman GM (1993) Neural Darwinism: Selection and reentrant signaling in higher brain function. *Neuron* 10: 115–125.

Edelman GM and Tononi G (2000) *A Universe of Consciousness: How Matter Becomes Imagination*. New York: Basic Books.

Elston GN (2003) Cortex, cognition and the cell: New insights into the pyramidal neuron and prefrontal function. *Cereb. Cortex* 13: 1124–1138.

Elston GN and Rosa MG (1997) The occipitoparietal pathway of the macaque monkey: Comparison of pyramidal cell morphology in layer III of functionally related cortical visual areas. *Cereb. Cortex* 7: 432–52.

Elston GN and Rosa MG (1998) Morphological variation of layer III pyramidal neurones in the occipitotemporal pathway of the macaque monkey visual cortex. *Cereb. Cortex* 8: 278–294.

Everitt BJ and Robbins TW (2005) Neural systems of reinforcement for drug addiction: From actions to habits to compulsion. *Nat. Neurosci.* 8: 1481–1489.

Fang F and He S (2005) Cortical responses to invisible objects in the human dorsal and ventral pathways. *Nat. Neurosci.* 8(10): 1380–1385.

Faure PH, Neumeister H, Faber DS, and Korn H (2003) Symbolic analysis of swimming trajectories reveals scale invariance and provides a model for fish locomotion. *Fractals* 11: 233–243.

Felleman DJ and Van Essen DC (1991) Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex* 1–47.

Fessard A (1954) *Nervous Integration and Conscious Experience, Symposium Sainte-Marguerite*. London: Blackwell.

Fiset P, Paus T, Daloze T, et al. (1999) Brain mechanisms of propofol-induced loss of consciousness in humans: A positron emission tomographic study. *J. Neurosci.* 19: 5506–5513.

Friston KJ (1998) The disconnection hypothesis. *Schizophr. Res.* 30: 115–125.

Friston KJ, Tononi G, Reeke GN Jr, Sporns O, and Edelman GM (1994) Value-dependent selection in the brain: Simulation in a synthetic neural model. *Neuroscience* 59: 229–243.

Frith C, Perry R, and Lumer E (1999) The neural correlates of conscious experience: An experimental framework. *Trends Cogn. Sci.* 3: 105–114.

Frith U and Frith CD (2003) Development and neurophysiology of mentalizing. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 358 (1431): 459–473.

Fuentes U, Ritz R, Gerstner W, and Van Hemmen JL (1996) Vertical signal flow and oscillations in a three-layer model of the cortex. *J. Comput. Neurosci.* 3: 125–136.

Gajraj RJ, Doi M, Mantzaridis H, and Kenny GN (1999) Comparison of bispectral EEG analysis and auditory evoked potentials for monitoring depth of anaesthesia during propofol anaesthesia. *Br. J. Anaesth.* 82: 672–678.

Giedd JN, Blumenthal J, Jeffries NO, et al. (1999a) Brain development during childhood and adolescence: A longitudinal MRI study. *Nat. Neurosci.* 2(10): 861–863.

Giedd JN, Blumenthal J, Jeffries NO, et al. (1999b) Development of the human corpus callosum during childhood and adolescence: A longitudinal MRI study. *Prog. Neuropsychopharmacol*. *Biol. Psychiatry* 23(4): 571–588.

Goldman-Rakic PS (1988) Topography of cognition: Parallel distributed networks in primate association cortex. *Annu. Rev. Neurosci.* 11: 137–156.

Goldman-Rakic PS (1994) Working memory dysfunction in schizophrenia. *J. Neuropsychiatry Clin. Neurosci.* 6: 348–57.

Goodale MA, Milner AD, Jakobson LS, and Carey DP (1991) Object awareness. *Nature* 352: 202.

Granon S and Changeux JP (2006) Attention-deficit/hyperactivity disorder: A plausible mouse model? *Acta Paediatr.* 95: 645–649.

Granon S, Faure P, and Changeux JP (2003) Executive and social behaviors under nicotinic receptor regulation. *Proc. Natl. Acad. Sci. USA* 100: 9596–9601.

Granon S, Poucet B, Thinus-Blanc C, Changeux JP, and Vidal C (1995) Nicotinic and muscarinic receptors in the rat pefrontal cortex: Differential roles in working memory response selection and effortful processing. *Psychopharmacology (Berlin)* 119: 139–144.

Granon S, Vidal C, Thinus-Blanc C, Changeux JP, and Poucet B (1994) Working memory, response selection, and effortful processing in rats with medial prefrontal lesions. *Behav. Neurosci.* 108: 883–891.

Grillner S and Graybiel AM (2006) *Microcircuits: The Interface Between Neurons and Global Brain Function (Dahlem Workshop Reports)*. Cambridge, MA: MIT Press.

Gross J, Schmitz F, Schnitzler I, et al. (2004) Modulation of long-range neural synchrony reflects temporal limitations of visual attention in humans. *Proc. Natl. Acad. Sci. USA* Aug 31; 101(35): 13050–13055.

Gusnard DA (2005) Being a self: Considerations from functional imaging. *Conscious Cogn.* 14: 679–697.

Gusnard DA and Raichle ME (2001) Searching for a baseline: Functional imaging and the resting human brain. *Nat. Rev. Neurosci.* 2: 685–694.

Gutkin BS, Dehaene S, and Changeux JP (2006) A neurocomputational hypothesis for nicotine addiction. *Proc. Natl. Acad. Sci. USA* 24: 103: 1106–1111.

Han CJ, O'Tuathaigh CM, Van Trigt L, et al. (2003) Trace but not delay fear conditioning requires attention and the anterior cingulate cortex. *Proc. Natl. Acad. Sci. USA* 100: 13087–13092.

Harris G.C, Wimmer M, and Aston-Jones G (2005) A role for lateral hypothalamic orexin neurons in reward seeking. *Nature* 437: 556–559.

Haynes JD and Rees G (2005) Predicting the stream of consciousness from activity in human visual cortex. *Curr. Biol.* 15: 1301–1307.

Haynes JD and Rees G (2006) Decoding mental states from brain activity in humans. *Nat. Rev. Neurosci.* 7: 523–534.

Heidmann T and Changeux JP (1982) Molecular model of the regulation of chemical synapse efficiency at the postsynaptic level. *C. R Seances Acad. Sci.* III. 295: 665–670.

Heinke W and Schwarzbauer C (2002) In vivo imaging of anaesthetic action in humans: Approaches with positron emission tomography (PET) and functional magnetic resonance imaging (fMRI). *Br. J. Anaesth.* 89: 112–122.

Heiss WD, Pawlik G, Herholz K, Wagner R, and Wienhard K (1985) Regional cerebral glucose metabolism in man during wakefulness, sleep, and dreaming. *Brain Res.* 327: 362–326.

Hommel B, Kessler K, Schmitz F, et al. (2006) How the brain blinks: Towards a neurocognitive model of the attentional blink [review]. *Psychol. Res.* 70(6): 425–435.

Huxley TH (1874) On the hypothesis that animals are automata and its history. In: *Collected Essays of T. H. Huxley*, vol. 1, p. 199. New York: Appleton.

Author's personal copy

Hyder F, Rothman DL, and Shulman RG (2002) Total neuroenergetics support localized brain activity: Implications for the interpretation of fMRI. *Proc. Natl. Acad. Sci. USA* 99: 10771–10776.

Imas OA, Ropella KM, Ward BD, Wood JD, and Hudetz AG (2005) Volatile anesthetics disrupt frontal-posterior recurrent information transfer at gamma frequencies in rat. *Neurosci. Lett.* 387: 145–150.

Jasper HH (1998) Sensory information and conscious experience. *Adv. Neurol.* 77: 33–48.

Jeannerod M and Jacob P (2005) Visual cognition: A new look at the two-visual systems model. *Neuropsychologia* 43: 301–312.

Johnson MH (2005) Subcortical face processing. *Nat. Rev. Neurosci.* 6: 766–774.

Jones EG (1984) Laminar distribution of output cells. In: Peters A and Jones EG (eds.) *Cerebral Cortex*, *Vol. 1: Cellular Components of the Cerebral Cortex*, pp. 521–553. New York: Plenum.

Jones BE (1998) The neural basis of consciousness across the sleep-waking cycle. *Adv. Neurol.* 77: 75–94.

Kenet T, Bibitchkov D, Tsodyks M, Grinvald A, and Arieli A (2003) Spontaneously emerging cortical representations of visual attributes. *Nature* 425: 954–956.

Kolb B (1990) Animal models for human PFC-related disorders. *Prog. Brain Res.* 85: 501–519.

Koch C (2004) *The Quest for Consciousness: A Neurobiological Approach*. Englewood, CO: Roberts and Company Publishers.

Kunde W (2004) Response priming by supraliminal and subliminal action effects. *Psychol. Res.* 68: 91–96.

Lagercrantz H (2005) *I Barnets Hjärna*. Stockholm: Bonnier fakta.

Lagercrantz H, Hanson M, Evrard P, and Rodeck C (eds.) (2002) *The Newborn Brain: Neuroscience and Clinical Applications.* Cambridge, UK: Cambridge University Press.

Lamme VA (2003) Recurrent corticocortical interactions in neural disease. *Arch. Neurol.* 60: 178–184.

Lamme VA and Roelfsema PR (2000) The distinct modes of vision offered by feedforward and recurrent processing. *Trends Neurosci.* 23: 571–579.

Lamme VA, Zipser K, and Spekreijse H (2002) Masking interrupts figure-ground signals in V1. *J. Cogn. Neurosci.* 14: 1044–1053.

Landmann C, Dehaene S, Pappata S, et al. (2006) Dynamics of prefrontal and cingulate activity during a reward-based logical deduction task. *Cereb. Cortex* 17: 749–759.

Langford DJ, Crager SE, Shehzad Z, et al. (2006) Social modulation of pain as evidence for empathy in mice. *Science* 312: 1967–1970.

Laplane D and Dubois B (2001) Auto-activation deficit: A basal ganglia related syndrome. *Mov. Disord.* 16: 810–814.

Laufs H, Krakow K, Sterzer P, et al. (2003) Electroencephalographic signatures of attentional and cognitive default modes in spontaneous brain activity fluctuations at rest. *Proc. Natl. Acad. Sci. USA* 100: 11053–11058.

Laureys S (2005) The neural correlate of (un)awareness: Lessons from the vegetative state. *Trends Cogn. Sci.* 9: 556–559.

Laureys S, Boly M, and Maquet P (2006) Tracking the recovery of consciousness from coma. *J. Clin. Invest.* 116: 1823–1825.

Laureys S, Faymonville ME, Degueldre C, et al. (2000) Auditory processing in the vegetative state. *Brain* 123: 1589–1601.

Laureys S, Faymonville ME, Peigneux P, et al. (2002) Cortical processing of noxious somatosensory stimuli in the persistent vegetative state. *Neuroimage* 17: 732–741.

Laureys S, Owen AM, and Schiff ND (2004) Brain function in coma, vegetative state, and related disorders. *Lancet Neurol.* 3: 537–546.

Léna C, Popa D, Grailhe R, Escourrou P, Changeux J-P, and Adrien J (2004) Beta2-containing nicotinic receptors contribute to the organization of sleep and regulate putative micro-arousals in mice. *J. Neurosci.* 24: 5711–5718.

Levin ED (1992) Nicotinic systems and cognitive function. *J. Neurobiol.* 53: 633–640.

Livingstone MS and Hubel DH (1981) Effects of sleep and arousal on the processing of visual information in the cat. *Nature* 291: 554–561.

Llinas RR and Paré D (1991) Of dreaming and wakefulness. *Neuroscience* 44: 521–535.

Llinas R and Ribary U (1993) Coherent 40-Hz oscillation characterizes dream state in humans. *Proc. Natl. Acad. Sci. USA* 90(5): 2078–2081.

Llinas R, Ribary U, Contreras D, and Pedroarena C (1998) The neuronal basis for consciousness. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 353: 1841–1849.

Llinas RR and Steriade M (2006) Bursting of thalamic neurons and states of vigilance. *J. Neurophysiol.* 95: 3297–3308.

Lumer ED, Edelman GM, and Tononi G (1997a) Neural dynamics in a model of the thalamocortical system. I. Layers, loops and the emergence of fast synchronous rhythms.

Lumer ED, Edelman GM, and Tononi G (1997b) Neural dynamics in a model of the thalamocortical system. II. The role of neural synchrony tested through perturbations of spike timing. *Cereb. Cortex* 7: 228–236.

Mackenzie IR, Baker M, Pickering-Brown S, et al. (2006) The neuropathology of frontotemporal lobe degeneration caused by mutations in the progranulin gene. *Brain* 129(Pt. 11): 3081–3090.

Maquet P and Phillips C (1998) Functional brain imaging of human sleep. *J. Sleep Res.* 7(supplement 1): 42–47.

Marois R, Chun MM, and Gore JC (2000) Neural correlates of the attentional blink. *Neuron* 28: 299–308.

Marois R, Yi DJ, and Chun MM (2004) The neural fate of consciously perceived and missed events in the attentional blink. *Neuron* 41: 465–472.

Maskos U, Molles BE, Pons S, et al. (2005) Nicotine reinforcement and cognition restored by targeted expression of nicotinic receptors. *Nature* 436: 103–107.

Mazoyer B, Zago L, Mellet E, et al. (2001) Cortical networks for working memory and executive functions sustain the conscious resting state in man. *Brain Res. Bull.* 54: 287–298.

McCarley C (1999) A model of chronic dyspnea. *Image J. Nurs. Sch.* 31: 231–236.

McCormick DA and Bal T (1997) Sleep and arousal: Thalamocortical mechanisms. *Annu. Rev. Neurosci.* 20: 185–215.

McLeod JD (1991) Childhood parental loss and adult depression. *J. Health Soc. Behav.* 32: 205–220.

Meltzoff AN (1990) Towards a developmental cognitive science. The implications of cross-modal matching and imitation for the development of representation and memory in infancy. *Ann. N.Y. Acad. Sci.* 608: 1–31.

Meltzoff AN and Gopnik A (1993) The role of imitation in understanding persons and developing a theory of mind. In: Baron-Cohen S, Tager-Flusberg H, and Cohen DJ (eds.) *Understanding Other Minds: Perspectives from Autism*, pp. 335–366. Oxford, UK: Oxford University Press.

Mesulam MM (1998) From sensation to cognition. *Brain* 121: 1013–1052.

Mishkin M and Ungerleider LG (1982) Contribution of striate inputs to the visuospatial functions of parieto-preoccipital cortex in monkeys. *Behav. Brain Res.* 6: 57–77.

Miyashita Y and Hayashi T (2000) Neural representation of visual objects: Encoding and top-down activation. *Curr. Opin. Neurobiol.* 10: 187–194.

Naccache L (2005) Visual phenomenal consciousness: A neurological guided tour. *Prog. Brain Res.* 150: 185–195.

Author's personal copy

Naccache L and Dehaene S (2001) The priming method: Imaging unconscious repetition priming reveals an abstract representation of number in the parietal lobes. *Cereb. Cortex* 11: 966–974.

Naccache L, Dehaene S, Cohen L, et al. (2005) Effortless control: Executive attention and conscious feeling of mental effort are dissociable. *Neuropsychologia* 43: 1318–1328.

Naccache L, Gaillard R, Adam C, et al. (2005) A direct intracranial record of emotions evoked by subliminal words. *Proc. Natl. Acad. Sci. USA* 102: 7713–7717.

Neary D, Snowden JS, Northen B, and Goulding P (1988) Dementia of frontal lobe type. *J. Neurol. Neurosurg. Psychiatry* 51: 353–361.

Newby EA and Rock I (1998) Inattentional blindness as a function of proximity to the focus of attention. *Perception* 27(9): 1025–1040.

Nordberg A and Winblad B (1986) Reduced number of [^{3}H]nicotine and [^{3}H]acetylcholine binding sites in the frontal cortex of Alzheimer brains. *Neurosci. Lett.* 72: 115–119.

Norman DA and Shallice T (1986) Attention to action: Willed and automatic control of behaviour. In: Davidson RJ, Schwartz GE, Shapiro D (eds.) *Consciousness and Self-Regulation*, vol. 4. New York: Plenum.

O'Doherty JP (2004) Reward representations and reward-related learning in the human brain: Insights from neuroimaging. *Curr. Opin. Neurobiol.* 14: 769–776.

Pardo JV, Pardo PJ, Janer KW, and Raichle ME (1990) The anterior cingulate cortex mediates processing selection in the Stroop attentional conflict paradigm. *Proc. Natl. Acad. Sci. USA* 87: 256–259.

Parkin AJ and Walter BM (1992) Recollective experience, normal aging, and frontal dysfunction. *Psychol. Aging* 7: 290–298.

Paus T (2000) Functional anatomy of arousal and attention systems in the human brain. *Prog. Brain Res.* 126: 65–77.

Paus T, Koski L, Caramanos Z, and Westbury C (1998) Regional differences in the effects of task difficulty and motor output on blood flow response in the human anterior cingulate cortex: A review of 107 PET activation studies. *Neuroreport* 9: 37–47.

Petersen CC, Hahn TT, Mehta M, Grinvald A, and Sakmann B (2003) Interaction of sensory responses with spontaneous depolarization in layer 2/3 barrel cortex. *Proc. Natl. Acad. Sci. USA* 100: 13638–13643.

Portas CM, Krakow K, Allen P, Josephs O, Armony JL, and Frith CD (2000) Auditory processing across the sleep-wake cycle: Simultaneous EEG and fMRI monitoring in humans. *Neuron* 28: 991–999.

Posner MI (1994) Attention: The mechanisms of consciousness. *Proc. Natl. Acad. Sci. USA* 91: 7398–7403.

Posner MI and Dehaene S (1994) Attentional networks. *Trends Neurosci.* 17: 75–79.

Poucet B (1993) Spatial cognitive maps in animals: New hypotheses on their structure and neural mechanisms. *Psychol. Rev.* 100: 163–182.

Poucet B and Herrmann T (2001) Exploratory patterns of rats on a complex maze provide evidence for topological coding. *Behav. Processes* 53: 155–162.

Premack D and Woodruff G (1978) Chimpanzee problem-solving: A test for comprehension. *Science* 202: 532–535.

Preyer W (1894) *Mental Development in the Child*. New York: Appleton.

Raichle ME (1994) Images of the mind: Studies with modern imaging techniques. *Annu. Rev. Psychol.* 45: 333–356.

Raichle ME and Gusnard DA (2002) Appraising the brain's energy budget. *Proc. Natl. Acad. Sci. USA* 99: 10237–10239.

Raichle ME and Gusnard DA (2005) Intrinsic brain activity sets the stage for expression of motivated behavior. *J. Comp. Neurol.* 493: 167–176.

Raichle ME, MacLeod AM, Snyder AZ, Powers WJ, Gusnard DA, and Shulman GL (2001) A default mode of brain function. *Proc. Natl. Acad. Sci. USA* 98: 676–682.

Raizada RD and Grossberg S (2003) Towards a theory of the laminar architecture of cerebral cortex: Computational clues from the visual system. *Cereb. Cortex* 13: 100–113.

Rauschecker JP and Tian B (2000) Mechanisms and streams for processing of ''what'' and ''where'' in auditory cortex. *Proc. Natl. Acad. Sci. USA* 97: 11800–11806.

Raymond CK, Howald-Stevenson I, Vater CA, and Stevens TH (1992) Morphological classification of the yeast vacuolar protein sorting mutants: Evidence for a prevacuolar compartment in class E vps mutants. *Mol. Biol. Cell* 3: 1389–1402.

Rees G, Russell C, Frith CD, and Driver J (1999) Inattentional blindness versus inattentional amnesia for fixated but ignored words. *Science* 286: 2504–2507.

Reuter F, Del Cul A, and Audoin B (2007) Intact subliminal processing and delayed conscious access in multiple sclerosis. *Neuropsychologia* Apr 19 [epub ahead of print].

Reynolds GD and Richards JE (2005) Familiarization, attention, and recognition memory in infancy: An event-related potential and cortical source localization study. *Dev. Psychol.* 41(4): 598–615.

Rochat P (2003) Five levels of self-awareness as they unfold early in life. *Conscious Cogn.* 12: 717–731.

Rondi-Reig L, Petit GH, Tobin C, Tonegawa S, Mariani J, and Berthoz A (2006) Impaired sequential egocentric and allocentric memories in forebrain-specific-NMDA receptor knock-out mice during a new task dissociating strategies of navigation. *J. Neurosci.* 26: 4071–4081.

Rose S (ed.) (1999) *From Brains to Consciousness? Essays on the New Sciences of the Mind*. Princeton, NJ: Princeton University Press.

Salin PA and Bullier J (1995) Corticocortical connections in the visual system: Structure and function. *Physiol. Rev.* 75: 107–154.

Schultz W, Dayan P, and Montague PR (1997) A neural substrate of prediction and reward. *Science* 275(5306): 1593–1599.

Sergent C, Baillet S, and Dehaene S (2005) Timing of the brain events underlying access to consciousness during the attentional blink. *Nat. Neurosci.* 8: 1391–1400.

Shallice T (1988) *From Neuropsychology to Mental Structure*. New York: Cambridge University Press.

Sergent C and Dehaene S (2004) Neural processes underlying conscious perception: Experimental findings and a global neuronal workspace framework. *J. Physiol. Paris* 98: 374–384.

Shulman GL, Hyder F, and Rothman DL (2003) Cerebral metabolism and consciousness. *C. R. Biol.* 326: 253–273.

Shulman RG and Rothman DL (1998) Interpreting functional imaging studies in terms of neurotransmitter cycling. *Proc. Natl. Acad. Sci. USA* 95: 11993–11998.

Shulman RG, Rothman DL, Behar KL, and Hyder F (2004) Energetic basis of brain activity: Implications for neuroimaging. *Trends Neurosci.* 27: 489–495.

Shytle RD, Silver AA, Sheehan KH, Sheehan DV, and Sanberg PR (2002) Neuronal nicotinic receptor inhibition for treating mood disorders: Preliminary controlled evidence with mecamylamine. *Depress. Anxiety* 16: 89–92.

Sigman M and Dehaene S (2005) Parsing a cognitive task: A characterization of the mind's bottleneck. *PLoS Biol.* 3: e37.

Sigman M and Dehaene S (2006) Dynamics of the central bottleneck: Dual-task and task uncertainty. *PLoS Biol.* 4: e220.

Simons DJ and Chabris CF (1999) Gorillas in our midst: Sustained inattentional blindness for dynamic events. *Perception* 28: 1059–1074.

Author's personal copy

Singer T (2006) The neuronal basis and ontogeny of empathy and mind reading: Review of literature and implications for future research. *Neurosci. Biobehav. Rev.* 30: 855–863.

Sleigh JW, Steyn-Ross DA, Steyn-Ross ML, Williams ML, and Smith P (2001) Comparison of changes in electroencephalographic measures during induction of general anaesthesia: Influence of the gamma frequency band and electromyogram signal. *Br. J. Anaesth.* 86: 50–8.

Squire LR (1987–1988) The organization and neural substrates of human memory. *Int. J. Neurol.* 21–22: 218–222.

Steinlein OK (2004) Genetic mechanisms that underlie epilepsy. *Nat. Rev. Neurosci.* 5: 400–408.

Steriade M (1999) Coherent oscillations and short-term plasticity in corticothalamic networks. *Trends Neurosci.* 22: 337–345.

Steriade M, Contreras D, Amzica F, and Timofeev I (1996) Synchronization of fast (30–40 Hz) spontaneous oscillations in intrathalamic and thalamocortical networks. *J. Neurosci.* 16: 2788–2808.

Steriade M, McCormick DA, and Sejnowski TJ (1993) Thalamocortical oscillations in the sleeping and aroused brain. *Science* 262: 679–685.

Strawson PF (1974) *Freedom and Resentment, and Other Essays.* London: Methuen.

Super H, Spekreijse H, and Lamme VA (2001) Two distinct modes of sensory processing observed in monkey primary visual cortex (V1). *Nat. Neurosci.* 4: 304–310.

Sutton RS and Barto AG (1981) Toward a modern theory of adaptive networks: Expectation and prediction. *Psychol. Rev.* 88: 135–170.

Thinus-Blanc C, Save E, Poucet B, and Foreman N (1996) Effects of parietal cortex lesions on spatial problem solving in the rat. *Behav. Brain Res.* 81: 115–121.

Thorndike EL (1898) *Animal Intelligence: An Experimental Study of the Association Processes in Animals*. New York: Macmillan (Psychological Review, Monograph Supplements, No. 8).

Tononi G and Edelman GM (1998) Consciousness and complexity. *Science* 282: 1846–1851.

Trivers R (1985) *Social Evolution*. Menlo Park, CA: Benjamin.

Tsodyks M, Kenet T, Grinvald A, and Arieli A (1999) Linking spontaneous activity of single cortical neurons and the underlying functional architecture. *Science* 286: 1943–1946.

Ungerleider LG (1995) Functional brain imaging studies of cortical mechanisms for memory. *Science* 270: 769–775.

Vogel EK, Luck SJ, and Shapiro KL (1998) Electrophysiological evidence for a postperceptual locus of suppression during the attentional blink. *J. Exp. Psychol. Hum. Percept. Perform.* 24: 1656–1674.

von Economo C (1929) *The Cytoarchitectonics of the Human Cerebral Cortex*. London: Oxford University Press.

Wang XJ (1993) Ionic basis for intrinsic 40 Hz neuronal oscillations. *Neuroreport* 5: 221–224 [erratum in: *Neuroreport* (1994): 5: 531].

Weiskrantz L (1991) Disconnected awareness for detecting, processing, and remembering in neurological patients. *J. R. Soc. Med.* 84: 466–470.

Weiskrantz L (1997) Fragments of memory. *Neuropsychologia* 35: 1051–1057.

Weissman DH, Roberts KC, Visscher KM, and Woldorff MG (2006) The neural bases of momentary lapses in attention. *Nat. Neurosci.* 9: 971–978.

Yerkes RM (1916) *The Mental Life of Monkeys and Apes*. New York: Holt.

Zéki S (2003) The disunity of consciousness. *Trends Cogn. Sci.* 7: 214–218.

Zelazo PD (1996) Towards a characterization of minimal consciousness. *New Ideas Psychol.* 14: 63–80.

Zelazo PD (2004) The development of conscious control in childhood. *Trends Cogn. Sci.* 8: 12–17.

Zeman A (2001) Consciousness. *Brain* 124: 1263–89.

Zeman A (2005) What in the world is consciousness? *Prog. Brain Res.* 150: 1–10.