Neurosciences and the Human Person: New Perspectives on Human Activities
Pontifical Academy of Sciences, *Scripta Varia* 121, Vatican City 2013
www.casinapioiv.va/content/dam/accademia/pdf/sv121/sv121-dehaene.pdf

The brain mechanisms of conscious access and introspection

Stanislas Dehaene

Introduction

γνῶθι σεαυτόν: know thyself. This famous maxim, inscribed in the pronaos of the Apollo temple in Delphi, draws our attention to a remarkable competence of the human brain: the capacity to bring to the forefront of our awareness, not just sensory information from the external world, but also aspects of our inner mental life. Indeed, a characteristic feature of *Homo sapiens sapiens* is that we are conscious of being conscious. A talented painter of introspection, Vladimir Nabokov lyrically summarized, in *Strong Opinions*, the fascinating reflections of this mirror seemingly turned onto itself:

> Being aware of being aware of being… if I not only know that I am but also know that I know it, then I belong to the human species. All the rest follows – the glory of thought, poetry, a vision of the universe. In that respect, the gap between ape and man is immeasurably greater than the one between amoeba and ape.

How does consciousness work? Can it be reduced to the operation of the brain? What are its neurobiological mechanisms? For a long period, these questions were considered beyond the realm of cognitive psychology and neuroscience. Consciousness was considered an unnecessary term. John Broadus Watson forcefully rejected introspection and consciousness from the science of psychology which he sketched in his 1913 manifesto *Psychology as the behaviorist views*:

> Psychology as the behaviorist views it is a purely objective experimental branch of natural science. Its theoretical goal is the prediction and control of behavior. Introspection forms no essential part of its methods, nor is the scientific value of its data dependent upon the readiness with which they lend themselves to interpretation in terms of consciousness.

Although cognitive science rejected behaviorism, the anti-introspection view left a durable mark. During the cognitive revolution (approximately 1960 to 1990), consciousness was barely mentioned, even less studied (with a few major exceptions, e.g. Bisiach, Luzzatti, & Perani, 1979; Frith, 1979; Libet, Alberts, Wright, & Feinstein, 1967; Marcel, 1983; Posner, Snyder, Balota, & Marsh, 1975/2004; Shallice, 1972; Weiskrantz, 1986).

Philosophical approaches failed to shed much light on the problem of how an assembly of nerve cells could produce conscious thoughts. René Descartes, although propounding a materialistic approach to perception, action, emotion and memory, conceived of human consciousness as belonging to an entirely different realm (*res cogitans*) (Descartes, 1648/1937). This dualist position was scientifically unproductive, since it essentially barred any experimental approach. Surprisingly, dualism remained an appealing intuition for some contemporary philosophers (Chalmers, 1996) and even neuroscientists (Eccles, 1994). More recently, other philosophers, capitalizing on their intuitions, introduce additional ill-defined concepts of "qualia", "phenomenal awareness", or "what it is like" to have a certain experience (Block, 1995; Nagel, 1974). Yet others sought a haven in quantum mechanics, in the hope that its mysterious non-deterministic rules would somehow leave room for a conscious observer and free will (Penrose, 1990; Penrose & Hameroff, 1998). It is fair to say, however, that such approaches have not yielded any scientific progress so far, but only theoretical constructs of a highly speculative nature (Eccles, 1994).

It is only in the past twenty years or so that the problem of consciousness recovered its status as a respectable empirical question in experimental psychology and neuroscience. A handful of philosophers (e.g. Churchland, 1986; Dennett, 1991), psychologists (e.g. Baars, 1989; Dehaene & Naccache, 2001), neuropsychologists (e.g. Weiskrantz, 1986) and neuroscientists (e.g. Crick & Koch, 1990; Logothetis, Leopold, & Sheinberg, 1996) argued that consciousness was, first and foremost, a well-defined experimental problem. Indeed, consciousness poses an urgent problem in the clinic where the loss of consciousness in coma, epilepsy or anesthesia is a frequent and yet ill-understood and poorly controlled phenomenon. Fortunately, consciousness can be easily monitored and even manipulated through many different paradigms (e.g. sleep, anesthesia, visual illusions, inattention, confidence reports, etc). The brain mechanisms underlying these manipulations can then be dissected using behavioral measures, neuroimaging and electrophysiology. Animal models of conscious and unconscious behavior may even be conceived (Cowey & Stoerig, 1995).

From this realization emerged a flurry of experimental results and theoretical models. Today, there is both a solid dataset on the brain mechanisms of conscious processing and some convergent theoretical proposals. In this chapter, I will briefly review them (for an in-depth review, see Dehaene & Changeux, 2011).

The multiple meanings of consciousness

Three ingredients permitted a solid line of empirical attack on the problem of consciousness: (1) better definitions of the terms; (2) minimal experimental paradigms; and (3) a careful quantification of introspection. I will consider them in turn.

The word "consciousness", as used in everyday language, is loaded with multiple meanings. Contemporary cognitive neuroscience made progress by recognizing the need to distinguish a minimum of three concepts.

1. **Vigilance**, also called wakefulness, is what varies when we fall asleep or wake up. It relates to the intransitive use of the word "consciousness" in everyday language (as when we say "the patient is still conscious"). It is a necessary but not sufficient condition for conscious access and conscious processing.
2. **Selective attention** is the focusing of mental resources on a subset of the available information. Attention selects some information, separates it from the background, and deepens its processing. Selective attention is typically a non-conscious process that gates access to consciousness.
3. **Conscious access** is the entry of some of the attended information into a second post-perceptual stage of cognitive processing which making it durable, available to many additional cognitive processes, and reportable to others. It relates to the transitive use of the word "consciousness" in everyday language (as when we say "The driver was conscious *of* the red light"). Information which has been consciously accessed can then be submitted to **conscious processing**: it can be channeled, in a typically serial manner, through a series of controlled information-processing stages.

Experiments indicate that the three concepts of vigilance, attention and conscious access are dissociable. For instance, vigilance (or wakefulness) may still exist when conscious access is gone: patients in vegetative state may still have a sleep-wake cycle, but their capacity to access, manipulate and report information is lost. Similarly, attention may exist without conscious access: in the laboratory, we can create conditions in which attention is demonstrably attracted by a flashed picture, and even selectively amplifies it, although the picture remains invisible (Koch & Tsuchiya, 2007; Naccache, Blandin, & Dehaene, 2002). Thus, conscious access is a distinct cognitive entity from both vigilance and selective attention.

Conscious access to sensory information is a simple and well-delimited construct that plays a central role in empirical studies of consciousness (Crick & Koch, 1990; Dehaene, Changeux, Naccache, Sackur, & Sergent,

2006). At any given moment, our brain is bombarded with sensory stimulation, which activates many peripheral sensory areas of the brain. Yet we only gain conscious access to one, or just a few, of these elements of information, while the rest remains unconscious. Conscious access has a limited capacity: if we attend to one object, we may transiently lose consciousness of the surrounding ones. The problem of conscious access consists in understanding what brain mechanisms underlie this limited capacity of consciousness.

There are yet other meanings of consciousness. **Self-consciousness** refers specifically to instances of conscious access in which the information being manipulated or reported is internal to the organism. Multiple aspects of self-consciousness may be distinguished: the capacity to represent our body and its limits; the separation of our actions from those of others (agency); and the formation of a "point of view" on the external world. All of these aspects can be and have been studied experimentally. We understand increasingly well how self-consciousness arises from a combination of brain circuits specializing in the representation of different aspects of our selves (sensory maps of the body, vestibular signals of head stability, programming of intentional movements, etc) (see e.g. Lenggenhager, Tadi, Metzinger, & Blanke, 2007).

There is also **recursive consciousness**, also known as **metacognition**. This is the capacity to "know oneself", i.e. to introspect and obtain information about one's own mental processes. Such information is called "metacognitive" because it provides a higher-order representation of the content, value or quality of some other information represented elsewhere in the system. We rely on metacognition when we evaluate our confidence in a past decision, or when we realize that we do not remember something. A broad array of experimental research, too large to be reviewed here, is available on this topic (Dunlosky & Metcalfe, 2008).

Importantly, research indicates that all of the above aspects of consciousness (vigilance, selective attention, conscious access, self-consciousness and metacognition) are not unique to humans, but are also available to many other animal species such as macaque monkeys. In particular, it is clearly incorrect to think of recursive consciousness as limited to the human species (as Nabokov did in the above citation): there are now well-defined animal models of this ability, in which animals act in ways that indicate some degree of knowledge of their own confidence and fallibility (Terrace & Son, 2009).

Some philosophers consider one last aspect of consciousness as worthy of a separate term: **phenomenal awareness** (Block, 1995; Chalmers, 1996). This term is used to refer to the subjective, feel of conscious experience (also

called *qualia*) – "what it is like" to experience, for instance, a gorgeous sunset or a terrible toothache. Introspectively, there is no doubt that these mental states are real and must be explained. However I share with the philosopher Dan Dennett the view that, as a philosophical concept, phenomenal awareness remains too fuzzily defined to be experimentally useful (Dennett, 2001). Whatever empirical content there is to *qualia* seems to be already covered by the concept of conscious access. A burning sensation, for instance, can be tracked as it makes its way into the brain and becomes transformed from a preconscious sensation in somatosensory cortex to a conscious feeling of pain in the anterior cingulate (Rainville, Duncan, Price, Carrier, & Bushnell, 1997). Whether there is anything left of phenomenal awareness once conscious access is taken care of is highly debatable: the other aspects that philosophers consider as central for the *qualia* concept, such as their ineffable character, remain largely untestable. In the rest of this chapter, I will thus primarily focus on the brain mechanisms of conscious access.

Minimal experimental paradigms for conscious access

The second ingredient that led to the contemporary science of consciousness was the recognition that a broad array of experimental paradigms was available to manipulate conscious access in the lab. With these tools, it became possible to create reproducible states of conscious and unconscious perception (Baars, 1989) (see figure 1).

One paradigm is provided by visual illusions such as **binocular rivalry** or **motion-induced blindness.** In these illusions, the stimulus is fixed, and yet the content of consciousness repeatedly changes. In motion-induced blindness, a visible disc, when touched by a cloud of moving dots, transiently vanishes from consciousness at seemingly random moments. In binocular rivalry, two pictures objectively presented to the two eyes alternate in awareness: subjectively, we never see them both at the same time. With such stimuli, it becomes feasible to ask a simple empirical question: Which aspects of brain activity vary in parallel to conscious experience? Neurons in the primary visual cortex typically discharge only in relation to the fixed, objective stimulus, but neurons in higher associative areas of the visual cortex show on and off responses in direct correlation to the subjective reports of visibility and invisibility, making them a neural correlate of conscious access (Logothetis, *et al.*, 1996).

Other visual illusions give scientists complete experimental control over the moment at which sensory stimuli vanish from conscious awareness. One such paradigm is **masking**: a target word or picture is briefly flashed on a computer screen, with an intensity clearly sufficient to make it visible. How-

ever, when the picture is followed, at a short interval, by another such stimulus (the "mask"), it may become totally invisible. Such a stimulus is called **subliminal**, i.e. below the threshold for conscious access. As reviewed below, psychological and brain-imaging experiments indicate that subliminal stimuli continue to be actively processed in the brain at multiple levels: the identity, the meaning, and even the action cued by a subliminal word can be partially activated without awareness. It is now very clear that a great variety of brain regions, located virtually everywhere in the cortex, can operate in a non-conscious mode (with the possible exception of dorsolateral prefrontal cortex). Studies of subliminal processing therefore help delimit what consciousness is *not*.

Inattention offers a third type of experimental paradigm. Here, the subject is temporarily absorbed by a demanding task on a first target T1. During this period, a second target T2 is briefly presented. Under such conditions, the limited capacity of conscious access is such that the second stimulus T2 may fail to be perceived at all, giving rise to **attentional blink** (Raymond, Shapiro, & Arnell, 1992) or **inattentional blindness** (Mack & Rock, 1998). The invisible T2 stimulus is said to be **preconscious**. This term specifically refers to a *temporary* invisibility: a preconscious stimulus, unlike a masked word, may become conscious if it is presented in the absence of any distracting or attention-grabbing thought. Preconscious stimuli are therefore useful in the study of consciousness because the very same stimulus may or may not be conscious at different times, under experimental control.

All of these experimental manipulations provide examples of **minimal contrasts** between conscious and non-conscious processing. In the laboratory, we can create experimental conditions that, in the ideal case, vary *only* in the presence or absence of consciousness. Not only the stimulus itself, but also the participant's responses, can be equated between conscious and non-conscious trials. Indeed, it is possible to exploit the fact that participants often respond at better-than-chance levels to non-conscious stimuli (such non-conscious performance is often called **blindsight**). Contrasting conscious and non-conscious trials in which the same stimulus is presented, and the same correct response is emitted, turns conscious access into a pure experimental variable that can be decorrelated from other input and output contingencies (Lamy, Salti, & Bar-Haim, 2009). The goal of the cognitive neuroscience of consciousness is precisely to understand what types of cognitive processes and brain activity distinguish conscious versus non-conscious trials, or reportable versus non-reportable trials, when everything else is kept identical.

The crucial role of introspection

Not only can sensory stimuli be made to vanish from conscious experience, but it is, in fact, possible to select fixed conditions of stimulation that are just at threshold, such that participants report seeing a stimulus on only half the trials (e.g. Sergent, Baillet, & Dehaene, 2005). By asking participants to report their subjective perception on each trial, we can later sort the trials into "seen" and "unseen", and probe the brain activation differences between them.

This approach illustrates the third key ingredient in the study of consciousness: taking introspection seriously. Introspective reports define the very phenomenon that a science of consciousness purports to study: the subjective, first-person mental states that occupy the mind of a given person and that only he or she knows about. The modern science of consciousness uses numerical scales and other devices to carefully register and quantity subjective introspective reports, such that they can be studied scientifically (Marti, Sackur, Sigman, & Dehaene, 2010; Overgaard, Rote, Mouridsen, & Ramsoy, 2006; Sergent & Dehaene, 2004; Sigman, Sackur, Del Cul, & Dehaene, 2008). The results indicate that illusions can be highly reliable across subjects. This is a crucial fact: although subjectivity is a private and first-person phenomenon, its reports obey psychological laws that are highly reproducible across individuals and can therefore be studied by the standard scientific method (e.g. Marti, *et al.*, 2010).

This realization took some time. As noted in the introduction, introspection has long had a bad reputation in cognitive neuroscience. It was long considered as a poor and unreliable measure that could not be used to found a solid psychological science (Nisbett & Wilson, 1977). This critique, however, conflated two different issues: introspection as a research method, and introspection as raw data. As a research method, introspection cannot be trusted to provide direct information about mental processes. Human subjects often supply inappropriate explanations for their behavior (Johansson, Hall, Sikstrom, & Olsson, 2005). We cannot count on them to tell us how their mind works, precisely because so much of mental computation occurs non-consciously. However, the introspections they provide, however weird or wrong, must still be explained. The correct view is to treat them as raw data in need of an explanation. Visual illusions, in this sense, are "real" phenomena in need of an explanation, and which have the potential to illuminate the mechanisms of consciousness.

Perhaps the best case in point is the "out-of-body" experience in which subjects report a feeling of leaving their body and watching themselves from above. We obviously cannot take them literally – but we can still examine

what brain processes cause this subjective experience. Olaf Blanke's research converges onto a cortical region in the right temporo-parietal junction which, when impaired or electrically perturbed, causes a systematic illusion of self displacement, which can now be systematically reproduced in normal subjects (Blanke, Landis, Spinelli, & Seeck, 2004; Blanke, Ortigue, Landis, & Seeck, 2002) (see Olaf Blanke's chapter in the present volume).

Cognitive signatures of consciousness

With those three ingredients at hand (a focus on conscious access, minimal paradigms contrasting conscious and non-conscious perception, and a careful quantification of introspection), the cognitive psychology and neuroscience of consciousness made huge strides in the past twenty years.

A first axis of research focused on the depth of unconscious processing. Using primarily masked priming and attentional blink paradigms, it was discovered that even stimuli that are totally unconscious can be processed up to a considerable depth (for review, see Kouider & Dehaene, 2007). An unseen picture, word or digit can be identified non-consciously. Even its meaning can be partially extracted. For instance, an unseen emotional word such as "rape", masked below threshold, still activates the amygdala, a brain center involved in fear and other emotions (Naccache, *et al.*, 2005). Even complex operations, such as computing the approximate average of several digits (Van Opstal, de Lange, & Dehaene, 2011) or the combination of multiple decision cues (de Lange, van Gaal, Lamme, & Dehaene, 2011; Dijksterhuis, Bos, Nordgren, & van Baaren, 2006), can unfold without consciousness. The guiding of our movements and the quick inhibition or correction of an inappropriate response also fall within the realm of non-conscious processing (Logan & Crump, 2010; Nieuwenhuis, Ridderinkhof, Blom, Band, & Kok, 2001). The exploration of the limits of non-conscious processing continues to this day, and it is likely that powerful yet non-conscious operations of the brain remain to be discovered. As a rule, we seem to constantly under-estimate the amount of non-conscious processing. It can be said that the vast majority of cognitive operations of the human brain occur without awareness.

While conscious processing thus appears only as the tip of the iceberg, are there cognitive operations can only be deployed when the information is consciously represented? It seems that the answer is positive. With a non-conscious target, cognitive operations can be launched, but they typically do not run to completion. Attaining a firm decision, developing a confident intention, and executing a strategy comprising multiple serial steps, are operations that seem to require conscious perception (de Lange, *et al.*, 2011;

Sackur & Dehaene, 2009). The quality of the extracted information, its durable maintenance and its flexible use in multiple tasks are drastically enhanced on conscious relative to non-conscious trials (Del Cul, Dehaene, Reyes, Bravo, & Slachevsky, 2009).

These data suggest that consciousness is not just an epiphenomenon or an illusion, but fulfills a specific role that may have been positively selected for in evolution: the amplification and global sharing of specific information selected for its likely relevance to the organisms' current goals.

Brain signatures of consciousness

At the neurophysiological level, contrasts between conscious and unconscious stimuli have revealed a number of signatures of consciousness.

Brain imaging techniques have been used, for instance, to track the fate of a flashed visual stimulus such as a word as it enters the retina and, depending on the trial, is or is not consciously perceived. Records of brain activity have revealed that the initial perceptual stages may remain almost strictly identical on conscious and non-conscious trials: the entry of the stimulus into visual areas and its feed-forward propagation into occipital, temporal and parietal cortices can proceed non-consciously (e.g. Sergent, *et al.*, 2005). The brain appears to accumulate evidence about the identity of a subliminal stimulus (Del Cul, Baillet, & Dehaene, 2007), and many specialized areas of the cortex, including motor areas, can receive these unconscious signals and bias their decisions towards the unperceived target (Dehaene, Naccache, *et al.*, 1998; Vorberg, Mattler, Heinecke, Schmidt, & Schwarzbach, 2003).

What seems to be unique to consciousness is a relatively late (~200-300 milliseconds), sudden and non-linear amplification of the incoming activation (Del Cul, *et al.*, 2007). After a brief transition period, the difference between conscious and unconscious trials quickly becomes qualitative, as many areas show a sudden activation ("ignition") only on conscious trials (Dehaene & Changeux, 2005; Del Cul, *et al.*, 2007; Fisch, *et al.*, 2009). When it is conscious, the incoming activation is suddenly amplified and reverberates bidirectionally (bottom-up and top-down) within a large network of distant brain areas, frequently including the original perceptual areas as well higher association cortices in the temporal, parietal and prefrontal lobes. This state of activity is meta-stable and can last for a long duration, long after the original stimulus is gone.

At the surface of the head, conscious ignition is characteristically accompanied by a broad component of the average electro-encephalogram (EEG) called the P300 wave (because its latency is typically 300 milliseconds

or more). The brain generators of the P300 have been shown by intracranial recordings to involve a highly distributed set of nearly-simultaneous active areas including hippocampus and temporal, parietal and frontal association cortices (Gaillard, *et al.*, 2009; Halgren, Marinkovic, & Chauvel, 1998).

Additional signatures of consciousness can be obtained by examining the spontaneous fluctuations of brain signals and whether they index a global, brain-scale state of synchronized activation. A late and distributed burst of local high-frequency activity in the gamma band (>30 Hz), a massive increase in the synchrony between distant brain signals in the beta band (13-30 Hz), and a bidirectional sharing of mutual information and causal links, when occurring in a late time window, all constitute markers of conscious access (Gaillard, *et al.*, 2009).

An important axis of recent research consists in probing the generality of these putative signatures of consciousness (review in Dehaene & Changeux, 2011). Beyond the perception of brief visual stimuli, these markers have begun to be replicated in auditory and tactile perception. Probing these markers during anesthesia and in brain-lesioned patients with loss of consciousness also confirms their tight association with conscious perception.

Importantly, a similar two-stage sequence, with non-conscious focal processing followed by a global synchronous conscious state, has also been observed in studies of conscious access to non-sensory information. For instance, when we are aware of having made an error, a focal and unconscious error-related negativity is followed by a late and global wave, the error positivity, which tightly resembles the sensory P300 (Nieuwenhuis, *et al.*, 2001). A similar sequence can also be evoked by direct brain stimulation: during the conscious state, a magnetic pulse induces activation that propagates to multiple distant brain areas for durations extending beyond 300 ms, while during the anesthetized or sleep state, the same pulse induces only a local activation that quickly dissipates (Ferrarelli, *et al.*, 2010; Massimini, Boly, Casali, Rosanova, & Tononi, 2009). New mathematical measures of information integration or non-linear dimensionality (Velly, *et al.*, 2007) are now being developed to provide improved markers of the global exchange of information across distant areas which characterizes consciousness.

Global workspace theory

My colleagues and I introduced the theory of a Global Neuronal Workspace (GNW) as a putative neurobiological architecture capable of accounting for cognitive and neuroscience observations on unconscious and conscious processing (Dehaene & Naccache, 2001). GNW theory assumes that cortical areas and subcortical nuclei contribute to a great variety of

specialized sub-circuits implementing unconscious and modular "processors" which operate in parallel. Non-conscious stimuli can thus be quickly and efficient processed along automatized or pre-instructed processing routes. However, GNW theory proposes that besides these encapsulated processors, the brain also comprises an architecture which allows a subset of the available information to be globally broadcasted. The GNW breaks the brain's modular organization by allowing selected information to be flexibly routed to various processes of verbal report, evaluation, memory, planning and intentional action (Baars, 1989; Dehaene & Naccache, 2001). Dehaene and Naccache (2001) postulate that "this global availability of information (…) *is* what we subjectively experience as a conscious state".

The hypothetical neurobiological mechanism for global availability is a set of large cortical pyramidal cells with long-range excitatory axons (GNW neurons), together with their relevant thalamo-cortical loops. These cells are present throughout the human cortex, yet they are particularly dense in prefrontal, cingulate, and parietal regions. They form a long-distance network that interconnects associative cortical areas and allows them to flexibly recruit, in a top-down manner, virtually any specialized area. Through their numerous reciprocal connections, GNW neurons are thought to amplify and maintain a specific neural representation for an arbitrary duration, thus keeping it "on line" or "in mind". At any given moment, a conscious content is assumed to be encoded in the sustained activity of a fraction of GNW neurons, the rest being inhibited. The long-distance axons of GNW neurons then broadcast it to many other processors brain-wide. Global broadcasting allows information to be more efficiently processed (because it is no longer confined to a subset of non-conscious circuits, but can be flexibly shared by many cortical processors) and to be verbally reported (because these processors include those involved in formulating verbal messages).

Artificial neuronal networks based on the workspace architecture have been explored in computer simulations (Dehaene & Changeux, 2005; Dehaene, Kerszberg, & Changeux, 1998; Shanahan, 2008; Zylberberg, Fernandez Slezak, Roelfsema, Dehaene, & Sigman, 2010). Their behavior has revealed dynamic electrophysiological phenomena very similar to the above experimental observations. When a brief pulse of sensory stimulation was applied to the model network, activation propagated according to two successive phases: (1) initially, a brief wave of excitation progressed into the simulated hierarchy through feedforward connections, with an amplitude and duration directly related to the initial input; (2) in a second stage, mediated by slower feedback connections, the network entered into a global self-sustained "ignited" state. This ignition was characterized by an increased

power of local cortico-thalamic oscillations in the gamma band, and by an increased synchrony across distant regions. This two-stage dynamics of the computer model reproduced most of the signatures of conscious access that have been empirically observed.

The model easily explains why conscious access exhibits a sharp threshold that separates supra- from sub-liminal stimuli. In GNW theory, the transition to the ignited or "conscious" state can be characterized as a phase transition in network activity. By amplifying its own incoming activity, the GNW exhibits a dynamic threshold with a fast non-linear divergence. Within a few tens of milliseconds, depending on stimulus strength, activity either rises to a high state, or decays to a low state. Even for a fixed stimulus, spontaneous activity and pre-stimulus oscillations impose a stochasticity on global ignition, explaining why the same stimulus can sometimes be perceived and sometimes remain unconscious. Computer simulations also exhibit analogs of the attentional blink and inattentional blindness phenomena: at any given moment, the ignition of the workspace by one cell assembly can prevent the simultaneous conscious access to a second piece of information.

An original feature of the GNW model, absent from many other formal neural network models, is the occurrence of highly structured spontaneous activity (Dehaene & Changeux, 2005). Just like real neurons, the simulated GNW neurons can fire spontaneously, with a fringe of variability, even in the absence of external inputs. In a GNW architecture, this spontaneous activity propagates in a top-down manner, starting from the highest hierarchical levels of the simulation, to form globally synchronized ignited states. The dynamics of such networks is thus characterized by a constant flow or "stream" of individual coherent episodes of variable duration In more complex network architectures, this stochastic activity can be shaped by reward signals in order to achieve a defined goal state, such as solving a logical problem (Dehaene & Changeux, 2000). These simulations provide a preliminary account of how higher cortical areas spontaneously activate in a coordinated manner during conscious effortful tasks.

In summary, the theoretical proposal is that conscious access corresponds to the selection and temporary maintenance of information encoded in the sustained activity of a distributed network of neurons with long-distance axons (the Global Neuronal Workspace). The GNW theory accounts for at least three aspects of subjective experience: (1) individuality: the same stimulus may or may not lead to conscious ignition, and whether such ignition occurs, in a given brain, is a stochastic event unique to each individual; (2) durability: thanks to its reverberating self-connectivity, the GNW network can maintain information "in mind" for an arbitrary duration, long after

the actual sensory stimulation has vanished; (3) autonomy: the shaping of spontaneous activity by GNW circuits leads to the stochastic endogenous generation of a series of activation patterns, potentially accounted for the never-ending "stream of consciousness".

Towards clinical applications

The discovery of the brain mechanisms of consciousness is not just an intellectual exercise. Our research is strongly motivated by the need to provide better experimental and conceptual tools to clinicians. Every year, due to stroke, head trauma or hypoxia, thousands of patients lose consciousness. The current clinical classification distinguishes several states:

- Brain death: complete and irreversible absence of brain function, marked by the durable absence of any detectable electro-encephalogram (EEG) and brain stem reflexes, which cannot be explained by hypothermia or drugs.
- Coma: prolonged loss of the capacity to be aroused, typically accompanied by slow-wave EEG and a variable preservation of cranial nerve and brain stem reflexes.
- Vegetative state: preserved sleep-wake cycle, yet with a total lack of responsiveness and voluntary action.
- Minimally conscious state: presence of rare, inconsistent, and limited signs of understanding and voluntary responding.
- Locked-in syndrome: fully preserved awakening and awareness, yet with complete or near-complete incapacity to report it due to paralysis (eye motion can be preserved).

Clinical scales, unfortunately, are not devoid of ambiguity. Brain imaging indicates that a few patients in apparent clinical vegetative state may, in fact, present residual consciousness. They exhibit complex and essentially normal cortical responses to speech, as well as a capacity to follow instructions such as "imagine visiting your apartment" (Owen, *et al.*, 2006). Functional magnetic resonance imaging (fMRI) can even be used to communicate with such patients, using very indirect instructions such as "if you want to respond yes, imagine visiting your apartment, otherwise imagine playing tennis", and monitoring the activity of the corresponding brain networks as a proxy for the patient's response (Monti, *et al.*, 2010).

In the near future, there is great hope that the current progress in understanding the signatures of consciousness will lead to easier and more theoretically justified clinical tools. Compared to fMRI, EEG should provide a simpler means to detect rare cases of residual awareness, but also to

improve the diagnosis of all coma and vegetative state patients and to sharpen the prediction of their awakening and future cognitive state. EEG is already used to monitor the depth of propagation of auditory signals in order to predict the recovery of coma patients (Fischer, Luaute, Adeleine, & Morlet, 2004; Kane, Curry, Butler, & Cummins, 1993). In our laboratory, we have developed a paradigm to specifically isolate the P300 wave which is evoked in response to novel auditory signals (Bekinschtein, *et al.*, 2009). In agreement with research in normal subjects, the detection of this wave facilitates the diagnosis of patients with residual awareness and/or imminent recovery (Faugeras, *et al.*, 2011). It is also possible to stimulate the brain with a pulse of external, magnetically induced activity. Again, as theoretically predicted, the duration, complexity, and distance of the propagation to other cortical sites indexes the recovery of consciousness (Rosanova, *et al.*, 2012). Other signatures, such as mathematical indices of the long-distance synchrony between brain areas, may prove to be even more sensitive (King, Dehaene, Naccache *et al.*, in preparation).

Conclusion

The subjective aspects of conscious experience no longer lie beyond the realm of an objective scientific inquiry. On the contrary, a solid body of scientific evidence links consciousness to specific cognitive computations and to the physical state of networks of neurons. Advances in brain imaging now make it possible to reliably detect electrophysiological signatures of consciousness. These signatures can be used to decide, with above-chance accuracy, whether a normal person is or is not aware of a given stimulus, or whether a patient still presents a residual form of consciousness.

While these advances are significant, it should be stressed that they concern primarily the simplest sense of the term "consciousness": the ability to gain conscious access to some information. The brain mechanisms underlying the capacity for self-consciousness (knowing that we know) are only starting to be studied with similar methods (e.g. Fleming, Weil, Nagy, Dolan, & Rees, 2010).

References

Baars, B.J. (1989). *A cognitive theory of consciousness*. Cambridge, Mass.: Cambridge University Press.

Bekinschtein, T.A., Dehaene, S., Rohaut, B., Tadel, F., Cohen, L., & Naccache, L. (2009). Neural signature of the conscious processing of auditory regularities. *Proc Natl Acad Sci U S A*, 106(5), 1672-1677.

Bisiach, E., Luzzatti, C., & Perani, D. (1979). Unilateral neglect representational schema

and consciousness. *Brain*, 102, 609-618.
Blanke, O., Landis, T., Spinelli, L., & Seeck, M. (2004). Out-of-body experience and autoscopy of neurological origin. *Brain*, 127(Pt 2), 243-258.
Blanke, O., Ortigue, S., Landis, T., & Seeck, M. (2002). Stimulating illusory own-body perceptions. *Nature*, 419(6904), 269-270.
Block, N. (1995). On a confusion about a function of consciousness. *Behavioral and Brain Sciences*, 18(2), 227-287.
Chalmers, D. (1996). *The conscious mind*. New York: Oxford University Press.
Churchland, P.S. (1986). *Neurophilosophy: toward a unified understanding of the mind/brain*. Cambridge MA: MIT Press.
Cowey, A., & Stoerig, P. (1995). Blindsight in monkeys. *Nature*, 373(6511), 247-249.
Crick, F., & Koch, C. (1990). Toward a neurobiological theory of consciousness. *Seminars in Neuroscience*, 2, 263-275.
de Lange, F.P., van Gaal, S., Lamme, V.A., & Dehaene, S. (2011). How awareness changes the relative weights of evidence during human decision-making. *PLoS Biol*, 9(11), e1001203.
Dehaene, S., & Changeux, J.P. (2000). Reward-dependent learning in neuronal networks for planning and decision making. *Prog Brain Res*, 126, 217-229.
Dehaene, S., & Changeux, J.P. (2005). Ongoing spontaneous activity controls access to consciousness: a neuronal model for inattentional blindness. *PLoS Biol*, 3(5), e141.
Dehaene, S., & Changeux, J.P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron*, 70, 200-227.
Dehaene, S., Changeux, J.P., Naccache, L., Sackur, J., & Sergent, C. (2006). Conscious, preconscious, and subliminal processing: a testable taxonomy. *Trends Cogn Sci*, 10(5), 204-211.
Dehaene, S., Kerszberg, M., & Changeux, J.P. (1998). A neuronal model of a global workspace in effortful cognitive tasks. *Proc Natl Acad Sci U S A*, 95(24), 14529-14534.
Dehaene, S., & Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework. *Cognition*, 79, 1-37.
Dehaene, S., Naccache, L., Le Clec'H, G., Koechlin, E., Mueller, M., Dehaene-Lambertz, G., *et al.* (1998). Imaging unconscious semantic priming. *Nature*, 395, 597-600.
Del Cul, A., Baillet, S., & Dehaene, S. (2007). Brain dynamics underlying the nonlinear threshold for access to consciousness. *PLoS Biol*, 5(10), e260.
Del Cul, A., Dehaene, S., Reyes, P., Bravo, E., & Slachevsky, A. (2009). Causal role of prefrontal cortex in the threshold for access to consciousness. *Brain*, 132, 2531–2540.
Dennett, D. (1991). *Consciousness explained*. London: Penguin.
Dennett, D. (2001). Are we explaining consciousness yet? *Cognition*, 79(1-2), 221-237.
Descartes, R. (1648/1937). *Traité de l'homme Descartes: Oeuvres et lettres*. Paris: Gallimard.
Dijksterhuis, A., Bos, M.W., Nordgren, L.F., & van Baaren, R.B. (2006). On making the right choice: the deliberation-without-attention effect. *Science*, 311(5763), 1005-1007.
Dunlosky, J., & Metcalfe, J. (2008). *Metacognition*: Sage Publications, Inc.
Eccles, J.C. (1994). *How the self controls its brain*. New York: Springer Verlag.
Faugeras, F., Rohaut, B., Weiss, N., Bekinschtein, T. A., Galanaud, D., Puybasset, L., *et al.* (2011). Probing consciousness with event-related potentials in the vegetative state. *Neurology*, 77(3), 264-268.
Ferrarelli, F., Massimini, M., Sarasso, S., Casali, A., Riedner, B.A., Angelini, G., *et al.* (2010). Breakdown in cortical effective connectivity during midazolam-induced loss of consciousness. *Proc Natl Acad Sci U S A*, 107(6), 2681-2686.

Fisch, L., Privman, E., Ramot, M., Harel, M., Nir, Y., Kipervasser, S., *et al*. (2009). Neural "Ignition": Enhanced Activation Linked to Perceptual Awareness in Human Ventral Stream Visual Cortex. *Neuron*, 64, 562-574.

Fischer, C., Luaute, J., Adeleine, P., & Morlet, D. (2004). Predictive value of sensory and cognitive evoked potentials for awakening from coma. *Neurology*, 63(4), 669-673.

Fleming, S.M., Weil, R.S., Nagy, Z., Dolan, R.J., & Rees, G. (2010). Relating introspective accuracy to individual differences in brain structure. *Science*, 329(5998), 1541-1543.

Frith, C.D. (1979). Consciousness, information processing and schizophrenia. *Br J Psychiatry*, 134, 225-235.

Gaillard, R., Dehaene, S., Adam, C., Clemenceau, S., Hasboun, D., Baulac, M., *et al*. (2009). Converging intracranial markers of conscious access. *PLoS Biol*, 7(3), e61.

Halgren, E., Marinkovic, K., & Chauvel, P. (1998). Generators of the late cognitive potentials in auditory and visual oddball tasks. *Electroencephalogr Clin Neurophysiol*, 106(2), 156-164.

Johansson, P., Hall, L., Sikstrom, S., & Olsson, A. (2005). Failure to detect mismatches between intention and outcome in a simple decision task. *Science*, 310(5745), 116-119.

Kane, N.M., Curry, S.H., Butler, S.R., & Cummins, B.H. (1993). Electrophysiological indicator of awakening from coma. *Lancet*, 341(8846), 688.

Koch, C., & Tsuchiya, N. (2007). Attention and consciousness: two distinct brain processes. *Trends Cogn Sci*, 11(1), 16-22.

Kouider, S., & Dehaene, S. (2007). Levels of processing during non-conscious perception: a critical review of visual masking. *Philos Trans R Soc Lond B Biol Sci*, 362(1481), 857-875.

Lamy, D., Salti, M., & Bar-Haim, Y. (2009). Neural Correlates of Subjective Awareness and Unconscious Processing: An ERP Study. *J Cogn Neurosci*, 21(7), 1435-1446.

Lenggenhager, B., Tadi, T., Metzinger, T., & Blanke, O. (2007). Video ergo sum: manipulating bodily self-consciousness. *Science*, 317(5841), 1096-1099.

Libet, B., Alberts, W.W., Wright, E.W., Jr., & Feinstein, B. (1967). Responses of human somatosensory cortex to stimuli below threshold for conscious sensation. *Science*, 158(808), 1597-1600.

Logan, G.D., & Crump, M.J. (2010). Cognitive illusions of authorship reveal hierarchical error detection in skilled typists. *Science*, 330(6004), 683-686.

Logothetis, N.K., Leopold, D.A., & Sheinberg, D.L. (1996). What is rivalling during binocular rivalry? *Nature*, 380(6575), 621-624.

Mack, A., & Rock, I. (1998). *Inattentional blindness*. Cambridge, Mass.: MIT Press.

Marcel, A.J. (1983). Conscious and unconscious perception: Experiments on visual masking and word recognition. *Cognitive Psychology*, 15, 197-237.

Marti, S., Sackur, J., Sigman, M., & Dehaene, S. (2010). Mapping introspection's blind spot: Reconstruction of dual-task phenomenology using quantified introspection. *Cognition*, 115(2), 303-313.

Massimini, M., Boly, M., Casali, A., Rosanova, M., & Tononi, G. (2009). A perturbational approach for evaluating the brain's capacity for consciousness. *Prog Brain Res*, 177, 201-214.

Monti, M. M., Vanhaudenhuyse, A., Coleman, M. R., Boly, M., Pickard, J. D., Tshibanda, L., *et al*. (2010). Willful Modulation of Brain Activity in Disorders of Consciousness. *N Engl J Med*, 362(7), 579-589.

Naccache, L., Blandin, E., & Dehaene, S. (2002). Unconscious masked priming depends on temporal attention. *Psychological Science*, 13, 416-424.

Naccache, L., Gaillard, R., Adam, C., Has-

boun, D., Clémenceau, S., Baulac, M., *et al*. (2005). A direct intracranial record of emotions evoked by subliminal words. *Proc Natl Acad Sci U S A*, 102, 7713-7717.

Nagel, T. (1974). What is it like to be a bat? *Philosophical Review* 83(4), 435-450.

Nieuwenhuis, S., Ridderinkhof, K.R., Blom, J., Band, G.P., & Kok, A. (2001). Error-related brain potentials are differentially related to awareness of response errors: evidence from an antisaccade task. *Psychophysiology*, 38(5), 752-760.

Nisbett, R.E., & Wilson, T.D. (1977). Telling more than we can know: verbal reports on mental processes. *Psychological Review*, 84(3), 231-259.

Overgaard, M., Rote, J., Mouridsen, K., & Ramsoy, T.Z. (2006). Is conscious perception gradual or dichotomous? A comparison of report methodologies during a visual task. *Conscious Cogn*.

Owen, A.M., Coleman, M.R., Boly, M., Davis, M.H., Laureys, S., & Pickard, J.D. (2006). Detecting awareness in the vegetative state. *Science*, 313(5792), 1402.

Penrose, R. (1990). *The emperor's new mind. Concerning Computers, Minds, and the Laws of Physics*. London: Vintage books.

Penrose, R., & Hameroff, S. (1998). The Penrose-Hameroff "Orch OR" model of consciousness. *Philosophical Transactions Royal Society London* (A), 356, 1869-1896.

Posner, M.I., Snyder, C.R.R., Balota, D.A., & Marsh, E.J. (1975/2004). *Attention and Cognitive Control Cognitive psychology: Key readings*. (pp. 205-223). New York, NY US: Psychology Press.

Rainville, P., Duncan, G.H., Price, D.D., Carrier, B., & Bushnell, M.C. (1997). Pain affect encoded in human anterior cingulate but not somatosensory cortex. *Science*, 277(5328), 968-971.

Raymond, J.E., Shapiro, K.L., & Arnell, K.M. (1992). Temporary suppression of visual processing in an RSVP task: an attentional blink? *J Exp Psychol Hum Percept Perform*, 18(3), 849-860.

Rosanova, M., Gosseries, O., Casarotto, S., Boly, M., Casali, A.G., Bruno, M.A., *et al*. (2012). Recovery of cortical effective connectivity and recovery of consciousness in vegetative patients. *Brain*, 135(Pt 4), 1308-1320.

Sackur, J., & Dehaene, S. (2009). The cognitive architecture for chaining of two mental operations. *Cognition*, 111(2), 187-211.

Sergent, C., Baillet, S., & Dehaene, S. (2005). Timing of the brain events underlying access to consciousness during the attentional blink. *Nat Neurosci*, 8(10), 1391-1400.

Sergent, C., & Dehaene, S. (2004). Is consciousness a gradual phenomenon? Evidence for an all-or-none bifurcation during the attentional blink. *Psychol Sci*, 15(11), 720-728.

Shallice, T. (1972). Dual functions of consciousness. *Psychol Rev*, 79(5), 383-393.

Shanahan, M. (2008). A spiking neuron model of cortical broadcast and competition. *Conscious Cogn*, 17(1), 288-303.

Sigman, M., Sackur, J., Del Cul, A., & Dehaene, S. (2008). Illusory displacement due to object substitution near the consciousness threshold. *J Vis*, 8(1), 13 11-10.

Terrace, H.S., & Son, L.K. (2009). Comparative metacognition. Curr Opin Neurobiol, 19(1), 67-74.

Van Opstal, F., de Lange, F.P., & Dehaene, S. (2011). Rapid parallel semantic processing of numbers without awareness. *Cognition*.

Velly, L.J., Rey, M.F., Bruder, N.J., Gouvitsos, F.A., Witjas, T., Regis, J.M., *et al*. (2007). Differential dynamic of action on cortical and subcortical structures of anesthetic agents during induction of anesthesia. *Anesthesiology*, 107(2), 202-212.

Vorberg, D., Mattler, U., Heinecke, A., Schmidt, T., & Schwarzbach, J. (2003). Different time courses for visual perception and action priming. *Proc Natl Acad Sci U S A*, 100(10), 6275-6280.

Weiskrantz, L. (1986). *Blindsight: A Case Study and Its Implications*. Oxford: Clarendon Press.

Zylberberg, A., Fernandez Slezak, D., Roelfsema, P. R., Dehaene, S., & Sigman, M. (2010). The brain's router: a cortical network model of serial processing in the primate brain. *PLoS Comput Biol*, 6(4), e1000765.

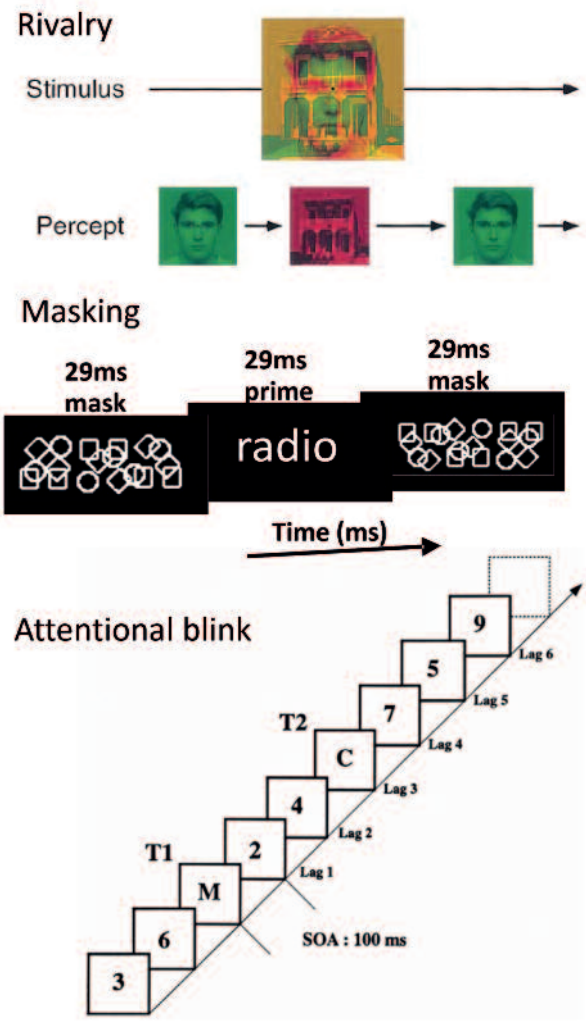

Figure 1. Examples of experimental paradigms to manipulate conscious perception. In rivarly, distinct images are presented to the two eyes, yet subjective perception alternates between seeing one and seeing the other. In masking, a visible word is made invisible by surrounding in time it with shapes that mask it. In the attentional blink, processing of a first target T1 prevents the perception of a second target T2.

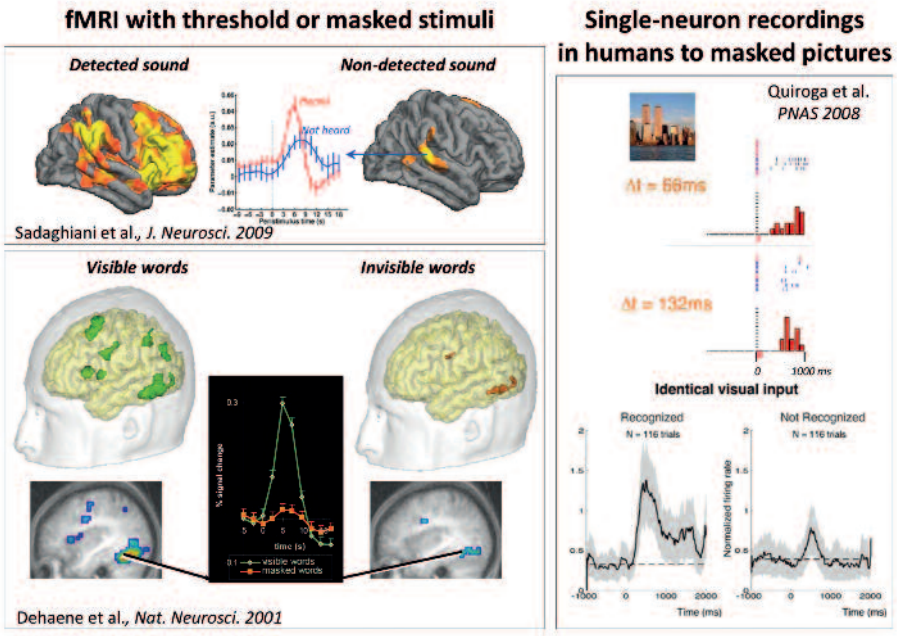

Figure 2. Converging evidence for cerebral signatures of consciousness. Conscious perception, compared to non-conscious processing, systematically involves a late and long-lasting "ignition": sensory activation is amplified and expands into a broad set of associative areas of the prefrontal, parietal and temporal lobes.