Available online at www.sciencedirect.com
ScienceDirect
Current Opinion in Neurobiology

Toward a computational theory of conscious processing

Stanislas Dehaene[1,2,3,4], Lucie Charles[2,3,4], Jean-Rémi King[2,3,4] and Sébastien Marti[2,3,4]

The study of the mechanisms of conscious processing has become a productive area of cognitive neuroscience. Here we review some of the recent behavioral and neuroscience data, with the specific goal of constraining present and future theories of the computations underlying conscious processing. Experimental findings imply that most of the brain's computations can be performed in a non-conscious mode, but that conscious perception is characterized by an amplification, global propagation and integration of brain signals. A comparison of these data with major theoretical proposals suggests that firstly, conscious access must be carefully distinguished from selective attention; secondly, conscious perception may be likened to a non-linear decision that 'ignites' a network of distributed areas; thirdly, information which is selected for conscious perception gains access to additional computations, including temporary maintenance, global sharing, and flexible routing; and finally, measures of the complexity, long-distance correlation and integration of brain signals provide reliable indices of conscious processing, clinically relevant to patients recovering from coma.

Addresses
[1] Collège de France, F-75005 Paris, France
[2] Cognitive Neuroimaging Unit, Institut National de la Santé et de la Recherche Médicale, U992, F-91191 Gif/Yvette, France
[3] NeuroSpin Center, Institute of BioImaging Commissariat à l'Energie Atomique, F-91191 Gif/Yvette, France
[4] Université Paris 11, Orsay, France

Corresponding author: Dehaene, Stanislas (stanislas.dehaene@gmail.com)

Current Opinion in Neurobiology 2014, **25**:76–84

This review comes from a themed issue on **Theoretical and computational neuroscience**

Edited by **Adrienne Fairhall** and **Haim Sompolinsky**

For a complete overview see the Issue and the Editorial

Available online 29th December 2013

0959-4388/$ – see front matter,

http://dx.doi.org/10.1016/j.conb.2013.12.005

Introduction

Consciousness is the only real thing in the world and the greatest mystery of all

Vladimir Nabokov, *Bend Sinister* (1947)

What brain mechanisms underlie our capacity to become aware of a specific piece of information, while many others remain non-conscious? Considerable empirical and theoretical progress has been made lately in answering this deceivingly simple question. This research gained leverage when it was recognized that visual illusions [1–3] and a great variety of other normal and pathophysiological conditions such sleep, anesthesia, blindsight or hemineglect provided empirical windows into this phenomenon, by providing minimal contrasts between conscious and non-conscious brain states [4]. Here we review the recent advances made possible by this contrastive approach. We specifically focus on how these findings inform present-day theories of conscious processing. At present, there is no accepted computational theory of this function. Our hope is that the present review may point to the key ingredients that will lead to one.

Defining the terms

It is useful to start by separating the diversity of concepts that the everyday term of 'consciousness' can refer to. The *content of consciousness* refers to the specific information that I am aware of at a given moment. For instance, I am currently aware of reading these words, but not of the music playing in the background (until I attend to it). *Conscious access* is the process by which a piece of information becomes a conscious content. *Conscious processing* refers to the various operations that can be applied to a conscious content (as when multiplying two numbers mentally). *Conscious report* is the process by which a conscious content can be described, verbally or by various gestures. Such *reportability* remains the main criterion for whether a piece of information is or is not conscious: by hypothesis, I can report something if and only if I am aware of it.

A great variety of representations can be consciously accessed, including perceptual states, abstract knowledge, memories, plans, and other internal states (e.g. feelings, confidence, and errors). *Self-consciousness* is a particular instance of conscious access where the conscious 'spotlight' is oriented toward internal states.

The *state of consciousness*, associated with fluctuations in *wakefulness* or *vigilance*, finally, refers to the brain's very ability to entertain a stream of conscious contents. During normal wakefulness, any information may be consciously accessed, but this ability is continuously modulated according to the level of vigilance, and ultimately vanishes during coma, vegetative state, anesthesia or deep sleep. Although this review concentrates primarily on the mechanisms of conscious access and conscious processing, in a final section, we consider how what has been learned about conscious access in normal subjects generalizes to the detection of the state of consciousness in brain-lesioned patients.

The boundaries of non-conscious processing

To clarify the nature of conscious processing, a first step consists in delineating what it is *not*. Using masking [5], crowding [6], inattention [7] or binocular rivalry [8], images can be presented under conditions such that they remain strictly invisible. Behavioral priming and brain imaging can then reveal how deep these stimuli are processed. Studies of non-conscious processing play an instrumental role in refuting specific theories of consciousness. The logic is simple: if a cognitive computation or neural marker, proposed by some theory to be uniquely associated with conscious processing, can be observed under demonstrably non-conscious conditions, then that theory is severely undermined.

Twenty years of research indicates that subliminal processing can be quite deep. Many cortical areas can be activated by an unseen stimulus, including areas of the visual ventral [9] and dorsal pathways [10]. The brain non-consciously recognizes the abstract identity of pictures, words and faces [9,11,12•], the quantity attached to a number symbol [10,13], the fact that two words are related or synonymous [6,14,15], the emotional meaning of a word [16•,17], or the reward value of a coin or an arbitrary symbol [18,19,20••].

In recent years, the frontiers of non-conscious processing have been pushed further. For instance, in chess experts, a brief non-conscious flash of a chessboard suffices to determine whether the king is in check [21]. Within the language domain, the grammatical fit of a masked word with the preceding sentence can be determined non-consciously [22•]. Transitive inferences can also be deployed non-consciously: after non-conscious exposure to arbitrary word pairs such as 'winter-red' and 'red-computer', word association effects generalize to non-adjacent pairs ('winter-computer'), a transitive link mediated by the hippocampus [23]. As another example of high-level computation, the approximate average of four masked numbers can be extracted non-consciously [13]. There is even a suggestion that multi-step operations such '9 − 5 + 2' may be mediated non-consciously [24], although this conclusion will require better control over the stimuli and the degree of non-consciousness.

All in all, these findings refute the idea that non-conscious processing stops at an early perceptual level: meaning and value can clearly be assigned non-consciously. There is also considerable evidence that attention can be deployed and enhance processing even if its target remains non-conscious [25–27]. At the brain level, attending to a stimulus and becoming conscious of it have distinct signatures that occur on distinct trials and at different times [28–30]. For instance, by orthogonally manipulating visibility and attention (using masked images presented at the threshold for conscious perception such that half were visible and half were invisible, and preceding them by valid or invalid attentional cues), Wyart and colleagues [29] found a double dissociation: attention, but not visibility, modulated early occipital activity, while visibility, but not attention, modulated later temporal and parieto-frontal activity. Under some circumstances, greater spatial attention may even lead to a reduced visibility [31••]. These findings refute theories that conflate attention and consciousness. William James' classical definition of attention ('the taking possession by the mind, in clear and vivid form, of one out of what seem several simultaneously possible objects or trains of thought') mixes up conscious access proper ('taking possession of the mind') with selection ('one out of several') which can be fully non-conscious. Selective attention may facilitate conscious access, even when the attentional cue comes long after the stimulus is gone [32•], but it operates largely non-consciously.

Recent findings also invalidate the idea that the central executive, which controls our strategies and inhibits unwanted behaviors, always operates consciously. A series of experiments with the go/no-go paradigm indicate that an unseen visual cue can trigger inhibitory control circuits in the pre-supplementary motor area and anterior insula [33,34,35•,36]. Error detection [37•,38••] and task switching [39•,40], which are typical executive functions, can be triggered non-consciously. Even the maintenance of a stimulus in working memory may remain above the chance level for subliminal stimuli [41•] — although this recent finding will need to be reconciled with the more frequent observation that subliminal priming drops to chance level after a second or less [42–44].

Overall, these findings support the view that virtually any cerebral processor may operate in a non-conscious mode. They challenge theories that associate conscious processing with a specific cognitive processor. For instance, the hypothesis that conscious perception coincides with the ability to deploy higher-order thoughts or metacognition (the brain's ability to represent its own knowledge states) [45] does not bode well with evidence that self-monitoring, error detection and confidence assignment partially operate non-consciously [38••,46•,47].

Findings from subliminal research also eliminate some physiological theories of conscious processing. It is now clear that early changes in gamma band power (>30 Hz), once postulated as a marker of consciousness, can be evoked by a non-conscious stimulus [48••,49••] and do not faithfully track variations in subjective reports [50]. Similarly, the views that recurrent interactions [51,52] and information integration [53,54] are necessary and