Neurobiological Theories of Consciousness

S Kouider, Laboratoire des Sciences Cognitives et Psycholinguistique, CNRS/EHESS/ENS-DEC, Paris, France

Glossary

Neural correlates of consciousness – They are defined by Christoph Koch as "The minimal set of neuronal mechanisms or events jointly sufficient for a specific conscious percept or experience." They allow to avoid the difficult problem of directly looking for neural bases.

Panpsychism – Reflects the philosophical doctrine that everything (in Greek, 'pan') has a mind ('psyche') and is therefore conscious. Some theories presented in this article endorse a certain form of panpsychism in which anything that transmits information is in a way conscious.

The hard problem – It is the problem of explaining how and why we have the subjective experience of consciousness. It is often contrasted with the easy problem, which consists of describing consciousness as the cognitive ability to discriminate, integrate information, focus attention, etc.

Introduction

Neuroscientists working on the issue of consciousness consider that it is a biological problem. They assume that we will understand how and why we are conscious by studying the cerebral and neuronal features of the brain. These theories have largely benefited from the recent advances in neuropsychology, neurophysiology, and brain imaging in particular. However, neurobiologists have also been influenced, on the one side, by cognitive theories aimed at characterizing the psychological determinants of consciousness, and on the other side, by philosophical issues related to the mind–body problem.

Cognitive Influences on Neurobiological Accounts

Regarding the influence of cognitive theories, the majority of neurobiological accounts can be seen, in fact, as extensions of preexisting cognitive theories (e.g., for instance global workspace theories). Indeed, one of the main tasks exercised by neurobiologists in the last two decades has been to search for cerebral or neuronal equivalents to the functional elements constituting cognitive models (e.g., the dorsolateral prefrontal cortex for voluntary control, or long range axons for connecting brain regions associated with 'unconscious' and 'conscious' processing). Of course, many neurobiologists disagree with this approach. Consciousness, because it is a biological problem, should be reframed the other way around, by focusing primarily on its structural basis rather than relying on cognitive theories, often considered too speculative. Therefore, many neurobiologists consider that an ideal neurobiological science of consciousness should focus on neural structures and mechanisms in order to understand how the organic matter constituting the brain creates consciousness.

The Hard Problem for a Neurobiology of Consciousness

Yet, studying the neural mechanisms 'leading to' consciousness, trying to explain the 'emergence' of consciousness, or focusing on how the brain 'creates' consciousness, as often described in neurobiological literature, sounds as if it involved an immaterial soul that would magically arise from the brain. This is not a new issue for philosophers who have also been wondering about the equivalent mind–body problem since antiquity. More than a century ago, the contemporary 'brain-consciousness' problem was well captured by Thomas Huxley's famous remark: "How it is that anything so remarkable as a state of consciousness

comes about as a result of irritating nervous tissue, is just as unaccountable as the appearance of the djinn when Aladdin rubbed his lamp in the story." The same issue applies today: understanding consciousness as an 'emergent' property 'arising' from functional elements of the neurocognitive architecture, without falling on a dualistic position where consciousness lies somewhere outside of the brain, poses serious epistemological difficulties and leads to the so-called hard problem of consciousness.

p0020 Indeed, many philosophers have concluded that there is not one single problem, but actually two problems that are faced by anyone trying to understand consciousness: they distinguish between the so-called easy problem and hard problem. In a nutshell, the easy problem consists in relying on objective measures of conscious processing in order to explain how one is able to discriminate sensory events, integrate information, report mental states, focus attention, etc. By contrast, the hard problem consists in explaining the first-order, subjective nature of qualias and phenomenal states, the 'what is it like to be conscious' as well as how and why we experience consciousness at all. Addressing the hard rather than the easy problem of consciousness constitutes an important epistemological constraint put forward by philosophers. In particular, contemporary philosophers such as Joseph Levine and later David Chalmers have argued that trying to resolve the hard problem leads to an 'explanatory gap' that science is unable to cross, at least today. Indeed, they stress the fact that it appears impossible to demonstrate that a neural structure leads to a conscious experience, while denying the reverse possibility. In addition, given how different they are, reducing phenomenal states to neural states appears almost impossible.

s0020 Looking for Neural Correlates, Not Neural Bases

p0025 Should neurobiologists then give up on addressing this issue? Most neurobiologists acknowledge the existence of a hard problem. However, they also endorse the principle that further scientific investigations will ultimately allow us to resolve it. Others explicitly deny the existence of a hard problem in the Chalmerian sense. For some, assigning too much importance to the explanatory gap might actually turn out to be counterproductive and impedes rather than facilitates scientific progress. Accordingly, neurobiologists have mostly focused on the easy problem, considering that this strategy will progressively get us closer to understanding the full issue. They extended the 'contrastive analysis,' originally put forward by Bernard Baars, from the cognitive to the neurobiological domain. While this method initially consisted in contrasting conscious and unconscious processes in order to characterize their cognitive features, the neurobiological approach aims at characterizing the neural features. A typical example consists in comparing the cerebral activity when subjects are presented with subliminal stimuli they cannot report (unconscious processing) with that of visible stimuli they can report (conscious processing).

p0030 In other terms, the current first step in trying to understand the link between consciousness and the brain consists in finding out which neural components are specifically involved during conscious processing, but importantly not during unconscious processing. Francis Crick and Christoph Koch have coined the term 'neural correlates of consciousness' (NCC; see Glossary) in order to describe this epistemological approach. According to them, the best strategy for a neurobiological science of consciousness is to search for the NCC. Underneath this approach is the crucial principle that 'correlates' do not imply any relation of causality between the occurrence of conscious mental events and their associated physiological structure. Consequently, this strategy has the advantage of leaving aside, at least for the moment, the hard problem of finding the neural 'bases' of consciousness.

p0035 In the following sections, I will provide an overview of the current most influential neurobiological theories of consciousness. These theories will be largely described in an independent manner, such that each of them can be understood individually, that is, without having to frame it in the context of alternative accounts. Only later, in the section labeled 'Neurobiological standpoints on the hard problem' will I evaluate their explanatory power by confronting them in relation to some important conceptual issues (e.g., dissociating access vs. phenomenal consciousness, dissociating attention vs.

consciousness, panpsychism). I will conclude by emphasizing how promising these theories are in getting us closer to resolve the issue of the hard problem.

s0025 ## From Globalist to Localist Accounts of Consciousness

p0040 Neurobiological theories of consciousness differ in many respects. One way to portray them in a coherent manner is to follow the large spectrum ranging from globalist to minimally localist accounts. By globalist or localist I refer to the size of the brain states that are assumed to be sufficient for consciousness (extended to large parts of the brain vs. focalized to specific and small brain areas, respectively). I will mainly restrict this review on the cerebral level and exclude alternative theories that are more globalist, in the sense that consciousness encompasses more than neural activity in the brain (e.g., theories proposing that consciousness reflects modification in the electromagnetic field surrounding the brain, as proposed by J. McFadden and by R. John), or, conversely, theories that are more localist, by focusing on single neurons or even lower structural levels (e.g., the quantum-level theory of microtubules by S. Hamerof and R. Penrose). Such theories remain excessively speculative and unspecified to be included in a serious review, at least for the moment.

s0030 ## The Reentrant Dynamic Core Theory

p0045 The Reentrant Dynamic Core theory, proposed by Gerald Edelman and Giulio Tononi, is arguably the most globalist account of consciousness. Indeed, in this framework, consciousness is not to be localized in dedicated brain areas or with particular types of neurons. Rather, it is the result of dynamic interactions among widely distributed groups of neurons in the entire thalamocortical network (the 'dynamic core'). This theory offers an interesting tentative of unifying the hard and easy problems, providing a neurobiological explanation for qualias, and explaining both phylogenetic and ontogenetic aspects of the development of consciousness in humans and other species. Yet, although this theory is very appealing, especially given its explanatory power, it is also highly speculative, and based on several assumptions that remain to be demonstrated. I will further discuss the speculative aspects of this theory later (see the section 'Neurobiological standpoints on the hard problem'). For now, I shall provide an overview of the core assumptions underlying this theory.

p0050 In order to appreciate the specificity of the Reentrant Dynamic Core theory, it is important to understand that, regardless of its explanation for consciousness, it offers an alternative view on brain structures, considering the wiring of the brain into neuronal assemblies as the result of variation and selection mechanisms that are analogous to those underlying evolutionary theories. This macrolevel account of brain development, formally developed by Edelman in 1978, and called the theory of Neuronal Group Selection (also called the theory of Neural Darwinism) constitutes a key element for understanding the development of consciousness. According to the Neuronal Group Selection theory, the brain is assumed to be a selectionist system in which variant groups of neurons are selected over others in three steps. The first one, termed developmental selection, happens during embryogenesis and early development, and it is largely influenced by epigenetic factors. It consists of several processes such as cell division, cell death, extension, which has the consequence of connecting neurons together into a large number of variant neuronal circuits (labeled primary repertoires), and elimination, which targets unconnected neurons in particular. The second step is called experiential selection and lasts from early infancy through all of the lifespan. It consists of reinforcing, through the influence of behavior and experience with the environment, the synaptic connections of some variants over others, leading to secondary repertoires of neurons. The final step, that of reentry, consists of the formation of massively parallel reciprocal connections among distant maps of neuronal repertoires, allowing them to exchange signals and be spatiotemporally coordinated.

p0055 In the Reentrant Dynamic Core theory, consciousness results exactly from this mechanism of reentry among distant groups of neurons within the dynamic core of thalamocortical connections. The spatiotemporal coordination provided by reentry allows the binding of several elements into a single and coherent object or event, providing a solution to the binding problem. It also allows to

explain why conscious experiences appear to be unified.

p0060 Particularly important in this theory is the reentry of information between groups of neurons dealing with perceptual categorization (i.e., in posterior areas) and the more frontally located systems responsible for value-category memory. Indeed, the latter will constrain the selectionist process by modulating or altering the synaptic connections within groups of neurons, as a consequence of their influence on behavior (pleasure, pain, etc.). Another key aspect to this whole framework is that groups of neurons are constantly in competition and their survival depends on creating or reinforcing synaptic links with other groups, such as to form large assemblies of reentrant neuronal maps in the dynamic core. The victorious assembly (or 'coalition' in Crick and Koch's terms) will lead to consciousness, at least for a few hundred milliseconds, until a new coalition of neurons bypass it. Indeed, because the variant groups of neurons are assumed to be degenerate, which means that different variants of neuronal assemblies can actually carry out the same function and have the same output, each integrated state in the dynamic core is followed by yet another and differentiated neural state in the core. As such, coalitions of neurons leading to consciousness are temporally transient by nature and widespread along variant regions of the whole dynamic core. Because interactions within the dynamic core are constantly moving, it explains the diversity of consciousness, yet the constant integration through reentry explains the unity of consciousness. Note then that the victory of a coalition of neurons in leading to consciousness is not just a matter of its size; rather it depends on its complexity, that is, its ability to generate at the same time an integrated scene (i.e., appearing as a unitary event) and a differentiated scene (i.e., which can be highly discriminated from other scenes). Measures of complexity in relation to consciousness have been further developed by Tononi in his recent Information Integration theory of consciousness (see the article 'cognitive theories of consciousness' in this volume). According to Edelman and Tononi, these highly discriminatory properties of neural complex systems are the qualias that have been torturing philosophers, nothing more, nothing less! Finally, they distinguish between primary consciousness that allows for a perceptual organization of the environment and whose characteristics have been presented above, and higher-order consciousness that is possessed by humans, and which is related to linguistic and symbolic mental activities. The latter requires further reentry with additional brain regions such as those involved in language production and comprehension.

p0065 In sum, this theory assumes that the brain is a dynamic complex system in which consciousness emerges from the interactivity itself. Instead of involving top-down communication between dedicated areas, the Reentrant Dynamic Core theory involves regions in the cortex in active communication with one another and with associated nuclei in the thalamus. Consciousness arises from the differentiated integration of activity in these areas, as information is transmitted recurrently, with local groups of neurons performing their specialized and discriminatory function, while at the same time being unified with other neuronal groups of the dynamic core. Before concluding on the characteristics of this theory, it is important to point out that although the Reentrant Dynamic Core can be considered as a globalist theory, in the sense that it can involve a large set of regions in the brain, it is also a gradual theory of consciousness that can result from minimal neural networks. Indeed, reentry among even a small number of neurons, as far as it reflects differentiated integration, will induce consciousness to a certain degree. We will come back to this theoretical aspect below.

The Global Neuronal Workspace Theory

s0035

p0070 The Global Neuronal Workspace theory proposed by Stanislas Dehaene with Lionel Naccache and Jean-Pierre Changeux is currently the most explicit and most functional account of the cerebral architecture underlying conscious access. Its functionality has rendered this theory very popular in neuroscience circles. Yet, as we shall see below in the section 'Neurobiological standpoints on the hard problem,' this theory has also been greatly criticized for probing functionality at the price of sacrificing some phenomenological aspects of consciousness. For now, let us focus on the main characteristics of this theory.

p0075 This theory is a perfect example of the neurobiological extension of a cognitive theory, that of a global workspace, originally proposed by Bernard Baars in 1988 (see the article 'Cognitive theories of consciousness' in this volume). Dehaene's theory assumes that neurocognitive architecture is composed of two qualitatively distinct types of elements. The first type is represented by a large network of domain-specific processors, in both the cortical and subcortical regions that are each attuned to the processing of a particular type of information. For instance, the occipitotemporal cortex is constituted of many such domain-specific processors, or 'cerebral modules,' where color processing occurs in V4, movement processing in MT/V5, face processing in the fusiform face area (FFA), etc. Although these neural processors can differ widely in complexity and domain specificity, they share several common properties: they are triggered automatically (i.e., mandatorily, by opposition to voluntarily), they are encapsulated (their internal computations are not available to other processors), and importantly, they largely operate unconsciously.

p0080 Conscious access involves only the second type of element, namely, the cortical 'workspace' neurons that are particularly dense in the prefrontal, cingulate, and parietal regions. These neurons are characterized by their ability to send and receive projections to many distant areas through long-range excitatory axons, breaking the modularity of the nervous system and allowing the domain-specific processors to exchange information in a global and flexible manner. The global workspace is thus a distributed neural system with long-distance connectivity that can potentially interconnect multiple cerebral modules through workspace neurons. It offers a common communication protocol, by allowing the broadcasting of information to multiple neural targets.

p0085 One important aspect of this theory is that encapsulation and automaticity are less rigid than traditional modularist (i.e., Fodorian) accounts of the cognitive system (see the article 'Cognitive theories of consciousness' in this volume). Indeed, once a set of processors have started to communicate through workspace connections, this multi-domain processing stream also becomes more and more 'automatized' through practice, resulting in direct interconnections, without requiring the use of workpace neurons and without involving consciousness. Another important aspect that follows from this theory is that information computed in neural processors that do not possess direct long-range connections with workspace neurons, will always remain unconscious. This idea is rooted in the work of Crick and Koch, postulating that neural activity in V1, because it does not project toward prefrontal neurons, does not participate directly in visual consciousness.

p0090 Note that for a mental object to become conscious, it is not sufficient that its activity gives input to the global workspace. Two other conditions have to be met. One is that the content of the mental object must be represented as an explicit firing pattern of neuronal activity, that is, a group of neurons that unambiguously indexes its relevant attributes. A final and important condition is that the top-down amplification mechanisms mobilizing the long-distance workspace connections render the content of consciousness accessible, sharpened, and maintained. A mental object, even if it respects the two first conditions (explicit firing and accessibility to workspace neurons) will still remain buffered in a 'preconscious' (and thus a nonconscious) store until it is attended to and its neural signal amplified. Therefore, top-down attention in this framework is a necessary condition for consciousness. Whether being conscious requires top-down attention constitutes, as we will see below, one of the most debated aspects among neurobiological theories of consciousness.

The Duplex Vision Theory

s0040

p0095 The Duplex Vision theory proposed by David Milner and Melvyn Goodale postulates that visual perception involves two interconnected, but distinctive pathways in the visual cortex, namely, the dorsal and the ventral stream. After being conveyed along retinal and subcortical (i.e., geniculate) structures, visual information reaches V1 and then involves two streams. The ventral stream projects toward the inferior part of the temporal cortex and, according to this theory, serves to construct a conscious perceptual representation of objects, whereas, the dorsal stream projects toward the posterior parietal cortex and mediates

the control of actions directed at those objects. Apart from these structural considerations, the two streams also differ at the computational and functional levels. On the one side, the ventral stream conveys information about the enduring (i.e., long-lasting) characteristics that will be used to identify the objects correctly, and subsequently to link them to a meaning and classify them in relation to other elements of the visual scene. Computing these enduring characteristics involves relatively long and costly computations. On the other side, the dorsal stream can be regarded as a fast and online visuomotor system dealing with the moment-to-moment information available to the system, which will be used to perform actions in real time.

p0100 It should be noted that this dissociation between ventral and dorsal pathways has sometimes been misunderstood as equivalent to conveying the 'what' and 'where' information in the visual cortex, respectively. However, structural and spatial attributes can conjointly be used by both systems. For instance, visual information such as the size, geometrical structure, and location of a target object might be computed both by the dorsal steam, in order to grasp and reach the object, and by the ventral stream in order to segregate the object from a complex visual scene. While the ventral steam involves object perception by comparison with visual perceptual attributes stored in memory, the dorsal steam, because it is fast and updated in real time, involves no storage of the visuomotor attributes extracted from the object, nor the motor program resulting from actions upon that object.

p0105 In other terms, while the ventral stream can be seen as conveying vision for perception, the dorsal stream is concerned with vision for action. Both systems work together in the production of adaptative behavior, but they can also function independently as revealed by clinical studies. Indeed, the development of this theory mainly results from neuropsychological investigations, demonstrating a double dissociation between vision for perception and vision for action. On the one side, patients with a lesion in the posterior parietal cortex, the area terminating the visual dorsal stream, suffer from 'optic ataxia,' a deficit in the control of reaching and grasping objects. Despite this deficit, these patients are perfectly able to verbally describe the unreachable objects, either as a whole or in terms of their attributes. In other terms, they can use vision for consciously perceiving objects in their environment, but not for controlling real-time actions directed at those objects. On the other side, patients with damage to a ventral region known as the lateral occipital complex are unable to recognize everyday objects or even simple geometric forms, a deficit labeled 'visual form agnosia.' Yet, such patients are strikingly efficient at grasping objects correctly (for instance by opening their hand as a function of the object size, or by rotating their hand according to the object orientation). Although such patients cannot consciously perceive the object and even its visual attributes (size, shape, or orientation), they can use the same object attributes to control their object-directed actions.

p0110 One important consequence of this theory, according to Milner and Goodale, is that because the processes performed by the dorsal stream are very fast and largely automatic, they are largely unconscious, while, by contrast, those conveyed by the ventral stream are assumed to constitute the core of conscious perception. The visual phenomenology generated in ventral regions will in turn be transferred to working memory components in order to use information off-line, when the objects are not stimulating the visual system anymore. According to this theory, although we are typically aware of the actions we perform, no phenomenology is associated with the visual information used by the dorsal stream to control those actions. Hence, the neural computations performed in the dorsal stream remain quite inaccessible to consciousness.

p0115 It is of note that when this theory was proposed more than a decade ago, evidence from binocular rivalry (in which conscious perception alternates between two images, say a face and a house presented to each eye) revealed a very strong correlation with conscious perception in the ventral stream (e.g., in the FFA for faces). Therefore, this evidence misled many researchers at that time to deny the possibility of unconscious perception in the ventral stream. However, Geraint Rees and colleagues found that patients with unilateral neglect, a deficit in which they fail to pay attention and then report stimuli on half of their visual field, still exhibit FFA activity for faces in the neglected field. Since neglected stimuli are perfectly reported when cued carefully, neglect is considered as an inability to

report efficiently because of a lack of attention. Given the possibility that consciousness and attention might be distinct (see below), it still remains unclear whether. Rees's finding reflected ventral neural activity without consciousness or just without attention. Yet, more recent evidence with visual masking, obtained in my laboratory has revealed unconscious neural activity in ventral regions, including the FFA, during subliminal face perception, thus clearly demonstrating that the ventral stream is not exclusively related to conscious perceptual processes.

p0120 This type of evidence is problematic for the Duplex Vision theory, since this theory predicts that conscious perception should be proportional to neural activity in the ventral stream. Although the possibility of unconscious ventral processing was not taken into account in the original theory, it can be accommodated by assuming a threshold mechanism, as proposed by Zeki for the Minimal Consciousness theory reviewed below. However, including this threshold leads the theory to lose it former appeal, since consciousness is 'only partially' correlated with activity in the ventral stream. Conversely, various recent evidences have shown that the dorsal stream can, under some circumstances, be associated with the consciousness of actions. Thus, although this theory might be effective for distinguishing the neural mechanisms that are responsible for vision for perception and vision for action, this dissociation might turn out to be orthogonal to the dissociation between conscious and unconscious processing.

s0045 The Local Recurrence Theory

p0125 The Local Recurrence theory put forward by Viktor Lamme is mostly concerned with vision for perception rather than action. It distinguishes between three hierarchical types of neural processes related to consciousness. The first stage involves a 'feedforward sweep' during which the information is fed forward from striate visual regions (i.e., V1) toward extrastriate areas as well as parietal and temporal cortices, without being accompanied by any conscious experience of the visual input. Only during the second stage, involving the 'localized recurrent processing,' is the information fed back to the early visual cortex. It is these recurrent interactions between early and higher visual areas which, according to this theory, lead to visual experience (i.e., phenomenal consciousness, see below). The third and final stage consists of 'widespread recurrent processing,' which involves global interactions (similar to the global workspace model) and extends toward the executive (i.e., the frontoparietal network) and language areas. This final step also involves the attentional, executive, and linguistic processes that are necessary for conscious access and reportability of the stimulus.

p0130 An interesting aspect of the Local Recurrence theory is that it offers an explanation for the difference between conscious and unconscious perception in mechanistic rather than in architectural terms. Here, the distinction between subliminal and conscious perception involves the same regions. Subliminal perception reflects the fact that the visual signal is fed forward to higher visual areas without being fed back to early areas (for instance a second stimulus replaces the first one in V1 and then prevents the setting up of recurrence between higher regions and V1). Another interesting aspect of this theory, although provocative and speculative, is that consciousness should not be defined by behavioral indexes such as the subject's introspective reports. Instead, according to Lamme, one should rely on neural indexes of consciousness, one of which is neural recurrence. Indeed, Lamme is concerned with the question of defining phenomenological consciousness when a report is impossible. According to Lamme's theory, involvement of the second step (local recurrence) without involving the third step (global recurrence) represents exactly this situation. We will come back to this aspect when discussing the neurobiological standpoints on the hard problem.

p0135 In sum, the theory of Local Recurrence explicitly stipulates that recurrent activity is sufficient for consciousness. Yet, one main difficulty with this theory is that is fails to take into account the recurrent connections that exist between regions that are not associated with consciousness (for instance between V1 and the thalamus). It remains possible that consciousness involves local recurrence 'between some specific cerebral components.' However, local recurrence cannot then be considered as a sufficient condition for consciousness anymore, since it requires the involvement of

additional mechanisms for explaining why it only applies to a restricted set of brain regions.

The Microconsciousness Theory

p0140 The microconsciousness theory put forward by Semir Zeki and colleagues is arguably the most localist account of the cerebral architecture underlying consciousness. It is assumed in this theory that instead of a single consciousness, there are multiple consciousnesses that are distributed in time and space. This theory has also been mainly developed in the context of vision research and reflects the large functional specialization of the visual cortex. For instance, evidence from various sources (clinical, brain imaging, etc.) have shown that while the perception of colors is associated with neural activity in area V4, motion perception reflects neural activity in MT/V5. In particular, neuropsychological investigations have demonstrated that the respective cerebral lesions in these two functional sites lead to dissociated forms of conscious perception. Lesions in V4 lead to 'achromatopsia,' the inability to see the world in colors, while motion remains intact, and conversely lesions in MT/V5 result in visual 'akinetopsia,' the inability to perceive visual motion, while color perception remains unaffected. Furthermore, because of the existence of direct connections between subcortical structures and MT/V5 (i.e., without having to be mediated by V1), patients with a lesion in V1 are unable to perceive objects, while they can still experience motion when these objects are moving.

p0145 Zeki takes these findings as evidence that consciousness is not a unitary and singular phenomenon, but rather that it involves multiple consciousnesses that are distributed across processing sites (also called essential nodes in this theory), which are independent from each other. Another form of evidence in favor of this theory is that the conscious perception of different attributes is not synchronous and can respect a temporal hierarchy. For instance, psychophysical measures have shown that color is perceived a few tens of milliseconds before motion, reflecting, according to Zeki, the fact that neural activity during perception reaches V4 before reaching MT/V5. This observation is congruent with the microconsciousness theory, where it is postulated that microconsciousnesses are not only distributed in space, but also in time.

p0150 A critical characteristic of the microconsciousness theory is that it considers these processing sites to be equivalent to perceptual sites, which means that conscious perception of one specific attribute (e.g., color) is proportional to the strength of activity in its respective processing site (i.e., V4). According to this theory, a microconsciousness associated with a processing site does not necessitate top-down influences from higher (i.e., frontal) areas although these regions might play a role in visual consciousness. Note that although the correlation between neural activity in a processing site and conscious perception is predicted to be highly positive, it cannot be perfect in this theory. Indeed, in order to take into account some clinical and experimental evidences revealing unconscious processing, notably those obtained by Zeki himself, neural activity in a processing site must reach a certain height for a conscious correlate to be generated. Therefore, this theory does not deny the existence of unconscious processing. In fact, a consequence of this postulate is that it obviates the need to separate brain regions that are linked to consciousness from those that are linked to unconscious processing, as we saw above if the Local Recurrence theory. The transition between unconscious and conscious perception reflects the crossing of a threshold 'within' the processing site, though the neural and behavioral characteristics of this transition remain to be specified.

p0155 Note that this theory does not deny that the attributes of each processing site are bound together at one point. However, it assumes that binding is a post-conscious process occurring 'after' consciousness of the specific attributes to be bound together has taken place. This second step is termed 'macroconsciousness.' Although microconsciousness involves only one perceptual attribute (e.g., color), macroconsciousness reflects the phenomenal experience associated with the bound object (i.e., including its form, color, motion, etc.). It occurs higher and later in the hierarchy, and depends upon the presence of the previous one. In addition, Zeki proposes that there is also a third and last level coined as 'unified consciousness,' reflecting a more global form of consciousness,

which involves linguistic and communication skills. This third level remains largely unspecified, but it is roughly equivalent to the brain regions leading to consciousness in the global workspace model.

p0160 One main difficulty with this theory is that any processing region in the brain should, at first glance, constitute an NCC in the multiple-consciousness theory. As such, it remains unclear why conscious perception is not associated with activity in most brain regions, including the cerebellum and subcortical regions, especially those conveying visual signals (e.g., the Lateral Geniculate Nucleus). Zeki takes this problem into account, at least by admitting it, and tries to solve it by explaining that neural activity in a processing site probably also necessitates the involvement of additional systems, in particular the reticular activating system maintaining arousal. Yet, the theory loses its force, since neural activity, in certain perceptual sites, is not a sufficient condition for microconsciousness anymore (not mentioning the fact that the reticular system must be involved during unconscious processing, since it also influences the subcortical regions). In addition, another difficulty for the Multiple Consciousness theory is that visual regions can lead to the binding of several attributes in the absence of consciousness, as shown both by Dehaene and colleagues using visual masking, and by Wojeiulik and Kanwisher in a patient with a bilateral parietal damage suggesting that the binding mechanisms that are supposed to lead to macroconsciousness can in fact operate in the absence of consciousness. Therefore, empirical evidence contradicts the claim that binding has to be a post-conscious process. Finally, and maybe more crucially, the inclusion of a threshold mechanism between unconscious and conscious processing, and hence the possibility of unconscious processing, is far from being obvious in this framework. This constitutes, as we saw above for the Duplex Vision theory, an empirical constraint (i.e., there has to be a threshold even if its nature remains unspecified) that implies a radical change in the main message conveyed by these two theories (i.e., consciousness is 'only partially' correlated with activity in the ventral/perceptual sites). In particular, it then becomes difficult to understand the frequent claim by Zeki that processing sites in the visual brain are also perceptual sites.

Neurobiological Standpoints on the Hard Problem

s0055

p0165 Now that we have seen the main characteristics of these competitive theories individually, the next sections will describe how they face (or deny) conceptual issues related to the hard problem of explaining conscious experience.

Relation to Access and Phenomenal Consciousness

s0060

p0170 The neuroscientific and philosophical issues of consciousness have never been so closely related. A consequence of this close interplay is that conceptual issues raised by philosophers are progressively influencing neurobiological theories. Arguably, the most influential issue in recent years has been the potential distinction between phenomenal and access consciousness proposed by Ned Block. In short (and in the context of visual perception), this dissociative approach assumes that the NCC for phenomenal consciousness reflects the qualitative subjective experience (i.e., the qualia) associated with the percept. According to Block, the phenomenological richness that one can experience when seeing a complex visual scene goes far beyond what the observer can report. Hence, conscious access is assumed to reflect another NCC that is more linked to reportability, stimulus discrimination, introspection, etc. This distinction has been appealing to many neuroscientists in recent years, and several research programs are currently investigating the possibility to dissociate different NCCs for these two forms of consciousness.

p0175 An important contrastive feature regarding the five theories reviewed above is whether they accept or reject this dissociative approach. It turns out that all of them, except the Global Neuronal Workspace theory, have recently acknowledged or even incorporated this distinction. In the Reentrant Dynamic Core theory, the distinction between phenomenal and access consciousness is analogous to the one between primary consciousness and higher-order consciousness, respectively. In a recent extension of the Duplex Vision theory, neural activity in the ventral stream has been directly associated with phenomenal consciousness by Goodale. Similarly, Zeki has recently

linked micro- and macroconsciousness with the phenomenal consciousness of specific attributes (colors, contrasts, etc.) and bound objects, respectively, while unified consciousness is analogous to access consciousness. In the Local Recurrence theory, phenomenal consciousness has been a very central concept at the origin of the construction of this framework. It is assumed to be a by-product of local recurrent loops, while access consciousness involves additional widespread loops between posterior and anterior regions of the brain. Importantly, all these theories assume that the NCC for access consciousness involves more or less the network of brain regions constituting the workspace in Dehaene's theory. Therefore, the Global Neuronal Workspace theory is often considered as a restrictive theory of access consciousness, and has been criticized for confounding the subjective experience of consciousness with the 'subsequent' executive processes that are used to access its content.

p0180 Dehaene and his colleagues not only reject this dissociation in terms of two NCCs, but they also put doubts on its psychological reality. In particular, they deny the possibility of phenomenal consciousness without access and consider this dissociation to be flawed for at least two important reasons. First, following Larry Weiskrantz, they assume that reportability is the only reliable index of consciousness and thus stimuli that cannot be reported/accessed are by definition not conscious in any way. Second, following Kevin O'Regan and Alva Noë, they consider that apparent cases of a rich phenomenal experience without access actually reflect the so-called illusion of seeing. For instance, in 'change blindness' situations, observers are usually overconfident about their capacity of seeing an entire visual scene, while actually, they fail to notice when important modifications occur at unattended locations. As such, the illusion of phenomenal richness reflects the fact that observers think that they can see more than they actually do. Because of this overconfidence, it remains unclear how one can empirically probe observers to describe the content of unattended or inaccessible perceptual events. As one can see here, this conceptual issue of phenomenal consciousness without access is closely linked to the empirical issue of consciousness without attention. This latter issue is addressed in the next section.

Relation to the Distinction between Attention and Consciousness

s0065

p0185 Although attention and consciousness have long been considered to be similar, if not identical phenomena, it is largely acknowledged today that they can be dissociated. Yet, a double dissociation remains to be demonstrated. Indeed, since the seminal study by Naccache, Blandin, and Dehaene showed that top-down attention modulates subliminal priming, there is now a general consensus regarding the fact that attention is independent of consciousness and that subjects can attend to objects even if they do not consciously perceive them. However, the reverse dissociation where consciousness is itself independent of attention remains highly debated. More precisely, the possibility that subjects can be conscious of objects without any attention is at the center of a very intense controversy.

p0190 Neurobiologists acknowledging Block's dissociative approach actually assume that top-down attention is driven by workspace regions and correspond to a central mechanism, which, along with language and working memory, define access consciousness. For instance, in the case of visual perception, the signal in the posterior visual regions (i.e., in the occipitotemporal cortex) is assumed to be attentionally amplified in a top-down fashion by the same widespread parietofrontal areas as the workspace regions in Dehaene' theory. Thus, similarly to the possibility of phenomenal consciousness without access, all the neurobiological theories seen above, except that of the Global Neuronal Workspace, acknowledge the possibility of consciousness without attention. Some *a priori* support for the dissociative approach can be found in studies showing that visible (i.e., supraliminal) though unattended stimuli do not involve workspace regions, but rather an increase of activity in posterior regions, as I have recently evidenced myself. Yet, as we will see below, it remains extremely difficult to demonstrate that there is any form of consciousness for visible but unattended stimuli.

p0195 Here also, Dehaene and colleagues disagree with this dissociation and consider, on the contrary, that attention embodies conscious processing. According to them, only attended stimuli can be reported and are thus consciously

processed: demonstrating that an observer is conscious of an unattended stimulus, without relying on any sort of report by the subject, seems extremely difficult or even impossible. In other words, it appears that in order to assess consciousness of the stimulus, one necessarily needs to direct the observer's attention on the stimulus, leading to the conclusion that consciousness without attention is an illogical possibility! It is of note that a few attempts, notably by Koch and colleagues, have been made to demonstrate consciousness without attention. In particular, they have relied on situations termed 'near-absence of attention' and where a stimulus is presented in the periphery while the subject is performing the task on a central target. Under these conditions, Koch and colleagues have found that subjects can still consciously perceive the peripheral stimulus, at least indistinctively. Yet, unfortunately, it remains impossible to demonstrate that there has not been any residual attentional amplification in this situation (additionally, the mere fact that this situation is called 'near-absence of' and not 'full-absence of attention' appears to be highly symptomatic and suggestive of the difficulty to demonstrate consciousness without attention).

p0200 Importantly, the majority of neurobiologists including Dehaene and colleagues, acknowledge that the processing of supraliminal but unattended stimuli has a special status, in-between the first step of subliminal processing, and the last step of conscious access. However, they totally disagree on a crucial question which is, 'whether there is any form of consciousness associated with this intermediate level of processing or rather whether it is just another form of nonconscious perception.' Proponents of the Global Neuronal Workspace theory have been very explicit on this issue and consider that this intermediate level involves just another, more elaborated form of nonconscious processing. They termed it the 'preconscious' stage. According to them, a stimulus reaching this preconscious state becomes potentially accessible (it could quickly gain access to conscious report if it was attended), but it is not consciously accessed at the moment, mainly because attention is maintained away (e.g., by processing a concurrent stimulus as in situations of attentional blink, inattentional blindness, or change blindness).

p0205 It is interesting to note that since the recent inclusion of the preconscious stage in the Global Neuronal Workspace theory in 2006, it is becoming harder to distinguish it from the Local Recurrence theory on a purely structural basis. Indeed, a tripartite taxonomy is provided in both accounts. The initial stage reflects local feedforward activity at the neural level and corresponds to unconscious (i.e., subliminal) processing, while the third stage involves a global form of recurrence in the brain and reflects conscious access. Yet, the intermediate stage is also structurally analogous in these two theories, as it involves a local form of recurrence in both cases. Rather, it is on the psychological dimension that these two theories diverge, Dehaene considering this intermediate level to reflect a nonconscious stage, while Lamme takes it as reflecting phenomenal consciousness before access. This shows us how much neurobiological theories are still dependant on psychological (i.e., subjective) indexes of consciousness.

p0210 In sum, apart from proponents of the Global Neuronal Workspace theory, there is an increasing tendency in most neurobiological accounts to consider that there is an intermediate level associated with phenomenal consciousness without access. Consequently, new indexes that do not rely on subjective reports and that mainly reflect neural mechanisms are being proposed. For instance, Lamme proposes that consciousness should be indexed by the observation of local recurrent loops. In Edelman and Tononi's theory, consciousness is indexed in terms of neural complexity reflected by differentiated integration. Before addressing these approaches in the next section, it is necessary to remind the reader that the dissociation between consciousness with and without attention, similarly to the distinction between access and phenomenal consciousness, remains highly difficult to tackle experimentally. This difficulty is mainly due to the fact that subjective reports, which necessarily involve access/attention components, remain so far the best index of whether someone is conscious. Unless a more reliable index of consciousness is found, these dissociations might possibly turn out to be immune to scientific investigation. Here also, maybe we should wait and see; or maybe new and less constrained epistemological directions should be explored.

s0070 ## Relation to Neurocognitive Panpsychism

p0215 As we have seen in the previous section, there is an increasing tendency to consider that subjective reports cannot be trusted when indexing consciousness. Among the theories we have seen above, proponents of two theories in particular, that of Local Recurrence and that of the Reentrant Dynamic Core (and in particular the resulting Information Integration theory put forward by Tononi), have proposed alternative indexes that are supposedly more reliable than subjective reports. Both theories consider consciousness as an emergent property of any system sharing specific core mechanisms of either recurrence or differentiated integration, respectively. Therefore the most reliable index of consciousness should, according to these accounts, reflect the quantification of these core mechanisms. In particular, these neural indexes are assumed to offer a more reliable estimation of phenomenal experience, well above the poor estimation provided by subjective reports.

p0220 This approach can be appealing since it could potentially lead to a quantification of consciousness in nonhuman species, in preverbal infants, and in artificial intelligence systems. Yet, it leads to a form of panpsychism (see Glossary) that one can term as 'neurocognitive panpsychism.' Both the Local Recurrence theory and the Reentrant Dynamic Core theory are leading to a similar, though more restrictive panpsychist views: any system in the world can be conscious as long as it shares information following the core mechanisms that are specific to these theories. They constitute neurocognitive versions of panpsychism, because these theories are primarily aimed at offering neurobiological accounts of consciousness, while they can also be extended to any cognitive system in the world that shares the same mechanical properties for information-processing. As such, both theories follow the principle that consciousness arises when information is transmitted among neurons interacting along the core mechanisms of these theories.

p0225 Yet, there are several conceptual difficulties with these theoretical approaches (the term conceptual difficulties and not impossibilities is preferred here because none of these issues is impossible to resolve *per se*; however, they might turn out to be extremely difficult to demonstrate as well). A first conceptual difficulty is that this doctrine leads to the conclusion that any system displaying these specific core mechanisms will experience consciousness, whatever implementation supports it and, more importantly, whatever its size. This fact relates to Chalmers' argument that even a system as small as a thermostat, because it recurrently shares information between a temperature sensor and a controller for switching the heater on or off, will experience consciousness to some degree. Indeed, the theory of Local Recurrence leads to the fact that any two neurons in interaction would be sufficient for consciousness. The Reentrant Dynamic Core theory would require a few more neurons, but not that many to generate consciousness (precisely nine neurons would be sufficient according to Herzog, Esfeld, and Gerstner). Although such possibilities can hardly be totally rejected, it remains as unclear how they could ever be demonstrated.

p0230 A second main issue lies in the paradox of trying to prove that neural indexes are more respectable because they supposedly probe phenomenal and not access consciousness. Indeed, although it is clear that neural indexes offer interesting possibilities when report is impossible (for instance in cases of locked-in syndromes or for prelinguistic babies), they still cannot be taken as reflecting more than conscious access. This principle follows from the fact that neural indexes of any sort have to be validated by confronting them at one point with some kind of report, hence with access and not phenomenal consciousness. For instance, demonstrating that recurrent processing is sufficient for consciousness initially requires consciousness to be indexed by probing whether the system is indeed conscious, and the best way to do so is still to rely on subjective reports in humans. Then, a neural index can only be validated when a correlation has been established with subjective reports. Assuming that one can demonstrate that recurrence, even among two neurons, correlates with some degree of consciousness, this can be achieved only through report and that degree of consciousness will only reflect a partial form of access, not phenomenal consciousness. Once again, the interesting aspects of this approach cannot be neglected, since it can allow to subsequently probe whether the core mechanisms represented by this index can be found in

systems or species that cannot report, suggesting that they 'might probably' be conscious. However, neural indexes cannot be taken as reflecting a superior or richer phenomenal form of consciousness since they can only be validated through reports. In other terms, these theories paradoxically emphasize that their core mechanisms do not reflect access consciousness, while actually only this form of consciousness can be used to demonstrate their validity.

p0235 A third conceptual issue is that even if it turns out that one of these neural indexes turns out to be perfectly correlated with consciousness, and thus becomes a perfectly reliable measure of consciousness, then one might still ask whether we have made any progress. This issue is less problematic and might turn out to be resolved in the future. However, it is worth mentioning, since it remains unclear whether we are searching in the right direction. Indeed, although neural indexes might turn out to be very useful in many circumstances, they do not escape the limitation of NCCs. They would still primarily inform us on whether someone (or something) is or is not conscious. Yet, they would not directly help us understand why they are conscious. If it turns out that recurrence or differentiated integration are sufficient for consciousness, we would still have to accept that consciousness is 'emerging' from the brain each time these core mechanisms come into play. This issue is more moderate because once we have found a perfect neural index of consciousness, it might get us closer to the explanatory gap. Yet, it remains unclear how it will allow us to cross that gap.

s0075 Conclusion

p0240 By the end of this article, one might wonder where we stand with regard to the original issue of how the brain leads to consciousness. Over the last two decades, dozens of theories of the NCC have been proposed (leading David Chalmers to ironically speak about an 'NCC zoo') and a kind of trial-and-error strategy has been applied to evaluate them, and in most cases, to reject them one by one. Here we have focused more specifically on five recent neurobiological frameworks that are the most popular accounts in this field of research. We have seen that, actually, consciousness can hardly be restricted to neural activity in the ventral stream, nor is it simply reflecting recurrent activity or minimal activity in processing sites. It remains unclear whether consciousness reflects or does not reflect differentiated integration within a dynamic core, as emphasized by Edelman and Tononi's speculative, but powerful theory, or whether it simply reflects activity in workspace regions, as outlined by Dehaene's well-specified, but apparently restrictive account of conscious experience. One main difficulty in disentangling this issue and contrasting these theories, is that it remains unknown whether neurobiological frameworks should be restricted to conscious access and subjective reports, or whether they should extend to other (i.e., inaccessible, unattended, phenomenal) forms of consciousness. Therefore, future research will unavoidably be bound to decide whether neurobiological accounts should take into account or reject the hard problem of understanding a subjective form of conscious experience that cannot be simply defined as conscious access. Thus, facing the hard problem will probably not be any more a main issue on its own. Instead, it will be the problem of deciding whether there is a hard problem at all, leading to a 'super-hard problem' of consciousness and obviously complicating the whole issue. Whether we should stay focused on variants of the NCC until the solution pops out and reduces the hard problem to an easy one, or whether we actually need new laws (of physics, life, or anything else) to cross the explanatory gap, only time will tell. Yet, although none of the neurobiological theories provided so far have been able to provide a 'conclusive' explanation regarding this matter, the search for a neurobiological explanation of consciousness still constitutes one of the most exciting challenges of contemporary science.

See also: (00030); The Neurochemistry of Consciousness (00056).

Suggested Readings

Baars BJ (1988) *A Cognitive Theory of Consciousness*. New York: Cambridge University Press.

Block N (2005) Two neural correlates of consciousness. *Trends in Cognitive Science* 9: 46–52.

Chalmers D (1996) *The Conscious Mind*. New York: Oxford University Press.
Crick F and Koch C (1990) Towards a neurobiological theory of consciousness. *Seminars in Neuroscience* 2: 263–275.
Dehaene S, Changeux JP, Naccache L, Sackur J, and Sergent C (2006) Conscious, preconscious, and subliminal processing: A testable taxonomy. *Trends in Cognitive Sciences* 10: 204–211.
Dehaene S and Naccache L (2001) Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework. *Cognition* 79: 1–37.
Edelman GM and Tononi G (2000) *A Universe of Consciousness: How Matter Becomes Imagination*. New York, NY: Basic Books.
Gallese V (2007) The ''conscious'' dorsal stream: Embodied simulation and its role in space and action conscious awareness. *Psyche* 13(1): 1–20.
Goodale M (2007) Duplex vision: Separate cortical pathways for conscious perception and the control of action. In: Velmans M and Schneider S (eds.) *The Blackwell Companion to Consciousness*, pp. 616–627. Oxford: Blackwell.
Koch C (2004) *The Quest for Consciousness: A Neurobiological Approach*. Denver, CO: Roberts.
Koch C and Tsuchiya N (2007) Attention and consciousness: Two distinct brain processes. *Trends in Cognitive Sciences* 11: 16–22.
Kouider S and Dehaene S (2007) Levels of processing during non-conscious perception: A critical review. *Philosophical Transactions of the Royal Society of London B* 362: 857–875.
Lamme VA (2006) Towards a true neural stance on consciousness. *Trends in Cognitive Sciences* 10: 494–501.
Tononi G (2004) An information integration theory of consciousness. *BMC Neuroscience* 5: 42.
Zeki S (2007) A theory of micro-consciousness. In: Velmans M and Schneider S (eds.) *The Blackwell Companion to Consciousness*, pp. 580–588. Oxford: Blackwell.

Biographical Sketch

Sid Kouider is a cognitive neuroscientist working at the Ecole Normale Supérieure (Paris, France) on the neurobiological and psychological foundations of consciousness. His work focuses on contrasting conscious and unconscious processes, both at the psychological and neural level, using various behavioral and brain imaging methods. Recently, he extended this line of research to study the neural correlates of consciousness in prelinguistic babies.