

Review



Cite this article: Naccache L. 2018 Why and how access consciousness can account for phenomenal consciousness. *Phil. Trans. R. Soc. B* **373**: 20170357.

<http://dx.doi.org/10.1098/rstb.2017.0357>

Accepted: 21 May 2018

One contribution of 17 to a theme issue 'Perceptual consciousness and cognitive access'.

Subject Areas:

neuroscience, cognition, theoretical biology

Keywords:

consciousness, access consciousness, phenomenal consciousness, global workspace

Author for correspondence:

Lionel Naccache

e-mail: lionel.naccache@aphp.fr

Why and how access consciousness can account for phenomenal consciousness

Lionel Naccache^{1,2,3,4}

¹Department of Neurology, and ²Department of Neurophysiology, AP-HP, Groupe hospitalier Pitié-Salpêtrière, 75013 Paris, France

³Sorbonne Université, INSERM, U 1127, 75013 Paris, France

⁴Institut du Cerveau et de la Moelle épinière, ICM, PICNIC Lab, 75013 Paris, France

LN, 0000-0002-2874-1009

According to a popular distinction proposed by the philosopher Ned Block in 1995, our conscious experience would overflow the very limited set of what we can consciously report to ourselves and to others. He proposed to coin this limited consciousness 'Access Consciousness' (A-Cs) and to define 'Phenomenal Consciousness' as a much richer subjective experience that is not accessed but that would still delineate the extent of consciousness. In this article, I review and develop five major problems raised by this theory, and show how a strict A-Cs theory can account for our conscious experience. I illustrate such an A-Cs account within the global workspace (GW) theoretical framework, and revisit some seminal empirical findings and neuropsychological syndromes. In this strict A-Cs perspective, subjective reports are not conceived as the mere passive broadcasting of information to the GW, but as resulting from a dynamic and active chain of internal processes that notably include interpretative and belief attribution stages. Finally, I list a set of testable predictions, of unsolved questions and of some counterintuitive hypotheses.

This article is part of the theme issue 'Perceptual consciousness and cognitive access'.

1. Introduction

In this article, I review a series of five problems raised by the 'Phenomenal Consciousness' (P-Cs) theory initially framed by Ned Block. While some of these issues have been addressed previously—by myself and others (as mentioned below)—I found it useful to gather them with additional original problems within the same synthetic paper. Note also, that since the seminal version proposed by Ned Block, more recent and nuanced versions of P-Cs theory have been elaborated (see, for instance, the Discussion section of [1]). While some ideas proposed in the present article may suggest some convergence between 'Access Consciousness' (A-Cs) theories and these nuanced versions of P-Cs theory (see, for instance, §2d of the present article), it is important to underline some major issues that still distinguish A-Cs and P-Cs accounts of perception.

2. Five problems raised by phenomenal consciousness theory

(a) How do we know we are phenomenally conscious?

The core definition of P-Cs postulates the existence of unreported conscious experiences. As proposed by Block in his seminal 1995 article: 'Phenomenal consciousness is experience; the phenomenally conscious aspect of a state is what it is like to be in that state. The mark of access-consciousness, by contrast, is availability for use in reasoning and rationally guiding speech and action' [2, p. 228]. Later in the same article, Block confirmed his view that A-Cs is required for reasoning, reporting and enabling rational control of action. Our rich subjective (phenomenal) experience would not have to be restricted

to the very limited set of representations that we are accessing, and that we self-report. This proposal fits very nicely with the immediate intuition most of us share about conscious experience: we do experience much more than what we are actually able to report to ourselves or to others.

However, once we go beyond this immediate and intuitive agreement, we may wonder: how do I know I experience something? The response to this question is so univocal that I will cite Block's own answer: 'When one has a phenomenally conscious experience, one is in some way aware of having it' [3, p. 484]. In other words, we know we are phenomenally conscious because we access this experience and self-report it. The transitive action of being aware of something clearly defines P-Cs as a form of access to this something. Otherwise we would not have any subjective ground (valid or not) to posit the existence of P-consciousness. Consequently, P-Cs appears necessarily as a type of self-report, and seems, therefore, included within the realm of A-Cs. Note that this remark also unifies within A-Cs both the functions of consciousness (e.g. executive control machinery of self-report) and the subjective experience of consciousness [4].

It is noteworthy that, within this framework, two apparently very distinct reports in terms of content such as 'I see X' (typical 'A-Cs' report) and 'I see much more than X' (typical 'P-Cs' report) do belong to this same class of explicit meta-reports (see below). Both are explicit self-reports of the current subjective experience. P-Cs contents can be accounted for as typical A-Cs mental contents.

This first problem seems insoluble for P-Cs theory, but also requires us to define more carefully the concept of 'subjective report' (see below §2).

(b) How can we distinguish a P-conscious unreported representation from a non-conscious representation?

Under the assumption that the content of A-Cs would only correspond to a narrow subset of the mental perimeter of conscious experience, one crucial question then arises: how can we distinguish a P-conscious representation (i.e. conscious but unreported) from a non-conscious representation? If subjective reports are supposed to be of no help to address this major question, the frontier between P-conscious and unconscious (or non-conscious) representations appears as very loose and fuzzy, and cannot be defined on firm grounds. Actually, the ending point of this slippery slope corresponds to what we could name 'pan-consciousism', by analogy with panpsychism and its pitfalls. Under the P-Cs theory, any aspect of our mental life could be labelled as being P-conscious, and the very notion of unconscious cognition and mentation could be simply discarded. Consider, for instance, a visual stimulus presented to a subject unable to report its presence. Such a situation can occur in patients or in healthy individuals in various pathologic or experimental conditions such as blindsight, neglect, subliminal perception, attentional blink, change blindness or attentional blindness for instance. How would P-Cs theory categorize this non-reported stimulus: as P-conscious or unconscious? If one would rely on the fact that the individual 'is in some way aware of having it' to label it as P-conscious, then we are back to the previous issue we raised: P-conscious status would exclusively rely

on subjective report, and would, therefore, be included within the content of A-Cs. Alternatively, it is easy to see that no solid criterion enables us to tease apart these two categories: P-conscious versus unconscious.

In sharp contrast, reportability theory provides a clear delineation of conscious representations (accessed) and can be used in the framework of the global workspace (GW) model to propose a precise and testable taxonomy of various types of unconscious processing defined on their corresponding psychological and neural mechanisms [5,6].

(c) From 'pan-consciousism' of contents to 'pan-consciousism' of states

In addition to the risk of pan-consciousism of content (defining the whole perceptual content as P-conscious), P-Cs theory also paves the way to a pan-consciousism of states.

Under the reportability definition of conscious experience, being conscious is defined as the ability to self-report, irrespective of the content of such reports [7]. An individual unable to self-report is, therefore, univocally defined as being in an unconscious state. Here are some situations that correspond to such unconscious states: most periods of the deep sleep stage, general anaesthesia, comatose state, vegetative state (also coined unresponsive wakefulness syndrome) and complex partial epileptic seizures.

In sharp contrast, P-Cs theory is facing a difficult challenge here. If the reportability criterion would only capture A-Cs, and if P-Cs would rely on other psychological properties and on a distinct neural machinery from A-Cs, then may subjects unable to self-report still be considered as P-conscious?

In other words, the dissolution of the concept of 'unconscious contents' (see above) would be followed by the dissolution of the concept of 'unconscious states'. In the end, we would not have any solid argument to discard P-Cs in all these situations lacking A-Cs. Once crediting all these subjects with P-Cs, there is no reason not to extend this P-Cs credit to all living creatures, including vegetables, or even to microtubules or elementary particles. Obviously, P-Cs theorists may add other criteria (somehow arbitrarily within the P-Cs framework) to restrict P-Cs credit, such as the presence of a functional complex and differentiated cortical or computational network, but my point here is to underline the risks of pan-consciousism that seem inherent to P-Cs theory.

Moreover, while GW reportability theory provides a clear justification of its core postulate—according to which complex, coherent and differentiated brain-scale patterns of cortical activity would be necessary to conscious states—the neural requirements for P-Cs do not seem as well delineated (see, for instance, microconsciousness and local-recurrent loops hypotheses proposed, respectively, by Zeki [8] and Lamme [9], among the various possible neural substrates of P-Cs).

(d) Risk of psychological impoverishment of P-Cs

The pan-consciousism risk inherent to P-Cs theory comprises several problems, including the dissolutions of unconscious contents and states that we addressed above, as well as semantic and translational problems (e.g. in medicine when taking care of patients suffering from disorders of consciousness [10]).

In addition to these important issues, pan-consciousism can also be described as a 'poisoned kiss': by crediting all

living creatures with P-Cs, one severely impoverishes the psychological dimension of consciousness. Once having extended P-Cs beyond the limits of subjective reports that constitute the quintessence of a psychological view of subjectivity (i.e. 'What is it like to experience something?'), one is left with a kind of a-psychological ascription of consciousness that is disconnected from subjective reports.

Ironically, the P-Cs concept—that was initially introduced in order to take into account fine qualitative properties of conscious subjective experience—may turn out as an approach that has lost interest for psychological properties of subjective experience. The core origin of this apparent paradox has to be found in the ambiguous relation of P-Cs to A-Cs (see §1a). On the one hand, if P-Cs is ultimately defined on the basis of subjective reports, then there is no reason to tease it apart from A-Cs. On the other hand, if P-Cs is not to be based on subjective reports, there is no reason to link it with psychology [11].

(e) Gullibility of P-Cs theory: taking seriously subjective reports does not mean taking them at face value

Studying subjective constructs from a third person perspective—a project called hetero-phenomenology by Dennett [12]—exposes us to a difficult problem that we may define as finding the adequate distance to subjective reports. Indeed, this hetero-phenomenology approach exposes us to two major errors. Let us consider such an extreme report as a complex hallucination: 'I see a white rabbit wearing a waistcoat, and muttering: Oh dear! Oh dear! I shall be too late!'.

On the one hand, one can deliberately choose to ignore the richness of this subjective construct given the obvious absence of external reality of its content. This posture can obviously be extended to valid subjective reports that share with the hallucination the notion that they primarily reflect a subjective meaning rather than an objective description of external reality (even when the two do match quite nicely!). This first error can be described as an infinite distance between the observer and subjective reports, and leads straightforwardly to a posture of radical behaviourism that is deliberately ignoring the psychology of subjective states.

On the other hand, one can deliberately choose to take this subjective report at face value, in the name of researcher's interest for subjective constructs. This second error that corresponds to a null distance between the observer and subjective reports explains in part the pitfalls of active introspectionism, and leads to multiple blind spots in the mechanism underpinning our conscious reports and conscious experience. P-Cs theory seems to be exposed to this second risk. Indeed, the contents of typical subjective reports used to found P-Cs expose us to this gullibility pitfall when they are taken at face value. In a previous paper [13, p. 520], I illustrated the importance of this argument by applying it to the experiment of Rayner and Bertera [14] who used the moving window paradigm:

In the 'moving window' paradigm, for instance, where a computerized display is changed in synchrony with eye movements, viewers claim that they see a normal page of text even when all parafoveal information is replaced by strings of X's.

This seminal experiment reveals univocally the illusion of visual completeness, which constitutes the core argument of P-Cs theory (i.e. the subjective report and belief of seeing everything that is out there). Indeed, this experiment demonstrates how taking a sincere subjective report at face value to

build a theory of phenomenality without reportability may be extremely risky, and can lead to an incorrect theorization of what really happens in the mind and brain of the individual. In this experiment, subjects clearly do not experience the whole visual scene—as P-Cs theory would credit the subjects, by stating they are really P-Cs of the whole visual scene—but they really do believe they see everything that is out there. The difference between these two claims illustrates the need to use the adequate distance to analyse subjective reports: neither ignorance of subjective reports (infinite distance) nor gullibility (null distance). We have to create a theory of consciousness that can explain why subjects do experience and believe what they experience and believe, but we do not have to build a theory trying to take subjective beliefs for granted. The possibility of adopting such a correct distance with subjective reports has been emphasized for decades in psychology [15].

We may cite here the great G.K. Chesterton, in the light of this last problem raised by P-Cs theory: 'Do not be so open-minded that your brains fall out'!

As an interim conclusion of this section, we obtained two major results that are problematic to P-Cs theory. First, the existence of P-Cs necessarily originates from subjective reports that are, by definition, accessed and self-reported by the subjects. In other words, irrespective of the validity of its contents, P-Cs theory is grounded on A-Cs theory. Second, the content of subjective reports typically used to propose the existence of P-Cs do not have to be taken for granted because many of them reflect subjective beliefs that can be largely invalid.

3. Why access consciousness may be all there is to consciousness

At this stage, we still need to show how A-Cs theory can explain phenomenal experience.

(a) What a subjective report IS NOT

To clarify some frequent misunderstandings, I propose to begin by a 'negative' definition of subjective reports.

It is important not to confound the mental operation of self-reporting with the behavioural act (verbal or non-verbal) used to communicate the content of this report to an external observer. For instance, conscious but fully paralysed patients (e.g. locked-in syndrome patients before establishing an ocular code, or affected with a multisensory disconnection [16], or patients affected with severe forms of amyotrophic lateral sclerosis or of Guillain-Barré syndrome) are still self-reporting when they are awake, even if they cannot communicate their subjective reports. As a consequence, such patients do have preserved A-Cs.

A report is not limited to the visual modality, but can obviously address any sensory content, and, more broadly, any conscious content: reporting an emotional feeling, a memory, a behaviour, an intention, a belief, a desire, a fantasy ... As such, reportability can address any possible content of consciousness. It is not improper to state that subjective self-reports are the core content of our subjective experience: what it is like, for us, to experience what we experience.

A report is not necessarily verbal, as illustrated by the collection of subjective reports in aphasic patients, in the

mute disconnected right hemisphere of a split-brain patient [17], in preverbal infants or in non-human primates. Consider, for instance, the remarkable demonstration by Cowey and Stoerig of the possibility to collect non-verbal subjective perceptual reports in monkeys [18]. After a lesion of primary visual cortex in one hemisphere, human patients typically report the absence of conscious reportable vision in the contralateral hemi-field, while they can show reliable unconscious processing of these unreported stimuli, a phenomenon coined blindsight. Until the Cowey & Stoerig study, it was known that monkeys affected with a similar lesion do show residual visual processing, but the following question remained open: are they lacking, like human patients, subjective reportability in the corresponding visual hemi-field? To answer this question, three monkeys who showed excellent detection in tasks where a visual stimulus was presented on every trial, albeit at different positions, were tested in a signal-detection task in which half the trials were blank trials, with no visual stimulus. They then classified all visual stimuli presented in the impaired hemi-field as blank trials, demonstrating that they lacked the subjective experience of perceiving consciously these stimuli. In other words, this seminal study illustrates that it is possible to design ‘commentary keys’ paradigms enabling the collection of non-verbal subjective reports [19].

A subjective report is not a passive broadcasting of an initially non-conscious or preconscious [5] representation to the conscious content, but it is rather an active internal process that solicits many high-level cognitive functions such as postulated by the GW models of consciousness [7,20]. In particular, interpretative, narrative, belief-construct and belief criticism processes are engaged in this ability to self-report [6]. This can be illustrated by the rich dynamics of daily life subjective reports during which we typically correct, change and update continuously the current content of our stream of consciousness. This fictionalization inherent to self-report is often revealed by the impairments of such interpretative and belief-related processes in various neurological or psychiatric conditions. For instance, a patient suffering from a Capgras delusion who is unable to reject the incorrect interpretation of his wife being a look-alike that replaced her, or a left-paralysed patient with asomatognosia who believes his left arm does not belong to him, illustrate how self-reports do integrate primarily such active interpretative and belief processes [21]. Interestingly, some properties of these interpretative and belief-related processes are currently explored in theoretical models such as the dual model of delusion of Coltheart and colleagues [22]. Note however that the distinction between a subjective report—experienced at a first-person level—and objective reports is not always obvious to make and easy to define. This question constitutes a field of research in itself.

Finally, given that subjective reports are not to be confounded with the behaviour used to communicate them to an external observer, it becomes obvious that they can be collected using non-behavioural methods. For instance, once you identify a neural signature (e.g. EEG, MEG, fMRI) that is present when conscious subjects can report a visual [23] or an auditory stimulus [24], and that is absent otherwise, it becomes possible to probe the presence of such a signature in the absence of behavioural communication. We used such an approach to probe conscious access to an auditory regularity with scalp EEG in behaviourally vegetative state

patients [25]. This allowed us to correct this behavioural diagnosis in 2 out of 30 patients. Crucially, these two patients recovered behavioural evidence of consciousness a few days after EEG recording. The same approach was recently used to probe conscious access to a gradually unmasked visual stimulus [26] in 5-month-old infants [27].

(b) Conscious access is an all-or-none unified active multidimensional process

We emphasized the notion that subjective reports are not limited to the passive broadcasting of unconscious representations into a GW, but that they also include internal active processes that encompass a very large set of mental operations, and that they are not necessarily neither verbal nor overt in behaviour. Describing subjective reports as active constructs that result from the dynamic contribution of many cognitive abilities interacting within a conscious GW enlarges the dimensionality of subjective reports. Thereby, the respective subjective reports of a healthy literate adult, of a 5-month-old baby, of a chimpanzee, and of patients affected by various neurological and psychiatric diseases are expected to differ according to this last property.

Interestingly, the multidimensionality of conscious experience has been recently developed by Bayne, Hohwy and Owen in order to replace the too narrow and too limited current description of states of consciousness framed in terms of ‘levels of consciousness’, by a multidimensional view [28]. While I agree with them on some aspects of their criticism, I would argue that a multidimensional conception of consciousness is not incompatible with the existence of one core process showing an ‘all-or-none’ property (reportability present or absent), combined with multiple other components (e.g. language, episodic memory, executive functions, etc.) the functionality of which would contribute to a taxonomy of various conscious states. Such a multidimensional view of patients can be derived from the GW theory of consciousness [29]. Therefore, adopting a multidimensional view on consciousness does not preclude the ability to differentiate between various conscious and non-conscious fine-grained states on the basis of the presence/absence of self-reportability in behavioural and brain imaging data.

(c) An A-Cs account of the visual completeness illusion

I will now update our previous explanation [13] of the strong visual completeness illusion, within the seminal context of a typical Sperling’s iconic memory experiment [30]. When an array of 12 letters is briefly presented (approx. half a second), subjects have the ability to consciously report only a subset of letters. However, they also claim that they have a strong phenomenal impression of having seen all the letters. This experiment captures the mismatch between limited reportability of precise items, and an apparent much richer visual experience of the whole visual scene. This mismatch is frequently interpreted in terms of the respective capacities of A-Cs (limited) and P-Cs (much larger capacity) [3].

However, we must first note that the visual completeness belief is nothing else but a subjective report of the individual: ‘I saw the whole array’. In other words, and as mentioned above (see §2(a)), the intuition of being P-Cs of many more objects than we can report (A-Cs) originates from subjective reports. Therefore, the apparent mismatch we have to explain

is not between A-Cs and P-Cs, but between two apparently contradictory subjective reports that belong to the content of A-Cs.

Here is a very plausible scenario of what happens:

When facing this array with attention, a subject's brain represents unconsciously the precise identity of each individual letter of the array in the ventral pathway, in the V4 area and/or in the visual word form area or closely related areas of the left occipital cortex and left fusiform gyrus [31,32]. In parallel, a representation of the visual background, providing raw array structure and items' location, is probably also coded in the dorsal visual pathway (e.g. LIP area of the parietal cortex). As soon as approximately 300 ms after array onset [23,33], the subject consciously accesses a visual representation combining the global background description (array location and its imprecise 'letterhood' elements) with the precise identity of the few letters that were foveated and attended, and that could enter into the GW and working memory (that is postulated as one of the key processors contributing to GW architecture). This consciously accessed representation would also include a kind of high-level filling-in process (comparable to the retinal blindspot filling-in process that has been studied extensively [34,35]). This interpretative process would build the illusory attribute that all individual stimuli are precise letters, given that they share a 'letterhood' aspect, and given that when each of them is individually accessed, it is indeed a precise letter (a cognitive variation around the classical 'refrigerator light illusion' [36,37]). This active interpretative process would require the coexistence of the ventral pathway (local, precise and partial) and dorsal pathway (global and imprecise) attributes of the visual representation of the array to occur. The net subjective report would be something like: 'I saw an array of letters, among which there was a K, an N, a G and an L'. In a way, this subjective report is quite correct, but only its extrapolation to the belief that all constituents were experienced as precisely as the few letters that were individually accessed would be incorrect and illusory. Future studies could aim at better exploring the detailed subjective experience reported by subjects in such situations: are they really experiencing a full illusion of visual completeness, or are they rather reporting a more shaded experience?

The individual letters are clear precise contents of A-Cs, but so is the report of perceiving a whole array of letters (access to the global and imprecise representation), as well as the belief that the visual experience overflowed the few identified letters (active interpretative process contributing to A-Cs). All these three components of the report can be accommodated within the strict A-Cs framework, and change blindness, Rayner & Bertera's illusion or other manipulations can be explained accordingly.

In contrast with this scenario, Block and other P-Cs theorists would suggest here that the unreported letters could still have been experienced precisely by the subject unable to report each of them. While we showed how problematic this proposal is, it also relies on the 'overflow' argument: the reason why subjects cannot report all their visual experience is because this rich experience is too large to be captured by the limited capacity of A-Cs. Crucially, an elegant series of retrospective cueing by Sergent *et al.* [38] showed that conscious perception of one single stimulus (and not of a whole 12-letter array), presented close to conscious threshold,

can still be triggered several hundreds of milliseconds after stimulus disappearance. This result demonstrates that post-cued attention can trigger the conscious perception of a single non-reported letter that would have otherwise escaped consciousness. In this case, the overflow argument cannot be used easily to explain why the transition from P-Cs to A-Cs did not work in spite of the absence of overflow (a single stimulus is presented). In contrast, this counterintuitive result is compatible with our radical proposal that P-Cs content is nothing else but a sub-set of A-Cs content. Note also that this retrospective cueing effect can provide a satisfactory A-Cs account of findings initially interpreted as a large P-Cs capacity [39].

An additional support to the scenario we proposed can be found in the Balint's syndrome associated with bilateral lesions of the dorsal pathway [40,41]. In such a condition, patients typically report serially and slowly the few items of the visual scene they could attend to, but they do not report the experience of seeing everything. This is noteworthy given that patients do not have visual scotomas or visual field defects. This loss of the visual completeness belief would fit well with the absence of the global imprecise representation of items' location and raw identity—due to the parietal cortex lesions—that is predicted to be necessary to the building of the experience of seeing everything. In other words, the visual completeness belief is not mandatory and irrepressible, but does require dorsal pathway areas to compute the global representation of the visual scene mentioned above. When such a representation is unavailable and, therefore, is not accessed consciously, the visual completeness report simply disappears.

(d) All subjective reports are not meta-reports

Under our exclusively A-Cs account of conscious experience, there is still one key problem to address. Try to keep your eyes open, without engaging into the explicit activity of self-reporting what you see and experience. Once you suddenly access what you experienced immediately before self-reporting, you are irrepressibly left with the impression that you actually experienced something of the scene before self-reporting.

What is the status of this pre-report experience?

This could either be an illusion of memory continuity, bridging temporally the current report (see also below §4(c)) with the subjective past. Under this hypothesis, no additional ingredient would be needed.

Alternatively, we could hypothesize that the content of A-Cs is not limited to the products of the explicit task of self-reporting (explicit statements such as: 'I report X'), but could also include accessed representations that are not necessarily manipulated by the machinery of self-reporting, but are nevertheless accessed and reportable. This second hypothesis would lead one to distinguish two forms of subjective reports: (i) those that concern accessed representations that are not explicitly reflexively reported (primary or simple reports), and (ii) meta-reports defined as reports used by the explicit self-report machinery to provide statements such as: 'I saw X'. A primary report would correspond to the representation that is broadcasted, accessed and actively interpreted by the GW constituents, whereas a meta-report would include only a subset of these reports: those that

also include processing by the explicit self-report machinery. Both forms of reports would share a common updating of the conscious access content within the GW, but only meta-reports would engage the explicit reflexive reporting processor that contributes to GW architecture. In other words, a meta-report would differ from a primary report by the type of specialized processing within the GW functional architecture, as is also the case for many other kinds of conscious processes that rely on various configurations of GW processors.

One may read this proposal as a semantic trick designed to incorporate P-Cs within A-Cs while preserving the attributes of P-Cs. However, as mentioned several times in this article, this distinction proposed between a primary report and a meta-report does not reflect a discussion between A-Cs and P-Cs, but a discussion regarding the content of A-Cs. In particular our proposal leads to the prediction that a neural signature of conscious access should be present both in primary reports and in meta-reports. Thus, the predicted neural signature of primary reports would not be early and local (e.g. early local recurrent loops predicted by several P-Cs theorists), but would rather be a late and global ignition of the GW.

This distinction between meta-reports and primary reports could be tested in healthy and educated adult human subjects, but also in patients, in infants and in non-human species. In the framework of the GW, this distinction emphasizes the importance of combining an all-or-none common process mandatory for conscious processing (access to GW) with the specific functional architecture of GW processors open to social and cultural factors [7].

4. Future perspectives

As a conclusion, I would like to state some predictions, key questions and hypotheses that could be addressed in future studies.

(a) A set of testable predictions

The main prediction stemming from my proposal is that the neural signature of conscious access should be present in all situations corresponding to the updating of the content of the GW, irrespective of their A-Cs or P-Cs label. Crucially, experimental conditions considered as typical illustrations of P-Cs experience are predicted to show the same neural signature of access, as in any other consensual A-Cs situations. To test this prediction, we first have to identify such a signature, and ideally to be able to identify it on single trials.

We previously proposed the P3b event-related potential component as a good candidate for this signature of conscious access [23–27], and described the spectral power and functional connectivity facets of this neural event [33,42]. We noted that the same global P3b event was present irrespective of the sensory modality (visual or auditory), and irrespective of the content that was accessed [43]. Our proposal is debated on several grounds ranging from its specificity to consciousness [44,45], to its potentially too late timing relative to conscious access [46–49]. However, in support of our hypothesis, Rockstroh and colleagues proposed that P300 that appears as a slow positive shift is a net inhibitory signal that can be explained in a Hebbian neural network assembly:

Activity should reverberate only in the cell assemblies actively involved in the specific information. The development and stabilization of these distinct synaptic connections require a large portion of the cells for the incoming concept to be shut off. This should be seen in a reduced depolarization or even inhibition of vast networks. The surface positivity corresponding to these inhibited networks would then dominate over the relatively smaller spots of negativity caused by the reverberating excitation. We may hypothesize that positive waves such as the P300 result from such a disfacilitation of widespread neural activity [50, pp. 175–176].

Applied to the GW framework, the hypothesis of Rockstroh *et al.* would, therefore, explain why a same unique signature would occur for any conscious access event. Indeed, if this neural signature is the net result of a massive inhibition to most components of the GW, and of a very small activation of the representation that is accessed, then one expects this global signature to be highly similar across the very numerous and distinct contents that can be accessed. However, multivariate decoding techniques as well as multi-unit recordings in human and non-human primates could be used to finely disentangle the largely common massive inhibition from the tiny and specific activation patterns that are predicted to differ according to the specific conscious content.

Importantly, a reliable neural signature of conscious access could be used to address recent original findings described by their authors as a plausible support to new versions of P-Cs theory. Indeed, Bronfman *et al.* [1] engaged subjects in a dual task with a Sperling-like array of coloured letters. Immediately after reporting letters from the cued row, participants had to estimate colour diversity of the non-cued rows of the array better. In a series of experiments, Bronfman *et al.* showed that subjects could perform better than chance level without a cost to letter report in this second task. The authors proposed that this finding could be explained in a P-Cs perspective:

One possible interpretation of the results is in agreement with the rich-phenomenal-experience hypothesis, which asserts that during exposure to an array of letters observers initially experience more visual information than is subsequently available for subjective report. This information is encoded in fragile visual short-term memory and decays before it can be encoded into durable working memory for later report [1, p. 1402].

Alternatively, this forced choice performance in the second task (that was not followed by an assessment of the subjective confidence in this response) could also be explained in terms of unconscious processing, as also mentioned by the authors. Probing the neural signature of conscious access to non-cued colour diversity in this paradigm could, therefore, disentangle between these two possible interpretations.

(b) Reported versus reportable representations?

Are we always conscious of a content, or should we define consciousness as the ability to consciously access and report a representation, without necessarily being conscious of any content? Framed in our contemporary language, this question reiterates the intentionality attribute of consciousness initially described by phenomenology since Brentano and Husserl.

This question appears in our proposal as two questions related to the distinction between reported and reportable representations. First, concerning the unreported representations that are potentially reportable, is there a way to propose an explanatory taxonomy of them: why did we not access them?

We previously proposed such a taxonomy of unconscious representations ranging from: (i) information that is not explicitly coded in a cell assembly, to (ii) information explicitly coded in a network unreachable by the GW such as some brainstem structures, to (iii) subliminal ventral pathway and transient visuo-motor dorsal pathway activations that are not lasting long enough to be accessed and, finally, to (iv) supraliminal representations that could have been accessed if they had been amplified. For this last category, we proposed to label them as pre-conscious representations so as to emphasize their proximity with GW potential content [5,6].

The second question regarding report versus reportability deals with intentionality: are we always conscious of a given content, or are there conscious states free of any content? The GW framework seems rather close to the central claim of phenomenology, as it states that ‘global availability of information through the workspace is what we subjectively experience as a conscious state’ [7]. In other words, being conscious requires a functional GW, the content of which corresponds to conscious self-reported states. In this view, there is no place for a content-free conscious state. Obviously, conscious contents are not restricted to perceptual contents, and can correspond to any kind of report. Note also that the proposed distinction between primary reports and meta-reports casts some light on the distinct types of content (see above).

(c) Discrete temporal islets of consciousness separated by brief periods of unconsciousness?

When addressing the report versus meta-report distinction (see above), we raised the possibility that between two successive self-reports, a subject may actually not be in a conscious state. This strange possibility, according to which we would be conscious only during temporal islets interspersed with unconscious states, may deserve more attention. Indeed, a recent empirical study investigated the dynamics of consciousness by decomposing time series of fMRI resting state recordings of awake monkeys into several discrete states defined by distinct patterns of functional connectivity [51]. Such a decomposition enabled the authors to isolate some brain states defined by long-distance positive and negative correlations. Interestingly, these patterns were the least correlated with structural anatomical connections, whereas other states, defined by exclusively positive correlations, closely resembled structural anatomy. Previous studies already insisted on the importance of positive and negative correlation patterns during conscious states in humans [52,53]. Under anaesthesia, when monkeys lost vigilance and awareness, the positive/negative patterns vanished while the exclusively positive correlation patterns dominated the whole time series. In contrast, during conscious wakefulness the patterns that included long-distance negative correlations dominated the time-series decomposition. We recently extended this analysis to human patients suffering from disorders of consciousness, and showed that only patients in the minimally conscious state showed the anti-correlation pattern that was absent in unconscious vegetative state patients [54]. Note, however, that even during conscious states, some patterns showing only positive correlations (and observed mostly during unconscious states) still contributed to the dynamics of resting state time series. In the same vein, Tagliazucchi and Laufs analysed 1147

resting-state functional magnetic resonance datasets of human volunteers, and discovered dynamic transitions between conscious and sleep patterns, with fundamental changes in the associated functional neuroanatomy [55]. Taken together, these results could suggest that during conscious wakefulness, a form of high-level filling-in process may join discrete conscious states separated by short periods of unconsciousness into what we subjectively experience as a continuous stream of consciousness.

(d) Conscious and unconscious editing of conscious content

Finally, the conception of conscious access exposed above, as well as the proposed definition of primary reports as active processes (as opposed to passive broadcasting), raise the challenging question of the conscious versus unconscious type of this editing process that build our conscious content. This question would deserve to be addressed specifically, but many arguments converge to the notion that both conscious and unconscious interpretative and belief attribution processes participate in it. Typically, when we consciously access representations, we access interpreted mental objects while we do not self-report being the agents of these interpretations (e.g. seeing a face, perceiving a word or accessing a memory). However, once we consciously access these representations, we also have the ability to update them voluntarily, to change their meaning, to modify our confidence and belief regarding their validity. This dynamic ballet between conscious and unconscious interpretative processes can be easily discovered in everyday experiences, such as in the classical ‘uncanny’ (worrying strangeness) episode of Freud misattributing his own mirror reflection to the identity of an unknown stranger, before correcting voluntarily this incorrect initial interpretation (probably processed unconsciously) he accessed without effort and agentivity [56].

In the same vein, the writer Nancy Huston describes such a typical accident [57, pp. 65–66]:

Sometimes we can ‘sneak up’ on our brains, as it were, and watch them in the act of spinning tales for us to believe in. The other day, for instance, I entered my building, saw that the elevator was stopped on one of the upper floors, heard someone enter it and start heading down. When the doors opened on the ground floor, I expected to see one of my neighbours emerge, which is what always happens in this situation. Not today, however. (Now I must make it clear that all the mental processes described in the following paragraph took place within a few milliseconds.) At eye level, I saw nothing; disconcerted, I thought, oh, it’s not an adult, it must be a small child; glancing downward, I saw that I was wrong – it was a woman, but her head was at my waist level. She’s defecating, I thought. . . . No, she’s shinnying up from an underground tunnel through a hole in the elevator floor. . . . No, she had to crouch down to rummage through her handbag for a key.

I would predict that in such an example, each new interpretation actually corresponds to a new access and should be associated with a similar neural signature (e.g. P3b, see above).

I recently explored how the analysis of such collisions between external symbolic signs and our subjective interpretation system can help to reveal the dynamics of these unconscious and conscious processes that participate to define our subjective identity, as well as their reciprocal influences [58]. In particular, several studies revealed that, in many situations, the current conscious posture of the subject (allocation of top-down spatial [59] and temporal attention

[60], expectations [61] and stimuli sets and task-related strategic processing [62,63]) largely shapes and influences unconscious processing. In other words, the conscious posture determines indirectly the nature of the primarily unconscious contents that will be then accessed [64,65].

As a conclusion, while defending a radical position postulating that conscious access is basically all there is to consciousness, and while proposing that an A-Cs account of P-Cs is possible and very plausible, I do hope this scientific and fruitful discussion initiated by Ned will remain as stimulating, during the next decades, as it proved to be until now, and that we will go on sharing our views about what is (and what is not) consciousness.

References

- Bronfman ZZ, Brezis N, Jacobson H, Usher M. 2014 We see more than we can report: 'cost free' color phenomenality outside focal attention. *Psychol. Sci.* **25**, 1394–1403. (doi:10.1177/0956797614532656)
- Block N. 1995 On a confusion about the role of consciousness. *Behav. Brain Sci.* **18**, 227–287. (doi:10.1017/S0140525X00038188)
- Block N. 2007 Consciousness, accessibility, and the mesh between psychology and neuroscience. *Behav. Brain Sci.* **30**, 481–548. (doi:10.1017/S0140525X07002786)
- Cohen MA, Dennett DC. 2011 Consciousness cannot be separated from function. *Trends Cogn. Sci.* **15**, 358–364. (doi:10.1016/j.tics.2011.06.008)
- Dehaene S, Changeux J-P, Naccache L, Sackur J, Sergent C. 2006 Conscious, preconscious, and subliminal processing: a testable taxonomy. *Trends Cogn. Sci.* **10**, 204–211. (doi:10.1016/j.tics.2006.03.007)
- Naccache L. 2006 *Le nouvel inconscient. Freud, Christophe Colomb des neurosciences*. Paris, France: Odile Jacob.
- Dehaene S, Naccache L. 2001 Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework. *Cognition* **79**, 1–37. (doi:10.1016/S0010-0277(00)00123-2)
- Zeki S. 2003 The disunity of consciousness. *Trends Cogn. Sci.* **7**, 214–218. (doi:10.1016/S1364-6613(03)00081-0)
- Lamme VA. 2004 Separate neural definitions of visual consciousness and visual attention; a case for phenomenal awareness. *Neural Netw.* **17**, 861–872. (doi:10.1016/j.neunet.2004.02.005)
- Naccache L. 2017 Minimally conscious state or cortically mediated state? *Brain* **141**, 949–960. (doi:10.1093/brain/awx324)
- Lamme VA. 2006 Towards a true neural stance on consciousness. *Trends Cogn. Sci.* **10**, 494–501. (doi:10.1016/j.tics.2006.09.001)
- Dennett DC. 1992 *Consciousness explained*. London, UK: Penguin.
- Naccache L, Dehaene S. 2007 Reportability and illusions of phenomenality in the light of the global neuronal workspace model. *Behav. Brain Sci.* **30**, 518–520. (doi:10.1017/S0140525X07002993)
- Rayner K, Bertera JH. 1979 Reading without a fovea. *Science* **206**, 468–469. (doi:10.1126/science.504987)
- Nisbett R, Wilson T. 1977 Telling more than we can know: verbal reports on mental processes. *Psychol. Rev.* **84**, 231–259. (doi:10.1037/0033-295X.84.3.231)
- Rohaut B, Raimondo F, Galanaud D, Valente M, Sitt JD, Naccache L. 2017 Probing consciousness in a sensory-disconnected paralyzed patient. *Brain Injury* **31**, 1398–1403. (doi:10.1080/02699052.2017.1327673)
- Gazzaniga MS, LeDoux JE, Wilson DH. 1977 Language, praxis, and the right hemisphere: clues to some mechanisms of consciousness. *Neurology* **27**, 1144–1147. (doi:10.1212/WNL.27.12.1144)
- Cowey A, Stoerig P. 1995 Blindsight in monkeys. *Nature* **373**, 247–249. (doi:10.1038/373247a0)
- Weiskrantz L. 1997 *Consciousness lost and found: a neuropsychological exploration*. New York, NY: Oxford University Press.
- Baars BJ. 1988 *A cognitive theory of consciousness*. New York, NY: Cambridge University Press.
- Naccache L. 2009 Visual consciousness: an updated neurological tour. In *The neurology of consciousness* (eds S Laureys, G Tononi), pp. 271–281. London, UK: Academic Press.
- Coltheart M, Langdon R, McKay R. 2011 Delusional belief. *Ann. Rev. Psychol.* **62**, 271–298. (doi:10.1146/annurev.psych.121208.131622)
- Sergent C, Baillet S, Dehaene S. 2005 Timing of the brain events underlying access to consciousness during the attentional blink. *Nat. Neurosci.* **8**, 1391–1400. (doi:10.1038/nn1549)
- Bekinschtein TA, Dehaene S, Rohaut B, Tadel F, Cohen L, Naccache L. 2009 Neural signature of the conscious processing of auditory regularities. *Proc. Natl Acad. Sci. USA* **106**, 1672–1677. (doi:10.1073/pnas.0809667106)
- Faugeras F et al. 2011 Probing consciousness with event-related potentials in the vegetative state. *Neurology* **77**, 264–268. (doi:10.1212/WNL.0b013e3182217ee8)
- Del Cul A, Baillet S, Dehaene S. 2007 Brain dynamics underlying the nonlinear threshold for access to consciousness. *PLoS Biol.* **5**, e260. (doi:10.1371/journal.pbio.0050260)
- Kouider S, Stahlhut C, Gelskov SV, Barbosa LS, Dutat M, De Gardelle V, Christophe A, Dehaene S, Dehaene-Lambertz G. 2013 A neural marker of perceptual consciousness in infants. *Science* **340**, 376–380. (doi:10.1126/science.1232509)
- Bayne T, Hohwy J, Owen AM. 2016 Are there levels of consciousness? *Trends Cogn. Sci.* **20**, 405–413. (doi:10.1016/j.tics.2016.03.009)
- Naccache L. 2018 Reply: minimally conscious state or cortically mediated state? *Brain*. **141**, e27 (doi:10.1093/brain/awy026)
- Sperling G. 1960 The information available in brief visual presentation. *Psychol. Monogr.* **74**, 1–29. (doi:10.1037/h0093759)
- Cohen L, Dehaene S, Naccache L, Lehéricy S, Dehaene-Lambertz G, Hénaff MA, Michel F. 2000 The visual word form area: spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. *Brain* **123**, 291–307. (doi:10.1093/brain/123.2.291)
- Vinckier F, Dehaene S, Jobert A, Dubus JP, Sigman M, Cohen L. 2007 Hierarchical coding of letter strings in the ventral stream: dissecting the inner organization of the visual word-form system. *Neuron* **55**, 143–156. (doi:10.1016/j.neuron.2007.05.031)
- Gaillard R, Dehaene S, Adam C, Clémenceau S, Hasboun D, Baulac M, Cohen L, Naccache L. 2009 Converging intracranial markers of conscious access. *PLoS Biol.* **7**, e61. (doi:10.1371/journal.pbio.1000061)
- Meng M, Remus DA, Tong F. 2005 Filling-in of visual phantoms in the human brain. *Nat. Neurosci.* **8**, 1248–1254. (doi:10.1038/nn1518)
- Komatsu H. 2006 The neural mechanisms of perceptual filling-in. *Nat. Rev. Neurosci.* **7**, 220–231. (doi:10.1038/nrn1869)
- Thomas NJ. 1999 Are theories of imagery theories of imagination? An active perception approach to conscious mental content. *Cogn. Sci.* **23**, 207–245. (doi:10.1207/s15516709cog2302_3)

37. O'Regan JK, Noe A. 2001 A sensorimotor account of vision and visual consciousness. *Behav. Brain Sci.* **24**, 939–1031. (doi:10.1017/S0140525X01000115)
38. Sergent C, Wyart V, Babo-Rebello M, Cohen L, Naccache L, Tallon-Baudry C. 2013 Cueing attention after the stimulus is gone can retrospectively trigger conscious perception. *Curr. Biol.* **23**, 150–155. (doi:10.1016/j.cub.2012.11.047)
39. Landman R, Spekreijse H, Lamme VA. 2003 Large capacity storage of integrated objects before change blindness. *Vision Res.* **43**, 149–164. (doi:10.1016/S0042-6989(02)00402-9)
40. Hécaen A, Ajuriaguerra J. 1954 Balint's syndrome (psychic paralysis of visual fixation) and its minor forms. *Brain* **77**, 373–400. (doi:10.1093/brain/77.3.373)
41. Coslett HB, Saffran E. 1991 Simultanagnosia: to see but not two see. *Brain* **114**, 1523–1545. (doi:10.1093/brain/114.4.1523)
42. El Karoui I *et al.* 2015 Event-related potential, time-frequency, and functional connectivity facets of local and global auditory novelty processing: an intracranial study in humans. *Cereb. Cortex* **25**, 4203–4212. (doi:10.1093/cercor/bhu143)
43. Rohaut B, Faugeras F, Chausson N, King JR, El Karoui I, Cohen L, Naccache L. 2014 Probing ERP correlates of verbal semantic processing in patients with impaired consciousness. *Neuropsychologia* **66**, 279–292. (doi:10.1016/j.neuropsychologia.2014.10.014)
44. Silverstein BH, Snodgrass M, Shevrin H, Kushwaha R. 2015 P3b, consciousness, and complex unconscious processing. *Cortex* **73**, 216–227. (doi:10.1016/j.cortex.2015.09.004)
45. Naccache L, Marti S, Sitt JD, Trübtschek D, Berkovitch L. 2016 Why the P3b is still a plausible correlate of conscious access? A commentary on Silverstein *et al.*, 2015. *Cortex* **85**, 126–128. (doi:10.1016/j.cortex.2016.04.003)
46. Sergent C, Naccache L. 2012 Imaging neural signatures of consciousness: 'what', 'when', 'where' and 'how' does it work? *Arch. Ital. Biol.* **150**, 91–106.
47. Aru J, Bachmann T, Singer W, Melloni L. 2012 Distilling the neural correlates of consciousness. *Neurosci. Biobehav. Rev.* **36**, 737–746. (doi:10.1016/j.neubiorev.2011.12.003)
48. Koivisto M, Grassini S. 2016 Neural processing around 200 ms after stimulus-onset correlates with subjective visual awareness. *Neuropsychologia* **84**, 235–243. (doi:10.1016/j.neuropsychologia.2016.02.024)
49. Koivisto M, Salminen-Vaparanta N, Grassini S, Revonsuo A. 2016 Subjective visual awareness emerges prior to P3. *Eur. J. Neurosci.* **43**, 1601–1611. (doi:10.1111/ejn.13264)
50. Rockstroh B, Müller M, Cohen R, Elbert T. 1994 Probing the functional brain state during P300-evocation. *J. Psychophysiol.* **6**, 175–184.
51. Bartfeld P, Uhrig L, Sitt JD, Sigman M, Jarraya B, Dehaene S. 2015 Signature of consciousness in the dynamics of resting-state brain activity. *Proc. Natl Acad. Sci. USA* **112**, 887–892. (doi:10.1073/pnas.1418031112)
52. Fox MD, Snyder AZ, Vincent JL, Corbetta M, Van Essen DC, Raichle ME. 2005 The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proc. Natl Acad. Sci. USA* **102**, 9673–9678. (doi:10.1073/pnas.0504136102)
53. Vanhaudenhuyse A *et al.* 2011 Two distinct neuronal networks mediate the awareness of environment and of self. *J. Cogn. Neurosci.* **23**, 570–578. (doi:10.1162/jocn.2010.21488)
54. Demertzi A *et al.* 2018 Dynamic inter-regional coordination patterns as specific predictors of consciousness. *4th Congress of the European Academy of Neurology, Lisbon, 17 June 2018*.
55. Tagliazucchi E, Laufs H. 2014 Decoding wakefulness levels from typical fMRI resting-state data reveals reliable drifts between wakefulness and sleep. *Neuron* **82**, 695–708. (doi:10.1016/j.neuron.2014.03.020)
56. Freud S. 2003 *The uncanny [1919]*. London, UK: Penguin Classics.
57. Huston N. 2008 *The tale-tellers: a short study of humankind*. Toronto, Canada: McArthur & Co Pub Ltd.
58. Naccache L. 2017 *Le chant du signe. Psychopathologie de nos interprétations quotidiennes*. Paris, France: Odile Jacob.
59. Kentridge RW, Heywood CA, Weiskrantz L. 1999 Attention without awareness in blindsight. *Proc. R. Soc. Lond. B* **266**, 1805–1811. (doi:10.1098/rspb.1999.0850)
60. Naccache L, Blandin E, Dehaene S. 2002 Unconscious masked priming depends on temporal attention. *Psychol. Sci.* **13**, 416–424. (doi:10.1111/1467-9280.00474)
61. Kunde W, Kiesel A, Hoffmann J. 2003 Conscious control over the content of unconscious cognition. *Cognition* **88**, 223–242. (doi:10.1016/S0010-0277(03)00023-4)
62. Kouider S, Dupoux E. 2004 Partial awareness creates the 'illusion' of subliminal semantic priming. *Psychol. Sci.* **15**, 75–81. (doi:10.1111/j.0963-7214.2004.01502001.x)
63. El Karoui I, Christoforidis K, Naccache L. 2017 Can application and transfer of strategy be observed in low visibility condition? *PLoS ONE* **12**, e0173679. (doi:10.1371/journal.pone.0173679)
64. Naccache L. 2008 Conscious influences on subliminal cognition exist and are asymmetrical: validation of a double prediction. *Conscious Cogn.* **17**, 1353–1359. (doi:10.1016/j.concog.2008.01.002)
65. Rohaut B, Alario F, Meadow J, Cohen L, Naccache L. 2016 Unconscious semantic processing of polysemous words is not automatic. *Neurosci. Conscious.* **2016**, niw010. (doi:10.1093/nc/niw010)