

Penultimate draft: January 2018

Forthcoming in *The Oxford Handbook of the Philosophy of Consciousness*, Uriah Kriegel (ed.), Oxford University Press.

The Neural Correlates of Consciousness

Jorge Morales¹ & Hakwan Lau^{2,3}

¹ Department of Philosophy, Columbia University

² Department of Psychology, UCLA

³ Brain Research Institute, UCLA

⁴ Department of Psychology, University of Hong Kong

Our understanding of the neural basis of consciousness has substantially improved in the last few decades. New imaging and statistical techniques have been introduced, experiments have become more sophisticated, and several unsuccessful hypotheses have been quite conclusively ruled out. However, neuroscientists still do not entirely agree on the critical neural features required for sustaining perceptual conscious experiences in humans and other primates. In this chapter, we discuss a selection of influential views of the neural correlates of consciousness (NCC) and the predictions they make. For example, neural activity synchronized at 40Hz used to be considered a serious candidate for the NCC. Among current views, some expect activity in the ventral stream of the visual processing pathway to be crucial for consciousness, others expect recurrent activity in visual areas, distributed activity across frontoparietal areas, or specific activity in prefrontal cortex (PFC). In particular, we focus on the predictions these views make with respect to the role of PFC during visual experiences, which is an area of critical interest and some source of contention. Our discussion of these views will focus mainly on the level of functional anatomy, i.e. the level at which we consider different brain regions,

rather than at the neuronal circuitry level. We take this approach because we currently understand relatively more about experimental evidence at this coarse level, and because these results are appropriate for arbitrating between current theoretical frameworks. For instance, while the Neural Synchrony Theory (Crick & Koch 1990), the Two-Visual-Systems Hypothesis (Milner & Goodale 1995; 2006), and the Local Recurrency Theory (Lamme 2010; Lamme 2006) predict that PFC activity is not critical for perceptual consciousness, the Higher Order (Lau 2008; Lau & Rosenthal 2011) and Global Workspace (Dehaene & Naccache 2001; Dehaene 2014; Baars 1997; Baars 2005) Theories confer activity in PFC a crucial role in enabling conscious perception. Moreover, while Global Workspace Theory requires global and elevated activity distributed in a frontoparietal network, Higher Order Theory expects specific computations in PFC to be responsible for visual conscious experiences.

While it is sometimes described as a “brain mapping” issue (for example, in the form of questions like “*Where* is the neural basis of consciousness?”), finding the NCC is hardly a simple “localization” job. This is not to say that identifying certain areas differentially involved during conscious experiences is not part of what is required for finding the NCC. But the theoretically interesting quest for the NCC goes beyond straightforward “brain mapping.” Success in finding the NCC is likely to involve describing how multiple brain areas work in conjunction to sustain conscious experiences, as well as the neural computations and the computational architecture behind them. Importantly, there are also important conceptual and experimental design issues that are relevant, where philosophy can play a key role. By highlighting some neurobiological and computational modelling results, we will argue that the currently available evidence favors

a hierarchical processing architecture that confers a crucial, if subtle and specific, role to PFC. After presenting the relevant results, we discuss some methodological and functional implications of this neural architecture supporting conscious experiences. To anticipate, we note that despite the apparent stark differences between conscious and unconscious perceptual processing, available evidence suggests that their neural substrates must be largely shared. This indicates that the difference in neural activity between conscious and unconscious perceptual processing is likely to be subtle and highly specialized. In consequence, imaging techniques that focus only on marked differences between conscious and unconscious level of activity are likely to be insensitive to the relevant neural activity patterns that underlie conscious experiences. Finally, it follows from the evidence we discuss that the functional advantages of conscious over unconscious perceptual processing may be more limited than commonly thought.

1. Finding the neural correlates of consciousness

Scientists study the neural difference between being conscious versus unconscious in at least two different ways. First, they may focus on the neural differences of being conscious versus unconscious overall (e.g. wakefulness, anesthesia, coma, sleep, etc.), also referred to as 'state-consciousness'. Alternatively, they may focus on the neural activity that determines whether someone is conscious of something or not (e.g. seeing or not seeing a face, seeing a face versus seeing a house, hearing or not hearing a sound,

feeling or not feeling pain, etc.). This is often referred to in the literature as ‘content-consciousness’ and it will be the main focus in this chapter.¹

When studying the NCC, scientists seek necessary and sufficient neural events that cause conscious experiences.² However, it has been acutely pointed out that finding necessary conditions for consciousness can be challenging (Chalmers 2000). First, after damage to a specific part of the brain (e.g. stroke, surgery, etc.), mental functions--including consciousness--may be lost. But they may also be recovered thanks to neuroplasticity: the brain’s capacity to “rewire” itself. In some rare cases, cognitive functions and consciousness are never lost at all, even after massive, albeit slow, destruction of neural tissue (Feuillet et al. 2007).

Second, redundancy makes finding necessary conditions for consciousness unlikely. It is not uncommon that the brain has redundant or backup mechanisms for performing the same function. This means that consciousness could be sustained by more than one neural mechanism. If mechanism x causally sustains consciousness, x is undoubtedly an NCC. But consciousness may be overdetermined if mechanisms x and y can cause the same type of conscious event independently. In this case, if x is damaged but y is spared, consciousness would still take place. This would demonstrate that x is not a necessary condition for that type of conscious event, even though it is *ex hypothesi* its neural correlate (or one of them). Thus, preservation of consciousness when a brain

¹ See (Hohwy 2009) for discussion of problems with the study of content- and state-NCC. See (Noe & Thompson 2004) for discussion of problems with the content-NCC approach. For a recent review of state-NCC research, see (Gosseries et al. 2014). Note that neuroscientists’ terms ‘content-’ and ‘state-consciousness’ are often described by philosophers as ‘state-’ and ‘creature-consciousness’, respectively (Rosenthal 1993).

² The term ‘correlate’ falls short from capturing necessary and sufficient conditions. We just follow the terminology used in the field at least since (Crick & Koch 1990).

region is destroyed, impaired or when it does not display any measurable activity does not in and of itself show that normal activity in that region is not an NCC.

Third, convergent evolution could have produced independent mechanisms for consciousness in two species whose common ancestor lacked either mechanism. It may be the case that something as complex as consciousness emerged during evolution just once, but it is not necessary. If different species (say, humans and octopuses) sustain conscious experiences via different types of neural mechanisms, neither would be necessary for consciousness in a strong metaphysical sense. For all these reasons, establishing strict necessary conditions for consciousness is unlikely to be successful. If anything, we can aspire to restricted necessity claims that include clauses like “in humans” or “in normal conditions”.³

Finding sufficient neural conditions for consciousness is not without challenges either. For instance, everything else being the same, the whole brain is likely to be sufficient for sustaining conscious experiences. Yet, postulating the whole brain as the NCC would not be informative. Instead, neuroscientists are interested in the “*minimal set of neural events jointly sufficient for a specific conscious experience (given the appropriate enabling conditions)*” (Koch 2004: 97); or “core realizers” of consciousness for short (see Shoemaker 1981). Delimiting what counts as a core realizer is far from straightforward (Aru et al. 2012; Chalmers 2000). For instance, when comparing a condition in which subjects report being conscious of a stimulus against a condition in which subjects report no consciousness of it, the difference between these two conditions should be conscious awareness only. Yet, distilling stimulation and cognition from

³ Establishing what counts as a normal condition is complicated too, but we sidestep this issue here.

consciousness is not easy. Controlling for stimulation, attention, and performance capacity (e.g. accuracy, reaction time, etc.), such that these are matched across conscious and unconscious conditions is hard to achieve experimentally (Lau 2008; Morales et al. 2015). During imaging experiments, prerequisites (e.g. stimulus processing, attention) and consequences (e.g. performance, attention, working memory, motor preparation, verbal report, etc.) of consciousness can be easily confounded with the actual NCC (Lumer & Rees 1999; Aru et al. 2012; Bachmann 2015; Tse et al. 2005). Using lesion patients for whom performance is constant across subjective judgments of awareness and unawareness without experimental manipulation does not eliminate all the problems. Not only these patients are rare and their deficits are often constrained in specific ways, their lesions are hardly ever limited to clear-cut anatomical or functional regions. Moreover, these patients' brains often rewire and recover functions in peculiar ways, which hinders making general inferences.

A practical limitation when studying the NCC is the methods currently available for detecting neural activity in the relevant functional networks. In the last few decades, sophisticated non-invasive imaging techniques such as functional magnetic resonance imaging (fMRI) have been added to decades-old technology like electroencephalography (EEG), magnetoencephalography (MEG), and positron emission tomography (PET). These technologies, however, have strong limitations with respect to either their spatial or temporal resolutions, or both. They are also indirect measurements of neural activity: oxygenated blood, electrical and magnetic signals measured outside the skull or glucose consumption detected via positron-emitting radioactive tracers. Electricocorticography (ECoG) allows making measurements with better signal-to-noise ratio and good temporal

resolution by placing electrodes directly over the cortex, but it requires risky surgical intervention. For obvious medical and ethical reasons, the use of this technology in humans is very limited. In contrast, direct single- and multi-unit recording of neural activity offers unsurpassable spatiotemporal resolution. Unfortunately, it requires inserting electrodes directly into or right next to neurons, making it an extremely invasive method. In consequence, it is available almost exclusively in other animals like monkeys or rats. Working with animal models offers multiple advantages (Passingham 2009), but the study of consciousness may be challenging even when ingenious solutions have been devised (Leopold & Logothetis 1996; Rigotti et al. 2013). We will come back to some of the limitations of these methods when assessing the available empirical evidence for the NCC.

Finally, restricted necessary and sufficient conditions should ideally be established via causal interventions. By directly manipulating neural activity, we may reveal the causal mechanisms underlying conscious states (Craver 2007; Neisser 2012). Manipulating the brain safely and effectively, however, is a major challenge--especially in humans. Genetic, chemical, and surgical interventions are risky, almost exclusively available in other animals and likely to affect more than just conscious awareness. More promising may be the use of non-invasive technology such as transcranial magnetic stimulation (TMS). TMS pulses project a small magnetic field on to the surface of the brain through a coil placed outside the skull. Depending on the number and frequency of pulses, the magnetic field can enhance or inhibit neural activity in the target region. This allows researchers to create reversible "virtual lesions" for short intervals and test whether the target region was subserving the function of interest, including conscious awareness.

While promising, the precise mechanisms of action of TMS are still poorly understood and its effects can only be coarsely controlled (Sandrini et al. 2011).

2. Theoretical predictions regarding the NCC

Different theories about the nature and localization of the NCC place their explanatory power at different levels (Hardcastle 2000). The emphasis has been sometimes laid on neurochemistry [e.g. activation of the NMDA neuroreceptor that forms large neural assemblies (Flohr 1995)], neuronal types [e.g. spindle neurons (Butti et al. 2013; Allman et al. 2005)], systemic properties [e.g. integrated information (Tononi 2008)], and functional neuroanatomy [e.g. specific neurophysiological markers and neural activity in specific regions or networks; for recent reviews see (Koch et al. 2016; Lau & Rosenthal 2011; Dehaene & Changeux 2011)]. In this section, we briefly discuss some of the main recent functional neuroanatomical theories. In no way this is an attempt at a thorough review. Not only we do not discuss other viable empirical theories of the NCC, we only make succinct presentations of the ones discussed. Rather, our goal is to show that the theories we discuss predict different neural implementations of consciousness, especially regarding the role of PFC, providing an opportunity to arbitrate empirically between several theoretical frameworks.

a. Neural Synchrony Theory

Much of the recent interest in finding the NCC was set off by the introduction of Neural Synchrony Theory (Crick & Koch 1990). According to it, at the psychological level consciousness depends on short-term memory and attention. At the neural level, attention makes groups of relevant neurons to fire in a coherent way giving rise to conscious percepts. Neurons in different areas often fire independently from each other. However, attention can make their firing rates to become synchronized in fast waves (between 40 and 70 times per second). This temporal coherence achieves a global unity imposed on different areas of the brain that activates short-term (working) memory. Crick and Koch hypothesize that this basic oscillatory mechanism underlies all kinds of consciousness (e.g. visual, auditory, tactile, or painful experiences). Thus, the NCC is identified in their theory with a special type of activity (i.e. neural firings oscillating at 40-70Hz). The specific contents of conscious experiences depend on the specialized cortex where the activity takes place. In the case of vision, different features of visual stimuli are processed by different areas of visual cortex (e.g. V1/orientation, V4/color, MT-V5/motion). The brain binds together all these features in a single, coherent, and conscious percept by synchronizing the neural activity in these areas. Moreover, this activity is coordinated by zones in sensory cortices that are rich in feedback neurons (i.e. neurons that project from a higher area to a lower area). These feedback projecting zones also exist in other regions, such as the thalamus or the claustrum, which may play a major coordination role (Crick & Koch 2005). Thus, synchronized firing at about 40-70Hz is proposed as a necessary and sufficient condition for consciousness (provided enabling

conditions such as attention and activation of working memory are met). Importantly, even though the NCC in Crick and Koch's proposal are highly distributed across brain areas, PFC is not predicted to play any significant role in sustaining conscious activity. At most, PFC may be relevant for attention, sustaining contents in working memory, and reporting conscious contents.

b. Two-Visual-Systems Hypothesis

According to an influential theory advanced by Milner and Goodale (1995; 2006), the neural correlates of visual awareness are restricted to activity in the ventral stream of the visual processing pathway. There are corticocortical projections from early visual cortex (V1) that later split into two processing streams (Ungerleider & Mishkin 1982). One stream is located dorsally and ends in parietal cortex, the other stream runs on a ventral pathway that ends in inferior temporal cortex. The Two-Visual-Systems Hypothesis relies on neurophysiological and anatomical evidence in monkeys, as well as neuropsychological evidence in humans, to suggest activity in the dorsal stream is associated with visually-based action (for example, saccades or visually guided hand movements) and egocentric representations (i.e. representations of objects from the subject's point of view). Despite involving complex computations, activity in this stream is not normally available to awareness according to this view. In contrast, activity in the ventral stream is typically associated with allocentric representations (i.e. objective representations independent of the subject's perspective) and visual object recognition. Objective visual representations have shape, size, color, lightness, and location constancies that allow subjects to re-

identify objects independently of viewpoint (Burge 2010). Milner and Goodale argue that “visual phenomenology...can arise only from processing in the ventral stream” (2006: 202). In other words, activity in the ventral stream is necessary for awareness. Additionally, attentional modulation that selects a represented object is required. Object representations in the ventral stream and attention are jointly sufficient for conscious awareness. Importantly, they think prefrontal cortex exert “some sort of top-down executive control...that can initiate the operation of attentional search” (2006: 232), guide eye movements and motor control. However, activity in prefrontal cortex would probably be in and of itself irrelevant for conscious awareness.

c. Local Recurrency Theory

Local Recurrency Theory (LRT) proposes three stages involved in visual information processing. First, after stimulus presentation there is a rapid, unconscious feedforward sweep (~100-200ms) of activity from visual cortex (V1) to motor and prefrontal cortex. Immediately after, in a second processing stage, an exchange of information within and across high- and low-level visual areas starts taking place. This fast and widespread information exchange is achieved by means of so-called recurrent processing, namely, neural activity in horizontal connections within a visual area, and activity in feedback connections from higher level areas back to lower levels (all the way back to V1). Local recurrent processing enables the exchange of information of different visual properties (e.g. orientation, shape, color, motion, etc.) that are processed independently in different visual areas. This facilitates the required “perceptual grouping” (Lamme 2006: 497) for

forming coherent conscious representations of objects. According to LRT, this second stage of recurrent processing is the NCC as it is both necessary and sufficient for phenomenal consciousness (Lamme & Roelfsema 2000; Lamme et al. 2002).⁴ Finally, in a late third stage, this reverberating activity becomes a widespread co-activated network involving visual and fronto-parietal areas through attentional amplification. Motor and prefrontal cortex activity enables response preparation, keeping information in working memory and other types of cognitive control like attending, changing response strategies or inhibiting response. For LRT, this later fronto-parietal activity is required exclusively for report and cognitive control (what Block (2007) calls 'access consciousness'), not for supporting conscious experiences themselves (what Block (2007) calls 'phenomenal consciousness'). One surprising consequence of the view is that conscious experiences take place even if they are not reportable or accessible to the subject (Block 2007; Landman et al. 2003; Sligte et al. 2008; Vandenbroucke et al. 2015). In other words, it would be possible to be conscious without knowing it and without any possible behavioral and cognitive manifestation of such phenomenal experiences.⁵ In many cases, according to LRT, when subjects report unawareness, they may just be reporting their lack of access to otherwise conscious experiences.

⁴ "That recurrent processing is necessary for visual awareness is now fairly well established, and supported by numerous experiments." (Lamme 2010: 216) "According to such empirical and theoretical arguments, [local recurrent processing] is the key neural ingredient of consciousness. We could even define consciousness as recurrent processing." (Lamme 2006: 499)

⁵ See (Kouider et al. 2012; Cohen & Dennett 2011) for criticisms of the scientific viability of this position.

d. Global Workspace Theory

According to Global Workspace Theory (GWT), after stimulus presentation, activity in visual areas starts accumulating in two independent processing streams, one that can lead to consciousness and another that supports unconscious processing (Del Cul et al. 2009; Charles et al. 2013; Charles et al. 2014).⁶ Evidence accumulation through visual information processing in each stream races to a threshold in a “winner-takes-all” fashion (Shadlen & Kiani 2013; Pleskac & Busemeyer 2010; Wald 1947). If activity in the conscious stream reaches its threshold first, a sudden ignition “mobilizes” perceptual representations to a widespread global workspace implemented in frontoparietal interconnected neurons. This global broadcasting makes visual representations available for report and cognitive control, which results in a visual conscious experience (Dehaene & Changeux 2011; Dehaene & Naccache 2001). It is this globally broadcasted activity that GWT identifies as the NCC (Dehaene et al. 2006). Simultaneously, an unconscious stream processes the same visual stimulus. In case global ignition fails, the perceptual representation in the unconscious stream can be used if the subject is forced to provide a response, accounting for the commonly-observed capacity of subjects to perform above chance even when they are unaware of stimuli. Global workspace theorists appeal to a wealth of studies showing that all sorts of cognitive processing can be performed unconsciously to a certain extent: visual judgments, word meaning extraction, performing simple arithmetic operations, cognitive control, etc. (Dehaene et al. 2014). Note that this

⁶ Not to be confused with the dorsal and ventral streams discussed by the Two-Visual-Systems Hypothesis. According to GWT, the conscious and unconscious streams may be implemented in largely overlapping anatomical regions in visual areas.

dual-stream approach makes the surprising assumption that every stimulus is processed twice simultaneously, which imposes stringent and possibly unnecessary computational requirements on the brain.

Global workspace theorists note that unconscious performance and neural activity associated to it are rarely at the same level as during conscious conditions. Thus, global ignition provides a necessary and minimally sufficient signature of consciousness, which according to the view, increases and maintains performance and cognitive flexibility. This signature is identified by GWT with frontoparietal activity in fMRI studies and with sudden, widespread activity in a late (~270-650ms) positive voltage in frontoparietal areas in EEG studies (also known as the P300 component) (Del Cul et al. 2007; Sergent et al. 2005; Lamy et al. 2009).

e. Higher Order Theory

The Higher Order Theory (HOT) of consciousness holds that a mental state is conscious by virtue of its relation to some higher-order state. A perceptual representation alone is never in and of itself conscious. Rather, it becomes conscious when it is somehow “tagged” or meta-represented by another, higher-order state. According to some versions of HOT, this relation is achieved by means of the higher-order state’s representing the first-order state in ways similar to thought or perception (Rosenthal 2004). What different versions of higher order theories have in common is that “a mere change in the higher order representation or process is sufficient to lead to a change in subjective awareness, even if all first-order representations remain the same” (Lau & Rosenthal 2011: 365).

HOT holds that first-order representations depend on neural activity in early visual areas, whereas higher-order processes are implemented mainly in prefrontal (and parietal) cortex in both human and other primates (Lau & Rosenthal 2011). More specifically, consciousness emerges from a hierarchical processing architecture in which unconscious visual information processed in early areas gets selected by downstream mechanisms in PFC. One of HOT's main predictions, then, is that disrupting the activity responsible for sustaining higher-order processes in prefrontal cortex should affect or eliminate visual experiences without affecting performance (because performance is driven mainly by unconscious first-order representations in early sensory cortex). Importantly, disruptions to PFC should affect conscious experiences themselves, not just report or access to visual experiences, as expected by LRT. In contrast to GWT, HOT does not expect global activity to be predictive of conscious awareness. PFC activity related to consciousness may be very subtle as it just needs to select relevant visual processes in early areas. Thus, HOT predicts that massive alterations to PFC may not be sufficient to disrupt consciousness as long as specific PFC activity is preserved. Perhaps more surprisingly, some versions of HOT predict that specific activity in PFC is necessary and minimally sufficient for consciousness. In other words, if the 'tagging' activity normally responsible for consciousness takes place in the absence of a 'tagged' state, conscious experiences may still occur.

In summary, these theories make very different general predictions about the nature and location of the NCC. They also make very different specific predictions regarding the role of PFC in consciousness, behavior, and the computational architecture underlying

conscious processing. Neural Synchrony Theory, Two Visual Systems Hypothesis, and Local Recurrency Theory focus on activity in sensory areas in fact denying any role in consciousness for PFC.⁷ GWT accepts PFC plays an important role, emphasizing the heightened level of activity and its distribution through frontoparietal areas. In contrast, HOT confers PFC a dominant role in consciousness because of the specific and subtle function it plays within a hierarchical processing architecture.

A clear sign of progress in the scientific quest for the neural correlates of consciousness is that despite their initial popularity, some theories are completely abandoned in light of subsequent evidence. The Neural Synchrony Theory, for example, has lost credibility thanks to multiple studies finding oscillations at 40Hz in the absence of awareness and failing to detect these same oscillations during reports of conscious experiences (for a brief review, see Koch et al. 2016). The Two-Visual-Systems Hypothesis (at least with respect to its commitment to the ventral stream being the NCC) has also been subject of strong skepticism after considering the mounting evidence against the independence of the dorsal and ventral streams and their proposed clear-cut roles (Wu 2014; Briscoe & Schwenkler 2015).

In the next two sections, we discuss neuroscientific and computational evidence relevant for arbitrating between the theoretical frameworks of the other three theories discussed in this section--LRT, GWT and HOT--and their predictions regarding the NCC and PFC's involvement.

⁷ Neural Synchrony Theory and Local Recurrency Theory further specify that consciousness is associated with a specific type of feedback activity.

3. The NCC: Evidence for PFC's involvement

Activity in PFC is crucial for supporting conscious perceptual experiences.⁸ Multiple neuroimaging studies have systematically found increased activity in prefrontal and parietal cortex when comparing conscious versus unconscious conditions, often even when performance capacity is controlled for (Lau & Passingham 2006; Sergent et al. 2005; Gross et al. 2004; Dehaene et al. 2001; for recent reviews see Dehaene & Changeux 2011; Lau & Rosenthal 2011; Odegaard et al. 2017; Boly et al. 2017). Some researchers minimize PFC's importance in the NCC arguing that it plays an important function in attention, report, and cognitive control, but that it has a negligible role in consciousness (Koch et al. 2016; Tsuchiya et al. 2015). While these ideas are not new (Lumer & Rees 1999; Tse et al. 2005), they have sparked a renewed interest in the topic.

Admittedly, interpreting imaging results can be challenging. During an imaging experiment, reasons other than a causal role in supporting conscious experiences might lead to statistically significant results (e.g. noise or different functions performed by the same areas). As discussed in section 1, a more robust way of determining if an area of the brain is necessary for supporting a function is to permanently or temporarily impair it. If the function is lost, a constrained necessity claim may be warranted. Relatedly, if the function is not lost, not only constrained necessity claims are harder to maintain, the non-affected areas become candidates for being sufficient for supporting that function.⁹ With

⁸ For simplicity we refer collectively to PFC, but activity relevant for consciousness is likely to be found in more specific areas, such as dorsolateral PFC, insula, and other orbitofrontal and rostrolateral regions.

⁹ Necessity claims or denials in this context have to be constrained for the reasons discussed in the first section. Other species may implement consciousness differently, preventing any unconstrained necessity claim. But, perhaps more importantly for the neuroscientific study of consciousness, failures to eliminate a function--consciousness in this case--need not imply that the area was not necessary (in a constrained

this logic in mind, recent studies with carefully controlled psychophysical methods have investigated how PFC lesions (Del Cul et al. 2009; Fleming et al. 2014) and temporarily induced impairments by transcranial magnetic stimulation (Rounis et al. 2010) impact visual experiences. The results of these studies have been univocal: permanent and temporary impairments to PFC do not abolish objective visual task performance capacity, while they affect subjective judgments. Either the percentage of visible stimuli decreased despite constant performance (Rounis et al. 2010; Del Cul et al. 2009) or these subjective judgments became less diagnostic of task performance (Fleming et al. 2014). In the case of lesion patients, the capacity to use subjective ratings to diagnose task performance (i.e. metacognitive capacity) was impaired by 50% (Fleming et al. 2014).

Nevertheless, several objections are often raised against this evidence. First, it is argued that these impairments only affect subjective judgments mildly, while damage to early visual areas like V1 abolish visual consciousness completely; second, that PFC does not represent conscious content specifically, which confers it a limited role (if any); and, third, that the activity detected in PFC during imaging studies pertain to attention and report, not consciousness per se. We address these objections in order.

way) for supporting the function. The impairment might not have been specific enough or the brain might have repurposed other circuits to implement that function which, otherwise, would have been implemented in the impaired area under normal conditions.

a. PFC activity related to consciousness is highly specific

Lesions to V1, in fact, can often completely abolish visual experiences (Melnick et al. 2016; Weiskrantz 1997). When V1 is affected, like in blindsight, the sensory signal is degraded to the point of preventing subjective judgments of consciousness. In blindsight patients, the lateral geniculate nucleus (LGN) is spared. This relay center of visual information from the retina to early visual areas in the occipital lobe is located in the thalamus, and is likely responsible for driving objective performance of blindsight patients (Schmid et al. 2010). This does not rule out that in normal cases proper functioning of early visual areas is necessary, even if not sufficient, for consciousness.

A second point to highlight is that PFC functions very differently from sensory cortices. For instance, neuronal coding in PFC is relatively distributed, is rarely linear and shows a high degree of mixed selectivity (Mante et al. 2013; Rigotti et al. 2013). This means that, unlike visual cortex whose function is highly specialized for processing visual information, PFC's role in consciousness is performed by highly specific patterns of activity as it is responsible for carrying out many other functions as well. Therefore, to exclusively produce a large disruption of perceptual experience, neural patterns of activity in PFC would need to be affected in highly specific ways.

Relatedly, frontal and parietal cortices are densely connected and frontal regions display high neuroplasticity (Miller & Cohen 2001; Andersen et al. 1985; Barbas & Mesulam 1981; Cavada & Goldman-Rakic 1989; Petrides & Pandya 1984; Croxson et al. 2005). This implies that the brains of patients with frontal impairments can rewire rapidly by the time they can be tested, often several months after the lesion. Lesions produced

by trauma, stroke or ablation are often too unspecific, but sometimes they are extended enough to likely include all regions responsible for consciousness. However, because these same regions support many central cognitive functions (Badre & D'Esposito 2009; Miller 2000; Duncan & Owen 2000; Passingham & Wise 2012), patients may be so generally impaired that testing them immediately following the brain damage may not be straightforward (Knight & Grabowecky 1995; Mettler 1949). As further support for this point, chemical inactivation in rodent and monkey PFC and regions strongly connected to PFC (e.g. pulvinar) lead to strong effects in subjective confidence judgments without affecting performance in perceptual and even memory tasks. In these cases, the animals are tested immediately after PFC or pulvinar are inactivated, preventing compensatory rewiring (Lak et al. 2014; Romanski et al. 1997; Shipp 2003; Komura et al. 2013; Pessoa & Adolphs 2010; Miyamoto et al. 2017). This background makes the specific effects of lesions or temporary impairments of PFC on subjective judgments indeed quite robust.

b. PFC encodes specific content

Another recent objection is that PFC activity does not encode specific content (Koch et al. 2016), making its role as the NCC likely to be limited. First, specific content representation of visual experiences in PFC is not explicitly predicted by all theories. For instance, PFC may enable conscious perception through connections to early visual areas where the specific content is supported (Lau & Rosenthal 2011). Second, and perhaps more importantly in terms of interpreting the available neuroscientific evidence

correctly, denying that PFC represents explicit contents of conscious experiences is empirically unsupported.

Researchers often perform simple contrastive univariate analysis with fMRI data. In this kind of analysis, the overall levels of activity belonging to one experimental condition is simply compared to (subtracted from) the overall levels of activity in another condition (e.g. conscious versus unconscious trials). But it is known that univariate fMRI analysis provides limited sensitivity. As mentioned above, activity in PFC is hardly linear and neurons exhibit mixed selectivity, which varies widely upon contextual changes. Measuring the overall levels of activity is at best a coarse approximation to total neural activity. Hence, visual content supported by specific patterns of activity may only be decoded effectively with careful analysis and sophisticated modelling strategies (Ester et al. 2015; Stokes 2015). This includes multivariate analyses that go beyond a simple subtraction of overall activity. One example of this is multi-voxel pattern analysis (MVPA), where a decoder is trained to classify the *patterns* of activity in two conditions of interest. For example, if subjects are presented with two types of stimuli in different trials, say, houses and faces, the decoder can be trained to distinguish between patterns of activity pertaining to houses and patterns pertaining to faces. A successful decoder classifies above chance a novel set of data (usually data from the same subject that was not used during training) as belonging to house- or face-trials. MVPA reveals that perceptual content can be decoded from PFC in a simple perceptual decision task (Cortese et al. 2016), and that the pattern of activity in PFC reflects specific perceptual content even under several straining conditions (Wang et al. 2013). In another recent study, patterns

of activity specific to subjective confidence judgments in perceptual and memory trials were successfully decoded from PFC (Morales et al. Forthcoming).

Finally, it could be objected that the spatiotemporal resolution of fMRI offers only a limited insight into neural activity, even when these sophisticated multivariate analyses are used. After all, it only gives us access to ~2 second snapshots of indirect blood-oxygen level dependent (BOLD) activity driven by the hundreds of thousands of neurons found in each voxel (i.e. the minimum resolution in fMRI, equivalent to a 3D pixel of approximately 3 x 3 x 3mm). However, direct single- and multi-unit neural activity recording in monkeys offer a significantly higher spatiotemporal resolution (i.e. in the order of milliseconds and down to a single neuron) and multiple studies have unambiguously confirmed that specific perceptual decisions can be decoded from PFC (Kim & Shadlen 1999; Rigotti et al. 2013; Mante et al. 2013).

c. PFC is crucial for consciousness, not just attention or report

Together, the aforementioned evidence indicates that activity in PFC is necessary for visual consciousness. However, most of the fMRI studies mentioned above involved subjects explicitly reporting their conscious experience. A legitimate worry is that this activity does not reflect conscious perception *per se* and that, rather, it is confounded by the task demand to report or attend the stimulus (Koch et al. 2016; Tsuchiya et al. 2015). Some of these concerns have been recent rekindled by neuroimaging studies where subjects were not required to make explicit subjective judgments about visual stimuli and

activity in prefrontal cortex previously related to consciousness was significantly diminished or undetected (Tsuchiya et al. 2015; Frässle et al. 2014).

The issues concerning limited sensitivity of methods commonly used in fMRI studies, specifically univariate analysis concerning PFC, are relevant here. Using more sensitive methods in humans, such as direct intracranial electrophysiological recording (electrocorticography, or ECoG), reveals activity related to visual consciousness in PFC even when subjects were not required to respond to the stimulus (Noy et al. 2015). Perhaps more importantly, in direct neuronal recordings in nonhuman primates who viewed stimuli passively, activity specifically related to the stimulus was detected in PFC (Panagiotaropoulos et al. 2012). It could be argued, however, that even under passive viewing an over-trained animal may still attend the stimuli or implicitly prepare a report (which could increase prefrontal activity for reasons unrelated to consciousness). But even unreported features of a visual stimuli can be decoded from PFC activity. That is, even when the animal had to report on a different, orthogonal stimulus feature, the unattended and unreported feature was encoded in PFC (Mante et al. 2013). It is very unlikely that the monkeys prepared to attend or report on both features, especially considering that the task was challenging and involved near-threshold stimuli.

It is important to note that this does not mean that in studies of conscious perception making explicit reports does not further drive activity in PFC. PFC activity is involved in all sorts of higher cognition, not just conscious awareness. But this is consistent with the hypothesis that most univariate imaging techniques will only reveal the most heightened activity. It is also consistent with the observation by Noy and colleagues (2015) that their positive ECoG findings in PFC were subtle when no report

was required. Still, in more direct recordings unreported stimulus features were robustly decoded, almost at the same level as attended and reported features (Mante et al. 2013). Thus, we conclude that objections from the so-called ‘no-report’ paradigms may have been exaggerated.

In summary, the important role of PFC in visual conscious experiences resists common objections. As anticipated in the first section, when looking for the NCC, methodological hurdles have to be considered with utmost care. When studying consciousness, non-invasive tools like fMRI may seem ideal for making inferences about neural function in humans. However, its spatiotemporal limitations as well as the prevalence of simple statistical approaches should give us pause, especially when confronted with null findings. When ECoG and single- and multi-unit cell recordings along with multi-voxel pattern decoding analysis are incorporated, the picture that emerges is that activity in PFC is a serious candidate for being the NCC. We note that this is incompatible with the main predictions made by LRT. Also, despite predicting an involvement of PFC during global ignition, GWT’s requirement of global, heightened activity does not fit well with the evidence presented in this section. This evidence points towards a more subtle and specific role of frontal activity during conscious awareness. HOT also predicts an important role of PFC as the NCC but, in contrast to GWT, it does not require the relevant activity to be particularly heightened or distributed.

4. The architecture of the NCC: computational considerations

Neuroimaging as well as direct cortical recordings offer evidence for determining where activity supporting conscious experiences is located in the brain. Multivariate analyses can even distinguish specific patterns of conscious and unconscious activity, rather than merely detecting a difference in levels of activity. Nevertheless, finding the NCC is not only a 'localization' problem. At the level of analysis we are focusing on, it also involves finding the computational architecture most likely to account for the available neurophysiological and behavioral evidence. Computational modelling offers a non-invasive, formal way of comparing different models' capacities to account for behavioral data obtained in normal experimental conditions. Unlike neuroimaging and neurophysiology, where different conditions prevail across different experiments, in computational modelling the same data from a single experiment can be fed to a range of models. This is especially important for comparing the likelihood of rival possible computational architectures of the NCC, giving them an equal chance to fit the data.

Some possible models of how perceptual processing and conscious processes interact in the brain are directly ruled out by the neurophysiological evidence. For example, a model that does not predict unconscious and conscious perceptual processing to take place in two distinct regions, like the one implied by LRT, is not particularly promising when evidence of the importance of frontal regions for visual consciousness is considered. Nevertheless, multiple computational architectures may be compatible with the extant neurophysiological evidence that privileges PFC. Unconscious and conscious processes could be instantiated in different fashions. For example, on one

model these distinct processes could operate in parallel. On another model, perceptual conscious processing could operate hierarchically such that later activity associated with consciousness operates as if evaluating the quality of unconscious visual processes.

We explore this issue with the illustrative case of experiments in which performance is matched while subjective judgments differ. Humans and some nonhuman animals make perceptual decisions about the external world all the time, and they are also capable of making subjective judgments regarding the quantity, quality or reliability of their evidence regarding such perceptual decisions (e.g. by making one decision over another, by extending or suspending a search for resources, by providing visibility or confidence ratings, by placing bets regarding their likelihood of being correct, etc.) (Smith 2009; Beran et al. 2012; Fleming & Frith 2014).

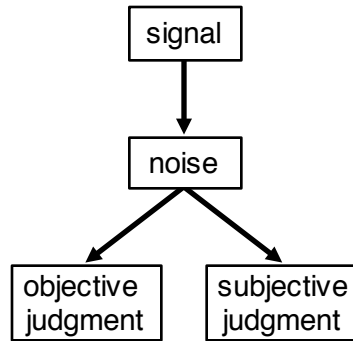
Notoriously, objective perceptual decisions and subjective judgments about the stimuli can come apart in the laboratory and in clinical contexts. For instance, blindsight patients can objectively discriminate visual stimuli while denying having any subjective experience of them (Weiskrantz 1997). In experimental conditions, humans (Maniscalco & Lau 2016; Rounis et al. 2010; Lau & Passingham 2006; Vlassova et al. 2014; Rahnev et al. 2011) and some other animals (Lak et al. 2014; Komura et al. 2013; Fetsch et al. 2014) can exhibit similar dissociations: subjects achieve comparable performance levels in a perceptual task while providing different subjective reports in different conditions. For example, in masking experiments (Maniscalco & Lau 2016; Del Cul et al. 2009; Lau & Passingham 2006), long and short gaps between stimulus presentation and the presentation of a mask allow subjects to identify the stimulus correctly at similar rates, while their subjective ratings of how visible the stimulus was differ significantly. These

dissociations offer a unique opportunity to assess the specific processes involved in consciousness while distinguishing them from mere perceptual processing.

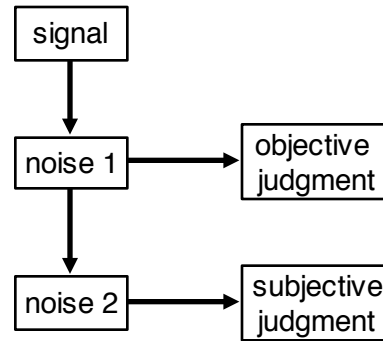
Here we consider three models recently used to fit data from a masking experiment (Maniscalco & Lau 2016): a single-channel, a dual-channel, and a hierarchical model (Figure 1). The single-channel model holds that subjective and objective judgments are different ways of evaluating the same underlying evidence generated by a single perceptual process. This sensory evidence consists on the sensory signal that arises in the brain after stimulus presentation plus the internal noise always present in neural processing. This sensory evidence is processed by the perceptual system and both objective and subjective systems tap into the same processing stream.

According to the dual-channel model, objective perceptual judgments are based on the same sensory evidence as subjective judgments when the subject is conscious of the stimulus, while unconscious perceptual judgments are based on an independent, parallel source of evidence. 'Conscious' and 'unconscious' streams receive the same sensory signal but this gets affected independently by different sources of noise. If the conscious processing stream reaches a threshold first, the stimulus is classified by the brain as 'seen' and the sensory evidence is amplified and made available in working memory for further cognitive control (e.g. making a perceptual judgment about the stimulus and report that it was consciously seen). If the consciousness threshold is not crossed, the stimulus is classified by the brain as 'not seen' and the evidence accumulated in the conscious channel is discarded. If the subject still has to provide an answer--for instance, if prompted by the experimenter--the sensory evidence accumulated in the unconscious channel is used to provide a forced response.

Single-channel



Hierarchical



Dual-channel

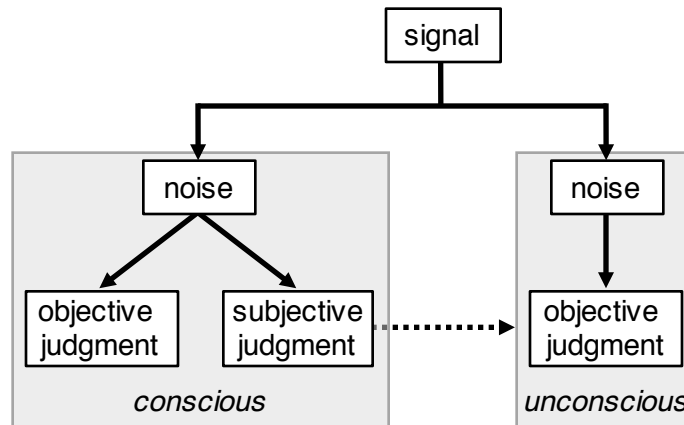


Figure 1. Diagrams of three computational models of objective and subjective judgments. *Single-channel model.* The same evidence (sensory signal + internal noise) gives rise to objective and subjective judgments. *Dual-channel model.* Two parallel streams of conscious and unconscious perceptual processing run simultaneously, each influenced by independent sources of noise. If the subjective judgment is given the lowest rating (e.g. “not seen”, “not confident”, “guess”) the unconscious stream is used for objective classification, otherwise the conscious stream is used. *Hierarchical model.* Objective and subjective judgments are driven by different processes organized in a serial hierarchy. An early stage produces objective judgments and a later stage of processing produces subjective judgments, as if evaluating the quality of the former. The second stage inherits the noise of the first, influenced by the early stage, but not vice versa.

Finally, according to the hierarchical model, the sensory evidence available for objective and subjective judgments differ, but it is not independent. The sensory signal (plus noise) is used to make objective perceptual judgments. Then, subsequent processing of this same evidence, in addition to a new source of noise, is used to make subjective judgments (Cleeremans et al. 2007; Fleming & Daw 2017). Thus, the accumulated evidence at the late stage might become degraded by the time it is tapped by subjective mechanisms due to signal decay or accrual of noise, or it may be enhanced due to further processing.

We note that these models have been proposed based on conceptually reasonable grounds. In other words, finding which fits the empirical data better provides us with substantial insight regarding the computational architecture behind conscious perception. After performing formal model comparison, Maniscalco & Lau (2016) found that the hierarchical model provided the best and more parsimonious fit to the data of the metacontrast masking experiment, and it was also superior in reproducing the empirical data pattern in a series of simulations. The hierarchical model was able to account for the dissociation between performance and subjective visibility ratings by supposing that early-stage perceptual processing is better transmitted to late-stage processing when the gap between stimulus and mask is longer. Since the early stage influences task performance and the late stage governs subjective judgments, longer gaps allow more evidence accumulation. This results in higher subjective visibility judgments in trials with longer gaps between the stimulus and the mask than in trials with short gaps, in spite of having similar task performance.

The last point is of importance for arbitrating between the theories of consciousness discussed in the previous sections. LRT does not make the prediction that the manipulation of the second processing stage changes subjective judgments without affecting task performance, bearing more functional resemblance to a single-channel model. Although GWT allows for unconscious above-chance performance, it does not predict unconscious performance will be at the same level if global workspace activity, likely implemented in fronto-parietal regions, is disrupted. Some global workspace theorists explicitly endorse this dual-channel model which, at least for the masking dataset reported above, does not account well for the dissociation of objective and subjective judgments (Del Cul et al. 2009; Charles et al. 2013; Charles et al. 2014). The dual-channel model espoused by GWT, then, does not aptly account for the data presented in the previous section, where altering PFC normal functioning affects subjective judgments but preserves performance at normal levels (Rounis et al. 2010; Fleming et al. 2014). In any case, the idea of perfectly parallel processing routes for conscious and unconscious visual stimuli is unlikely to reflect the real neural circuitry involved in visual processing. As discussed above, for a long time the dorsal and ventral streams of visual processing were taken to be exclusively involved in unconscious and conscious visual representation, respectively (Milner & Goodale 2006). However, information within both streams is likely to be integrated (Wu 2014), and unlikely to be sufficient for consciousness. In sharp contrast to LRT and GWT, HOT predicts that late stage activity can be disrupted without affecting task performance. HOT explicitly proposes that downstream brain areas like PFC render sensory activity conscious by

evaluating it. This puts HOT in close functional proximity to the hierarchical model, whose performance was far superior to the other two.

It is important to note that these results are limited to the analyzed dataset in Maniscalco & Lau (2016) and only further testing may confirm whether they generalize to other datasets, other experimental paradigms, or the hierarchical model outperforms other models. Nevertheless, it is also important to highlight that these results fit well with the data presented in the previous section according to which activity in PFC is crucial for conscious experiences. The second stage in the hierarchical model may be played by specific patterns of activity in PFC, while the earlier processing stage takes place in early visual areas.

5. Further implications for the NCC

The neuroscientific and computational evidence presented in the previous sections suggests that the NCC may be found in a hierarchical processing architecture of perceptual signals in the brain. In this section, we explore some relevant implications of this architecture of the NCC.

a. Conscious and unconscious neural circuitry is largely shared

The Hierarchical model favored by the formal model comparison results holds that unconscious and conscious objective performance is based on the same perceptual

evidence. Combined with available neuroscientific evidence, this suggests early visual and association areas support objective judgments while PFC taps onto this evidence later in the processing hierarchy, as if evaluating it, to give rise to consciousness. One consequence of this architecture is that, as far as visual information processing is concerned, unconscious and conscious mechanisms are mostly shared. PFC conscious-related engagement with visual representations constitutes only a late portion of the conscious processing stream, otherwise shared with unconscious representations. This important realization should impact how we study consciousness as well as how we think about the function of consciousness.

b. Distinguishing conscious and unconscious activity requires subtle methods

The fact that these mechanisms are largely shared points towards a subtle difference between conscious and unconscious processing. When controlling for stimulus strength and performance in an experimental setting, which is crucial for discovering the NCC, neural activity levels are not likely to differ greatly between conscious and unconscious trials. Activity in PFC is often not linearly correlated with behavior or stimulus properties and frontal neurons often have mixed selectivity properties that code distinct properties on a highly contextual manner (Rigotti et al. 2013; Mante et al. 2013). This suggests that we need to be very careful when interpreting results of purported elevated and distributed activity in conscious conditions (Lamy et al. 2009; Koivisto et al. 2016; Koivisto & Grassini 2016; Railo et al. 2011; Pitts et al. 2014; Dehaene 2014). In some of these experiments, it is often the case that stimulus strength and performance is inadequately controlled for

and, sometimes, dated conceptions of the nature of perception hinder the interpretation of these results (Morales et al. 2015). For instance, it is easy to mistakenly include activity related to objective stimulus processing as part of activity responsible for consciousness.

The interpretation of null findings also demands caution. Detecting subtle neural activity specifically involved in consciousness requires highly sensitive methods. Current, non-invasive imaging technologies like univariate fMRI, MEG, or EEG are not ideal for such task as they are only able to detect the strongest signals from the brain. Because of their particular limitations and their indirect nature, subtle yet critical activity in prefrontal cortex is easily missed when comparing activity from conscious and unconscious conditions. In other words, while there may be nothing wrong with positive results when these methods detect strong activity in prefrontal cortex, we should be conservative about the meaning of null findings. The computational and empirical evidence gathered from more powerful methods suggest that, for the most part, only subtle and highly specific patterns of activity are relevant for consciousness. It should not be surprising then, that crude methods--advanced as they are--turn out to be often unsuited for detecting critical activity for consciousness in PFC.

c. The function of consciousness may be limited

If the mechanisms for unconscious and conscious processing are mostly shared and their difference is expected to be subtle and specific, it is possible that consciousness per se does not contribute significantly to visual information processing, task performance or behavior in general (Rosenthal 2008; Rosenthal 2005; Robinson et al. 2015). It is hardly

contested that the brain can perform lots of perceptual and cognitive tasks unconsciously (Phillips 2016; Peters & Lau 2015): anything from stimulus detection (Tsuchiya & Koch 2005) and word identification (Dehaene et al. 2001), to processing word meanings (Luck et al. 1996; Gaillard et al. 2006) or performing basic arithmetic (Van Opstal et al. 2011). Even high-level cognitive functions, like cognitive control (Koizumi et al. 2015) or working memory (Samaha et al. 2016) show no apparent benefit from conscious awareness in controlled experimental conditions.

Denying the role of consciousness in behavior might strike as rather counterintuitive. Conscious experiences, it would seem, allow us to make fine-grained discriminations and to increase performance, and even to form beliefs, reason, and act (Tye 1996). In fact, in experiments showing above-chance performance in unconscious trials, the effects tend to be small and elicited only in forced-choice contexts. However, unconscious stimuli often differ from conscious ones in other ways besides consciousness. For instance, stimuli are often rendered unconscious by weakening perceptual stimulation (e.g. lower contrast, shorter presentation, higher noise, inattention, etc.), which has the effect of reducing the signal-to-noise ratio of the perceptual evidence. A lower signal-to-noise ratio alters first-order representations, expectedly decreasing performance capacity and the effect of attentional magnification. In these cases, it is the decreased signal-to-noise ratio elicited by the stimulation conditions rather than the stimulus being unconscious what accounts for the difference in performance capacity. This is why it is crucial to insist that performance capacity is a confound that needs to be controlled for when searching for the NCC (Morales et al. 2015). This, of course, is not to deny consciousness has *some* function; although it does invite to rethink what the

functions of consciousness might be (e.g. the initiation of action or availability for rational thought). Here we just point out that it is not a necessary trait of conscious experiences to enable better performance than during unconscious processing.

6. Conclusions

The current science of consciousness is gradually achieving maturity. Fair assessments of empirical evidence related to the NCC, however, require subtle and thorough theoretical work. Determining necessary and sufficient neural conditions for consciousness goes beyond merely 'mapping' conscious-related activity (or lack thereof) onto certain brain areas. First, detecting or failing to detect activity in a brain area is not immediately uncontroversial evidence in favor or against that area being the NCC. For instance, activity in certain areas during conscious conditions may be confounded with activity of some other cognitive capacities related to performance, attention or cognitive control. Also, activity supporting consciousness in normal situations may be subtle and, hence, hard to detect with traditional methods. In consequence, scientists and philosophers need to be cautious as a few null results may not be sufficient for ruling out certain area as an important NCC. Second, a simple mapping of relevant brain areas is insufficient for explaining the overall computational architecture supporting consciousness. Even if certain brain area is found to be related to consciousness, activity in that area could be consistent with different processing architectures. So, the NCC is probably better understood as brain-wide interconnected processing rather than isolated activity in a single brain area.

Importantly, the empirical efforts behind the search for the NCC go beyond functional localization as they can also shed light on theoretical issues. As different theories make distinct predictions regarding the neurofunctional and computational architecture involved in consciousness, we can use empirical findings to arbitrate between these theories. Here we found that the main predictions made by the Local Recurrency Theory regarding the NCC are not supported by current available evidence. A vast body of evidence using different methodologies privileges PFC as a crucial area for consciousness, which is incompatible with its central predictions. In contrast, both Global Workspace and Higher Order Theories predict PFC must have a major function in conscious awareness. We argued from a study involving a formal model comparison that a hierarchical computational model akin to HOT's prediction of a serial processing stream is better supported than a dual-channel model akin to some versions of GWT's prediction that objective and subjective processes are implemented in parallel. While this result is limited to the analyzed dataset, when considered along the systematic findings of PFC relevant role for consciousness, confidence in a hierarchical implementation of the NCC may be bolstered.

Finally, the data we presented point towards some important, although perhaps unexpected, features of the study of the NCC and consciousness itself. Firstly, we argued that the neural activity involved in conscious and unconscious perception may be largely shared. This suggests that the NCC involve subtle activity differences from unconscious processing which are detectable only by highly sensitive methods. Secondly, the function of consciousness may be limited. While a subtle difference in neural activity does not necessarily imply a subtle difference at the psychological, behavioral, or phenomenal

level, it does make it a possibility. Only future research will be able to confirm or reject this hypothesis.

References

- Allman, J.M. et al. (2005), Intuition and Autism: A Possible Role for Von Economo Neurons. *Trends in Cognitive Sciences*, 9(8): 367–373.
- Andersen, R.A., Asanuma, C. & Cowan, W.M. (1985), Callosal and Prefrontal Associational Projecting Cell Populations in Area 7A of the Macaque Monkey: A Study Using Retrogradely Transported Fluorescent Dyes. *The Journal of Comparative Neurology*, 232(4): 443–455.
- Aru, J. et al. (2012), Distilling the Neural Correlates of Consciousness. *Neuroscience and Biobehavioral Reviews*, 36(2): 737–746.
- Baars, B.J. (2005), Global Workspace Theory of Consciousness: Toward a Cognitive Neuroscience of Human Experience. In S. Laureys (Ed.), *Progress in Brain Research* (Elsevier), 45–53.
- Baars, B.J. (1997), In the Theatre of Consciousness. Global Workspace Theory, A Rigorous Scientific Theory of Consciousness. *Journal of Consciousness Studies*, 4(4): 292–309.
- Bachmann, T. (2015), On the Brain-Imaging Markers of Neural Correlates of Consciousness. *Frontiers in Psychology*, 6: 868.
- Badre, D. & D'Esposito, M. (2009), Is the Rostro-Caudal Axis of the Frontal Lobe Hierarchical? *Nature Reviews. Neuroscience*, 10(9): 659–669.
- Barbas, H. & Mesulam, M.M. (1981), Organization of Afferent Input to Subdivisions of Area 8 in the Rhesus Monkey. *The Journal of Comparative Neurology*, 200(3): 407–431.
- Beran, M.J. et al. (2012), *Foundations of Metacognition* (Oxford University Press).
- Block, N. (2007), Consciousness, Accessibility, and the Mesh Between Psychology and Neuroscience. *Behavioral and Brain Sciences*, 30(5–6): 481–99.
- Boly, M. et al. (2017), Are the Neural Correlates of Consciousness in the Front or in the Back of the Cerebral Cortex? Clinical and Neuroimaging Evidence. *The Journal of Neuroscience*, 37(40): 9603–9613.
- Briscoe, R. & Schwenkler, J. (2015), Conscious Vision in Action. *Cognitive Science*, 39(7): 1435–1467.
- Burge, T. (2010), *Origins of Objectivity* (Oxford University Press).
- Butti, C. et al. (2013), Von Economo Neurons: Clinical and Evolutionary Perspectives. *Cortex*, 49(1): 312–326.
- Cavada, C. & Goldman-Rakic, P.S. (1989), Posterior Parietal Cortex in Rhesus Monkey: II. Evidence for Segregated Corticocortical Networks Linking Sensory and Limbic Areas with the Frontal Lobe. *The Journal of Comparative Neurology*, 287(4): 422–445.
- Chalmers, D.J. (2000), What Is a Neural Correlate of Consciousness. In T. Metzinger (Ed.) *Neural Correlates of Consciousness* (MIT Press), 17–40.
- Charles, L. et al. (2013), Distinct Brain Mechanisms for Conscious Versus Subliminal Error Detection. *Neuroimage*, 73: 80–94.
- Charles, L., King, J.-R. & Dehaene, S. (2014), Decoding the Dynamics of Action, Intention, and Error Detection for Conscious and Subliminal Stimuli. *The Journal of Neuroscience*, 34(4):

1158–1170.

- Cleeremans, A., Timmermans, B. & Pasquali, A. (2007), Consciousness and Metarepresentation: A Computational Sketch. *Neural Networks*, 20(9): 1032–1039.
- Cohen, M.A. & Dennett, D.C. (2011), Consciousness Cannot Be Separated From Function. *Trends in Cognitive Sciences*, 15(8): 358–364.
- Cortese, A. et al. (2016), Multivoxel Neurofeedback Selectively Modulates Confidence Without Changing Perceptual Performance. *Nature Communications*, 7: 13669.
- Craver, C. (2007), *Explaining the Brain* (Oxford University Press).
- Crick, F. & Koch, C. (1990), Towards a Neurobiological Theory of Consciousness. *Seminars in the Neurosciences*, 2: 263–275.
- Crick, F.C. & Koch, C. (2005), What Is the Function of the Claustrum? *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 360(1458): 1271–1279.
- Croxson, P.L. et al. (2005), Quantitative Investigation of Connections of the Prefrontal Cortex in the Human and Macaque Using Probabilistic Diffusion Tractography. *The Journal of Neuroscience*, 25(39): 8854–8866.
- Dehaene, S. et al. (2001), Cerebral Mechanisms of Word Masking and Unconscious Repetition Priming. *Nature Neuroscience*, 4(7): 752–758.
- Dehaene, S. et al. (2006), Conscious, Preconscious, and Subliminal Processing: A Testable Taxonomy. *Trends in Cognitive Sciences*, 10(5): 204–211.
- Dehaene, S. (2014), *Consciousness and the Brain: Deciphering How the Brain Codes Our Thoughts* (Penguin).
- Dehaene, S. et al. (2014), Toward a Computational Theory of Conscious Processing. *Current Opinion in Neurobiology*, 25: 76–84.
- Dehaene, S. & Changeux, J.-P. (2011), Experimental and Theoretical Approaches to Conscious Processing. *Neuron*, 70(2): 200–227.
- Dehaene, S. & Naccache, L. (2001), Towards a Cognitive Neuroscience of Consciousness: Basic Evidence and a Workspace Framework. *Cognition*, 79(1–2): 1–37.
- Del Cul, A. et al. (2009), Causal Role of Prefrontal Cortex in the Threshold for Access to Consciousness. *Brain*, 132: 2531–2540.
- Del Cul, A., Baillet, S. & Dehaene, S. (2007), Brain Dynamics Underlying the Nonlinear Threshold for Access to Consciousness. *Plos Biology*, 5(10): E260.
- Duncan, J. & Owen, A.M. (2000), Common Regions of the Human Frontal Lobe Recruited By Diverse Cognitive Demands. *Trends in Neurosciences*, 23(10): 475–483.
- Ester, E.F., Sprague, T.C. & Serences, J.T. (2015), Parietal and Frontal Cortex Encode Stimulus-Specific Mnemonic Representations During Visual Working Memory. *Neuron*, 87(4): 893–905.
- Fetsch, C.R. et al. (2014), Effects of Cortical Microstimulation on Confidence in a Perceptual Decision. *Neuron*, 83(4): 797–804.
- Feuillet, L., Dufour, H. & Pelletier, J. (2007), Brain of a White-Collar Worker. *The Lancet*, 370(9583): 262.
- Fleming, S.M. et al. (2014), Domain-Specific Impairment in Metacognitive Accuracy Following Anterior Prefrontal Lesions. *Brain*, 137: 2811–2822.
- Fleming, S.M. & Daw, N.D. (2017), Self-Evaluation of Decision-Making: A General Bayesian Framework for Metacognitive Computation. *Psychological Review*, 124(1): 91–114.
- Fleming, S.M. & Frith, C.D. (2014), *The Cognitive Neuroscience of Metacognition* (Springer).
- Flohr, H. (1995), Sensations and Brain Processes. *Behavioural Brain Research*, 71(1–2): 157–161.
- Frässle, S. et al. (2014), Binocular Rivalry: Frontal Activity Relates to Introspection and Action But Not to Perception. *The Journal of Neuroscience*, 34(5): 1738–1747.
- Gaillard, R. et al. (2006), Nonconscious Semantic Processing of Emotional Words Modulates Conscious Access. *Proceedings of the National Academy of Sciences of the United States*

- of America*, 103(19): 7524–7529.
- Gosseries, O. et al. (2014), Measuring Consciousness in Severely Damaged Brains. *Annual Review of Neuroscience*, 37: 457–478.
- Gross, J. et al. (2004), Modulation of Long-Range Neural Synchrony Reflects Temporal Limitations of Visual Attention in Humans. *Proceedings of the National Academy of Sciences of the United States of America*, 101(35): 13050–13055.
- Hardcastle, V.G. (2000), How to Understand the N in NCC. In T. Metzinger (Ed.), *The Neural Correlates of Consciousness* (MIT Press), 259–264.
- Hohwy, J. (2009), The Neural Correlates of Consciousness: New Experimental Approaches Needed? *Consciousness & Cognition*, 18(2): 428–438.
- Kim, J.N. & Shadlen, M.N. (1999), Neural Correlates of a Decision in the Dorsolateral Prefrontal Cortex of the Macaque. *Nature Neuroscience*, 2(2): 176–185.
- Knight, R. & Grabowecy, M. (1995), Escape From Linear Time: Prefrontal Cortex and Conscious Experience. In M. Gazzaniga (Ed.), *The Cognitive Neurosciences* (MIT Press), 1357–1371.
- Koch, C. et al. (2016), Neural Correlates of Consciousness: Progress and Problems. *Nature Reviews. Neuroscience*, 17(5): 307–321.
- Koch, C. (2004), *The Quest for Consciousness: A Neurobiological Approach* (Roberts Publishers).
- Koivisto, M. et al. (2016), Subjective Visual Awareness Emerges Prior to P3. *The European Journal of Neuroscience*, 43(12): 1601–1611.
- Koivisto, M. & Grassini, S. (2016), Neural Processing Around 200 Ms After Stimulus-Onset Correlates with Subjective Visual Awareness. *Neuropsychologia*, 84: 235–243.
- Koizumi, A., Maniscalco, B. & Lau, H. (2015), Does Perceptual Confidence Facilitate Cognitive Control? *Attention, Perception & Psychophysics*, 77(4): 1295–1306.
- Komura, Y. et al. (2013), Responses of Pulvinar Neurons Reflect a Subject's Confidence in Visual Categorization. *Nature Neuroscience*, 16(6): 749–755.
- Kouider, S., Sackur, J. & Gardelle, V. De (2012), Do We Still Need Phenomenal Consciousness? Comment on Block. *Trends in Cognitive Science*, 16(3): 140-1.
- Lak, A. et al. (2014), Orbitofrontal Cortex Is Required for Optimal Waiting Based on Decision Confidence. *Neuron*, 84(1): 190–201.
- Lamme, V.A.F. & Roelfsema, P.R. (2000), The Distinct Modes of Vision Offered By Feedforward and Recurrent Processing. *Trends in Neurosciences*, 23(11): 571–579.
- Lamme, V.A.F. (2010), How Neuroscience Will Change Our View on Consciousness. *Cognitive Neuroscience*, 1(3): 204–220.
- Lamme, V.A.F. (2006), Towards a True Neural Stance on Consciousness. *Trends in Cognitive Sciences*, 10(11): 494–501.
- Lamme, V.A.F., Zipser, K. & Spekreijse, H. (2002), Masking Interrupts Figure-Ground Signals in V1. *Journal of Cognitive Neuroscience*, 14(7): 1044–1053.
- Lamy, D., Salti, M. & Bar-Haim, Y. (2009), Neural Correlates of Subjective Awareness and Unconscious Processing: An ERP Study. *Journal of Cognitive Neuroscience*, 21(7): 1435–1446.
- Landman, R., Spekreijse, H. & Lamme, V.A.F. (2003), Large Capacity Storage of Integrated Objects Before Change Blindness. *Vision Research*, 43(2): 149–164.
- Lau, H. & Rosenthal, D. (2011), Empirical Support for Higher-Order Theories of Conscious Awareness. *Trends in Cognitive Sciences*, 15(8): 365–373.
- Lau, H. (2008), A Higher Order Bayesian Decision Theory of Consciousness. *Progress in Brain Research*, 168: 35–48.
- Lau, H. & Passingham, R.E. (2006), Relative Blindsight in Normal Observers and the Neural Correlate of Visual Consciousness. *Proceedings of the National Academy of Sciences of the United States of America*, 103(49): 18763–18768.

- Leopold, D.A. & Logothetis, N.K. (1996), Activity Changes in Early Visual Cortex Reflect Monkeys' Percepts During Binocular Rivalry. *Nature*, 379(6565): 549–553.
- Luck, S.J., Vogel, E.K. & Shapiro, K.L. (1996), Word Meanings Can Be Accessed But Not Reported During the Attentional Blink. *Nature*, 383(6601): 616–618.
- Lumer, E.D. & Rees, G. (1999), Covariation of Activity in Visual and Prefrontal Cortex Associated with Subjective Visual Perception. *Proceedings of the National Academy of Sciences of the United States of America*, 96(4): 1669–1673.
- Maniscalco, B. & Lau, H. (2016), The Signal Processing Architecture Underlying Subjective Reports of Sensory Awareness. *Neuroscience of Consciousness*, 2016(1).
- Mante, V. et al. (2013), Context-Dependent Computation By Recurrent Dynamics in Prefrontal Cortex. *Nature*, 503(7474): 78–84.
- Melnick, M.D., Tadin, D. & Huxlin, K.R. (2016), Relearning to See in Cortical Blindness. *The Neuroscientist*, 22(2): 199–212.
- Mettler, F.A. (1949), *Selective Partial Ablation of the Frontal Cortex: A Correlative Study of Its Effects on Human Psychotic Subjects* (Hoeber).
- Miller, E.K. (2000), The Prefrontal Cortex and Cognitive Control. *Nature Reviews. Neuroscience*, 1(1): 59–65.
- Miller, E.K. & Cohen, J.D. (2001), An Integrative Theory of Prefrontal Cortex Function. *Annual Review of Neuroscience*, 24: 167–202.
- Milner, D. & Goodale, M. (1995), *The Visual Brain in Action* (Oxford University Press).
- Milner, D. & Goodale, M. (2006), *The Visual Brain in Action* (Oxford University Press), 2nd Ed.
- Miyamoto, K. et al. (2017), Causal Neural Network of Metamemory for Retrospection in Primates. *Science*, 355(6321): 188–193.
- Morales, J., Chiang, J. & Lau, H. (2015), Controlling for Performance Capacity Confounds in Neuroimaging Studies of Conscious Awareness. *Neuroscience of Consciousness*, 2015(1).
- Morales, J., Lau, H. & Fleming, S.M. (Forthcoming), Domain-General and Domain-Specific Patterns of Activity Support Metacognition in Human Prefrontal Cortex. *The Journal of Neuroscience*.
- Neisser, J. (2012), Neural Correlates of Consciousness Reconsidered. *Consciousness and Cognition*, 21(2): 681–690.
- Noe, A. & Thompson, E. (2004), Are There Neural Correlates of Consciousness? *Journal of Consciousness Studies*, 11(1): 3–28.
- Noy, N. et al. (2015), Ignition's Glow: Ultra-Fast Spread of Global Cortical Activity Accompanying Local "Ignitions" in Visual Cortex During Conscious Visual Perception. *Consciousness and Cognition*, 35: 206–224.
- Odegaard, B., Knight, R.T. & Lau, H. (2017), Should a Few Null Findings Falsify Prefrontal Theories of Conscious Perception? *The Journal of Neuroscience*, 37(40): 9593–9602.
- Panagiotaropoulos, T.I. et al. (2012), Neuronal Discharges and Gamma Oscillations Explicitly Reflect Visual Consciousness in The Lateral Prefrontal Cortex. *Neuron*, 74(5): 924–935.
- Passingham, R. (2009), How Good Is the Macaque Monkey Model of the Human Brain? *Current Opinion in Neurobiology*, 19(1): 6–11.
- Passingham, R.E. & Wise, S.P. (2012), *The Neurobiology of the Prefrontal Cortex: Anatomy, Evolution, and the Origin of Insight* (Oxford University Press).
- Pessoa, L. & Adolphs, R. (2010), Emotion Processing and the Amygdala: From a "Low Road" to "Many Roads" of Evaluating Biological Significance. *Nature Reviews. Neuroscience*, 11(11): 773–783.
- Peters, M.A.K. & Lau, H. (2015), Human Observers Have Optimal Introspective Access to Perceptual Processes Even for Visually Masked Stimuli. *Elife*, 4: E09651.
- Petrides, M. & Pandya, D.N. (1984), Projections to the Frontal Cortex From the Posterior Parietal Region in the Rhesus Monkey. *The Journal of Comparative Neurology*, 228(1): 105–116.

- Phillips, I. (2016), Consciousness and Criterion: On Block's Case for Unconscious Seeing. *Philosophy and Phenomenological Research*, 93(2): 419–451.
- Pitts, M.A. et al. (2014), Gamma Band Activity and the P3 Reflect Post-Perceptual Processes, Not Visual Awareness. *Neuroimage*, 101: 337–350.
- Pleskac, T.J. & Busemeyer, J.R. (2010), Two-Stage Dynamic Signal Detection: A Theory of Choice, Decision Time, and Confidence. *Psychological Review*, 117(3): 864–901.
- Rahnev, D. et al. (2011), Attention Induces Conservative Subjective Biases in Visual Perception. *Nature Neuroscience*, 14(12): 1513–1515.
- Railo, H., Koivisto, M. & Revonsuo, A. (2011), Tracking the Processes Behind Conscious Perception: A Review of Event-Related Potential Correlates of Visual Consciousness. *Consciousness and Cognition*, 20(3): 972–983.
- Rigotti, M. et al. (2013), The Importance of Mixed Selectivity in Complex Cognitive Tasks. *Nature*, 497(7451): 585–590.
- Robinson, Z., Maley, C.J. & Piccinini, G. (2015), Is Consciousness a Spandrel? *Journal of the American Philosophical Association*, 1(02): 365–383.
- Romanski, L.M. et al. (1997), Topographic Organization of Medial Pulvinar Connections with the Prefrontal Cortex in the Rhesus Monkey. *The Journal of Comparative Neurology*, 379(3): 313–332.
- Rosenthal, D.M. (2008), Consciousness and Its Function. *Neuropsychologia*, 46(3): 829–840.
- Rosenthal, D.M. (2005), *Consciousness and Mind* (Oxford University Press).
- Rosenthal, D.M. (1993), State Consciousness and Transitive Consciousness. *Consciousness and Cognition*, 2(4): 355–363.
- Rosenthal, D.M. (2004), Varieties of Higher-Order Theory. In R. J. Gennaro (Ed.) *Higher-Order Theories of Consciousness: An Anthology*, 17–44.
- Rounis, E. et al. (2010), Theta-Burst Transcranial Magnetic Stimulation to the Prefrontal Cortex Impairs Metacognitive Visual Awareness. *Cognitive Neuroscience*, 1(3): 165–175.
- Samaha, J. et al. (2016), Dissociating Perceptual Confidence From Discrimination Accuracy Reveals No Influence of Metacognitive Awareness on Working Memory. *Frontiers in Psychology*, 7, P.851.
- Sandrini, M., Umiltà, C. & Rusconi, E. (2011), The Use of Transcranial Magnetic Stimulation in Cognitive Neuroscience: A New Synthesis of Methodological Issues. *Neuroscience and Biobehavioral Reviews*, 35(3): 516–536.
- Schmid, M.C. et al. (2010), Blindsight Depends on the Lateral Geniculate Nucleus. *Nature*, 466(7304): 373–377.
- Sergent, C., Baillet, S. & Dehaene, S. (2005), Timing of the Brain Events Underlying Access to Consciousness During the Attentional Blink. *Nature Neuroscience*, 8(10): 1391–1400.
- Shadlen, M.N. & Kiani, R. (2013), Decision Making as a Window on Cognition. *Neuron*, 80(3): 791–806.
- Shipp, S. (2003), The Functional Logic of Cortico-Pulvinar Connections. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 358(1438): 1605–1624.
- Shoemaker, S. (1981), Some Varieties of Functionalism. *Philosophical Topics*, 12(1): 93–119.
- Sligte, I.G., Scholte, H.S. & Lamme, V.A.F. (2008), Are There Multiple Visual Short-Term Memory Stores? *Plos One*, 3(2): 1699.
- Smith, J.D. (2009), The Study of Animal Metacognition. *Trends in Cognitive Sciences*, 13(9): 389–396.
- Stokes, M.G. (2015), “Activity-Silent” Working Memory in Prefrontal Cortex: A Dynamic Coding Framework. *Trends in Cognitive Sciences*, 19(7): 394–405.
- Tononi, G. (2008), Consciousness as Integrated Information: A Provisional Manifesto. *The Biological Bulletin*, 215(3): 216–242.
- Tse, P.U. et al. (2005), Visibility, Visual Awareness, and Visual Masking of Simple Unattended

- Targets Are Confined to Areas in the Occipital Cortex Beyond Human V1/V2. *Proceedings of the National Academy of Sciences of the United States of America*, 102(47): 17178–17183.
- Tsuchiya, N. et al. (2015), No-Report Paradigms: Extracting the True Neural Correlates of Consciousness. *Trends in Cognitive Sciences*, 19(12): 757–770.
- Tsuchiya, N. & Koch, C. (2005), Continuous Flash Suppression Reduces Negative Afterimages. *Nature Neuroscience*, 8(8): 1096–1101.
- Tye, M. (1996), The Function of Consciousness. *Noûs*, 30(3): 287–305.
- Ungerleider, L.G. & Mishkin, M. (1982), Two Cortical Visual Systems. In D. J. Ingle, M. A. Goodale & R. J. W. Mansfield (Eds.) *Analysis of Visual Behavior* (MIT Press), 548–586.
- Van Opstal, F., De Lange, F.P. & Dehaene, S. (2011), Rapid Parallel Semantic Processing of Numbers Without Awareness. *Cognition*, 120(1): 136–147.
- Vandenbroucke, A.R.E. et al. (2015), Neural Correlates of Visual Short-Term Memory Dissociate Between Fragile and Working Memory Representations. *Journal of Cognitive Neuroscience*, 27(12): 2477–2490.
- Vlassova, A., Donkin, C. & Pearson, J. (2014), Unconscious Information Changes Decision Accuracy But Not Confidence. *Proceedings of the National Academy of Sciences of the United States of America*, 111(45): 16214–16218.
- Wald, A. (1947), *Sequential Analysis* (John Wiley & Sons).
- Wang, M., Arteaga, D. & He, B.J. (2013), Brain Mechanisms for Simple Perception and Bistable Perception. *Proceedings of the National Academy of Sciences of the United States of America*, 110(35): E3350-9.
- Weiskrantz, L. (1997), *Consciousness Lost and Found: A Neuropsychological Exploration* (Oxford University Press).
- Wu, W. (2014), Against Division: Consciousness, Information and the Visual Streams. *Mind & Language*, 29(4): 383–406.