Matthias Michel

Methodological Artefacts in Consciousness Science

The Case Against the Global Workspace Theory

Abstract: *Consciousness is scientifically challenging to study because of its subjective aspect. This leads researchers to rely on report-based experimental paradigms in order to discover neural correlates of consciousness (NCCs). I argue that the reliance on reports has biased the search for NCCs, thus creating what I call 'methodological artefacts'. This paper has three main goals: first, describe the measurement problem in consciousness science and argue that this problem led to the emergence of methodological artefacts. Second, provide a critical assessment of the NCCs put forward by the global neuronal workspace theory. Third, provide the means of dissociating genuine NCCs from methodological artefacts.*

1. Introduction

Consciousness is subjective: only the subject having a conscious experience has direct access to it. Contrary to a third-person observer, a subject having a conscious experience does not have to infer from her behaviour that she has that experience. Subjects have a first-person privileged access to their own conscious experiences: introspection is the cognitive capacity that enables this subjective access (Shoemaker, 1996; Bar-On, 2004). Introspection is a special kind of metacognition, defined broadly as the activity of thinking about our

Correspondence:
Matthias Michel, Sciences, Normes et Décisions, Université Paris-Sorbonne, 1 Rue Victor Cousin, 75005, Paris, France.
Email: matthias.Michel.2@paris-sorbonne.fr

Journal of Consciousness Studies, **24**, No. 11–12, 2017, pp. ?–?

own thoughts. I define introspection as the capacity to select specific information among consciously accessible contents, allowing for subsequent report of this information. Because introspection is subjective, a third-person observer can only *infer* that a subject has a conscious experience from her report or behaviour.

The subjective aspect of consciousness is what makes its scientific study so challenging. At least since Crick and Koch's seminal article (1990), which set out the agenda for current scientific research, one of the main goals of the scientific study of consciousness is to uncover the neural basis of consciousness, also called 'neural correlates of consciousness' (NCCs). Within this framework, research in consciousness science, such as the influential research programme of the 'global neuronal workspace theory' (Dehaene and Naccache, 2001; Dehaene and Changeux, 2011; Dehaene, 2014), relied on report-based methods in order to infer, from the subjects' reports, that they have conscious experiences caused by stimuli displayed by the experimenters (which is a necessary first step toward obtaining NCCs). However, mounting evidence suggests that NCCs discovered by consciousness science in the past twenty years could be artefacts due to the use of report-based methods, rather than proper NCCs (Aru *et al.*, 2012; De Graaf, Hsieh and Sack, 2012; Tsuchiya *et al.*, 2015).

In this paper, I argue that the requirement of reports has biased the search for proper NCCs, thus creating what I call 'methodological artefacts'. By comparing results from no-report paradigms, studies on the neural basis of introspection, and report-based paradigms, I argue that the main neural correlates of consciousness discovered by proponents of the global neuronal workspace theory of consciousness might be methodological artefacts. I then argue that, in order to study consciousness, one must begin with the study of the neural states that underlie introspection. Only then can we dissociate methodological artefacts from NCCs.

2. NCCs and Methodological Artefacts

In this section, I clarify the notion of NCCs and point to different ways in which experimenters could be misled when trying to uncover NCCs. I show that one of these problems could lead to the misidentification of neural states created by the act of introspection as genuine NCCs, I call this particular kind of misidentified NCC 'methodological artefacts'.

NCCs of a conscious experience C are defined as the minimal set of neural states N that are jointly sufficient for C, given appropriate enabling conditions (Chalmers, 2000; Koch, 2004). If a neural state N is an NCC, then N must be *sufficient* for a particular conscious experience C, because the activation of the relevant neural state must lead to C by itself.[1] If N is an NCC for a conscious experience C, no brain activity over and above N is required for C to occur. As Hohwy and Bayne (2015) point out, some other neural state N′ may be required for a creature to be in a neural state N. In such case, N′ is not an NCC, but a prerequisite for consciousness. Indeed, NCCs must be *minimally* sufficient for consciousness, because one wants to isolate *only* the neural features that are involved in a particular conscious state, and not neural prerequisites or neural consequences of consciousness (Aru *et al.*, 2012; 2015; De Graaf, Hsieh and Sack, 2012). A *prerequisite* of consciousness is a neural state that occurs upstream from the NCC, and may be necessary for a conscious experience, but is not a genuine NCC. For example, although activity on my retina is necessary for my having a conscious visual experience of a rose, it is not because of that activity that the visual percept of a rose is conscious rather than unconscious. Hence, activity on my retina does not qualify as an NCC. Rather, it is a prerequisite for my having the conscious experience of a rose. A *consequence* of consciousness is a neural state that results from a conscious experience and may often co-occur with this experience without being an NCC. For example, having a conscious visual experience of a rose could make me think about the smell of a rose, and therefore lead to the activation of a neural state that correlates with thinking about the smell of a rose, without these neural states having anything to do with the percept of a rose being conscious. One should therefore not conflate NCCs with prerequisites and consequences of consciousness.

In trying to establish the NCCs, consciousness science typically relies on ‘contrastive analysis’ (Baars, 1988; Aru *et al.*, 2012). Contrastive analysis consists in comparing behavioural characteristics or

[1] It is still debated whether the search for NCCs is a search for the minimally sufficient neural causes of consciousness, neural constituents of consciousness, or neural correlates of consciousness (Neisser, 2012; Miller, 2014). I will not enter into this debate in this paper: while it is sometimes difficult to avoid a ‘causal’ language on this matter, I will assume that the C in NCCs stands for ‘correlates’, as the use of this neutral language frees us from the debate on the mind/body relations and is widely used both in the philosophical and scientific literature on consciousness.

neural activity on trials in which a subject either consciously or unconsciously perceives a stimulus while holding the stimulus constant (for example, thanks to backward masking or binocular rivalry; see Kim and Blake, 2005). In a backward masking experiment, for example, a target stimulus becomes invisible for a subject because a second stimulus (i.e. the mask) is presented close in time and space to the target. If the stimulus is flashed around 50 ms before the mask, subjects tend to consciously perceive the target on some trials and not on others (in which it is only subliminally perceived), allowing experimenters to manipulate the subject's conscious perception while holding a stimulus constant. Experimenters then compare neural activity of a subject on trials in which she reported consciously perceiving the stimulus with neural activity on trials in which she reported no experience. Experimenters then infer neural states that correlate with a conscious experience from the difference in neural activity between conscious and unconscious trials. As noted by Overgaard (2006), in report-based paradigms, one obtains NCCs by matching two measures: first, an objective measure of the neural activity of the subject by a measuring apparatus (fMRI, EEG, MEG, etc.); second, a subjective detection of her own state of consciousness by the subject through introspection and subsequent report (verbal report, button pushes, etc.) (see Figure 1). From the measure of the neural activity, experimenters infer that a particular neural state is realized in the brain, and from the report of the subject, experimenters infer that she has or does not have a conscious experience of the stimulus. Overgaard thus concludes that:

> to derive the desired correlation, the 'NCC', from the actual data, one is fully dependent on the nature of the relation between the brain state and the measure hereof, on the one hand, and the relation between the conscious state and the report on the other. (*ibid.*)

Indeed, on this model, one obtains NCCs only if the following conditions are satisfied:

On the objective side of the process: (1) the measuring apparatus accurately measures neural activity, (2) experimenters correctly infer the actual brain state from the measured neural activity, and (3) they isolate neural states that correlate with the conscious experience by screening off both neural prerequisites and neural consequences of consciousness.

On the subjective side of the process: (4) the subject accurately detects the presence or absence of a conscious experience by

introspection; (5) she adequately reports what she introspects; (6) experimenters interpret the subject's report such that they correctly infer her conscious state.

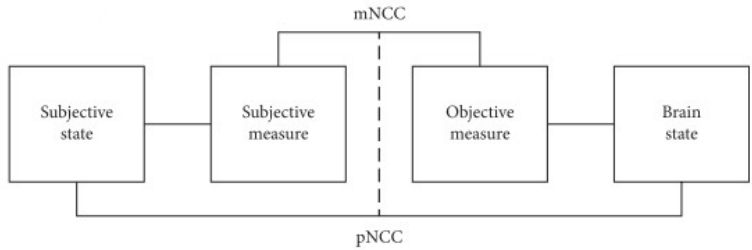

Figure 1. Source: Overgaard (2015). In order to obtain proper NCCs (pNCC), one has to obtain a correlation between a brain state and a conscious experience. But experimenters cannot directly access brain states nor conscious experiences. Hence, they correlate two measures instead: a subjective detection, the subject's report of having (or not having) a conscious experience, and an objective measure, obtained by a measuring apparatus, of the neural activity. The result is not a pNCC, but only a *measured* NCC (mNCC), from which experimenters then try to infer pNCC.

I already explained how the objective measure can go wrong: the neural state inferred from the measured neural activity may be a prerequisite or a consequence of consciousness, not a real NCC, and it is difficult to disentangle NCCs from these other factors. But, for consciousness science, the real trouble stems from the use of introspection and the nature of the relation between conscious states and subjective reports. I see three main ways in which the subjective measure can bias the search for NCCs: (1) introspection may fail to detect occurrent conscious episodes, or in other words, introspection might not be *exhaustive* (Overgaard and Sandberg, 2012; Timmermans and Cleeremans, 2015); (2) introspection may not be *exclusive*: it may misclassify information that is not conscious as conscious (Timmermans *et al.*, 2010); (3) neural prerequisites, neural correlates, and neural consequences of introspection may be conflated with NCCs. Neural correlates of introspection (NCIs), just as in the case of NCCs, are the set of neural states that are jointly minimally sufficient for introspection of a mental state. A neural prerequisite of introspection is a neural state that is necessary for introspection to detect a conscious mental state, but is not part of the neural states that are sufficient for this introspection to occur. A neural consequence of introspection is a neural state that may co-occur with introspection

without being necessary nor sufficient for it (such as neural states that underpin verbal reports or button pressings during an experiment for example).

The first kind of bias is the subject of a long lasting debate about whether or not phenomenal consciousness (what the experience is like for the subject) overflows cognitive access (what is available for a subject to report). Some claim that the scope of access is identical to the scope of consciousness, and hence that introspection is exhaustive, because all conscious contents would thus be available for introspection and subsequent report (Cohen and Dennett, 2011; Cohen, Dennett and Kanwisher, 2016; Dehaene, 2014), while proponents of phenomenal overflow claim that it is not (Block, 1995; 2007; Bronfman *et al.*, 2014). I won't take a stand on this debate in this paper.[2] The second kind of bias has been extensively studied by cognitive scientists and philosophers inspired by signal detection theory: experimenters try to uncover unconscious influences on metacognition, which might bias the subjective measure provided by the subjects' reports (Irvine, 2012; Newell and Shanks 2014). Here, I will be interested in the third kind of bias due to the use of introspection in consciousness science; namely, the mistake of conflating neural prerequisites, correlates, or consequences of introspection and NCCs. I use the term 'methodological artefacts' for NCCs that result from this latter kind of mistake.

> **Methodological artefact:** neural state that is a prerequisite, a correlate, or a consequence of introspection, but is misidentified by experimenters as a genuine NCC.

To illustrate this notion, let's imagine that activation of the prefrontal cortex is necessary for introspection of a conscious experience, without being part of the states that are jointly minimally sufficient for this experience (NCC). In such a case, if experimenters rely only on report-based paradigms, activation of the prefrontal cortex will always co-occur with the presence of a conscious experience. Every time the experimenter will require the subject to introspect her conscious experience in order to report it, the prefrontal cortex will be activated.

---

2 Nonetheless, I want to put forward an important distinction between access consciousness and introspection. While all access-conscious contents are *available* to be targets of introspection, not all access-conscious contents are introspected. As Block writes '[access]-consciousness… requires no introspection' (Block. 1995, p. 280).

This co-occurence could lead experimenters to mistake the activation of the prefrontal cortex for a genuine NCC. Activation of the prefrontal cortex would thus be a methodological artefact. The challenge of the measurement problem in consciousness science is to disentangle genuine NCCs from methodological artefacts.

Before I explain in more detail the challenge of the measurement problem, there is a point that needs some clarification. It is often presupposed in the literature that the mistake of conflating methodological artefacts and NCCs is identical to the mistake of conflating consequences of consciousness and NCCs (Aru *et al.*, 2012; 2015; De Graaf, Hsieh and Sack, 2012; Tsuchiya *et al.*, 2015). Although Aru *et al.* (2012) or Bachmann (2009) take into account prerequisites of consciousness as potential confounds in the search for NCCs, one should further distinguish between the prerequisites of consciousness and the prerequisites of introspection, as they could be different and could therefore constitute independent sources of errors. It is possible that asking subjects to report their experience modifies the processing of information downstream but also *upstream* from consciousness. In other words, asking the subjects for reports may modify both neural states that precede consciousness and neural states that follow it. Indeed, neural prerequisites of introspection may well be active before the conscious experience itself. A study by Overgaard (2006) shows that asking subjects for reports modifies early perceptual processes rather than only post-perceptual processes. Some neural prerequisites of introspection such as the neural states underlying attention may modify early perceptual processes independently of consciousness (Kentridge, Nijboer and Heywood, 2008; Norman, Heywood and Kentridge, 2013). Hence, if the requirement for reports can modify neural states upstream from consciousness, neural prerequisites and correlates of introspection should not be considered only as neural consequences of consciousness. The mistake of conflating NCCs and methodological artefacts is thus different from the mistake of conflating NCCs and consequences of consciousness.

Now that the two kinds of mistakes are disentangled, I will focus only on the mistake of conflating NCCs and methodological artefacts. The challenge raised by the potential presence of methodological artefacts in consciousness science stems from the fact that there is no *a priori* link between neural states that are necessary or sufficient for introspection and neural states that are minimally sufficient for a conscious experience (NCCs). Of particular difficulty is the task of disentangling genuine NCCs from necessary prerequisites of

introspection and NCIs. Indeed, if a neural state N is necessary for introspection of a conscious state C, then N is necessary for the subjective detection of C, and hence N will likely appear as an NCC. To see how the challenge generalizes to every NCC so far discovered through report-based paradigms, let me put the argument in a more formal way:

The methodological artefact argument:

(1) If introspection is necessary to detect the presence or absence of conscious experience C, and if a neural state N is necessary for introspection, then N is necessary to detect C.
(2) Hence, for every detection of C, there is a neural state N, such that N is necessary for detecting C.
(3) But N being necessary for *detecting* C does not imply that N is necessary nor sufficient for C.
(4) Hence, every N appearing to be necessary or sufficient for C may not be necessary nor sufficient for C (but may only be necessary for detecting C). Consequently, every N appearing to be necessary or sufficient for C on the basis of a report-based paradigm could be a methodological artefact.

Now, the argument only shows that NCCs so far discovered thanks to report-based methods *could be* methodological artefacts, it does not show that they are. Moreover, it is not sufficient to show that a neural state identified as an NCC is a neural prerequisite or a neural correlate of introspection in order to show that it is a methodological artefact. For it could very well be that some prerequisites or NCIs *are also* genuine NCCs. There could be some overlap between neural correlates and prerequisites of introspection and NCCs. Going back to the example of the activation of the prefrontal cortex: it could be that activation of the prefrontal cortex is a neural prerequisite or correlate of introspection while still being a genuine NCC. Hence, what is needed to demonstrate the existence of a methodological artefact is evidence that a neural state identified as an NCC is a neural prerequisite, a neural correlate, or a neural consequence of introspection, *and* evidence that the NCC has been misidentified as such from an independent source of knowledge that gathers data without relying on reports (i.e. no-report paradigms).[3]

[3] I want to stress that my line of argument should be distinguished from the traditional no-report argument, according to which neuroscientists confound the neural states

Despite the difficulty of showing that a given identified NCC really *is* a methodological artefact, it is still possible to make a case against the identification of certain neural states as NCCs by comparing data from report-based paradigms, no-report paradigms, and studies on the neural basis of introspection. The general argument that I put forward in the next section takes the form of an inference to the best explanation: if (1) N is identified as an NCC by report-based methods, (2) N is *not* identified as an NCC by no-report paradigms, (3) N is known to be a neural prerequisite, correlate, or consequence of introspection; then, the best explanation of the discrepancy between (1) and (2) is that N is a methodological artefact. Of course, the more (2) and (3) are firmly grounded, the more the argument is convincing. Although no-report paradigms and the study of the neural basis of introspection are still at an emerging stage of their development (Fleming and Dolan, 2012; Tsuchiya *et al.*, 2015), I argue in the next section that there is already convincing evidence that the main NCCs uncovered by the influential global neuronal workspace theory of consciousness are methodological artefacts.

Before applying the methodological argument to the global neuronal workspace theory, I want to address a potential concern. One might argue that the methodological artefact argument tacitly relies on a first-order view of consciousness: if it turns out that introspection and consciousness are the same, as might be supposed by proponents of higher-order thought theories of consciousness, then alleged methodological artefacts will also turn out to be genuine NCCs.

Higher-order thought theorists argue that consciousness of a first-order state arises from a second-order thought, representation, or perception of this first-order state (Armstrong, 1968; Lau, 2008; Lycan, 1996; Rosenthal, 2005). Now, could this second-order thought, representation, or perception be identical with introspection itself? It seems that most higher-order thought theorists would resist this claim. The leading higher-order thought theorist, David Rosenthal, explicitly

---

associated with *reports* and the neural states associated with consciousness itself. Although this is a genuine problem, my argument is more encompassing. On my view, a report is a *consequence* of introspection. One could still confound the NCCs and the NCIs without relying on reports, for it is perfectly possible that subjects are still introspecting and monitoring their experiences during experiments using no-report paradigms (Overgaard and Fazekas, 2016). Hence, although they can help disentangle NCCs from NCIs, no-report paradigms fail to do so by themselves.

distinguishes between a second-order thought rendering a first-order state conscious, and introspection:

> It is important to distinguish a mental state's being conscious from our being introspectively aware of that state… introspection is a more complex phenomenon than the ordinary consciousness of mental states. Intuitively, a mental state's being conscious means just that it occurs in our stream of consciousness. Introspection, by contrast, involves consciously and deliberately paying attention to our contemporaneous mental states. (Rosenthal, 2005, pp. 27–8)

If the relevant higher-order states are different from introspection, the methodological artefact argument still applies. Indeed, the methodological argument only requires consciousness to be different from the capacity to select an accessible content by top-down or voluntary attention in order to report being conscious of it (i.e. introspection).[4] Without entering further into this debate, I now apply the methodological artefact argument to the global neuronal workspace theory.

3. NCCs of the GNW Theory: The P3b and Prefrontal Cortex Activations as Methodological Artefacts

One of the most important theories of consciousness of the last twenty years, initially developed by Baars (1988), is the global neuronal workspace (GNW) theory of consciousness (de Gardelle and Kouider, 2009). The main claim of the GNW theory is that consciousness results from the global broadcast of information to many distant areas of the brain through a global neuronal workspace (Dehaene and Naccache, 2001; Dehaene and Changeux, 2011; Dehaene, 2014). Accordingly, consciousness is identical to 'the selection, amplification, and global broadcasting, to many different areas, of a single piece of information selected for its salience or relevance to current goals' (Dehaene and Changeux, 2011). On the GNW theory, a content is consciously experienced if and only if it is selected by attention to be in the GNW for subsequent global broadcast. The GNW is thought to be implemented by a prefronto-parietal network (Dehaene and

[4] Moreover, suggesting that consciousness is identical with introspection would commit one to the claim that voluntary attention is necessary for consciousness, which seems inconsistent with the growing consensus in cognitive science that it is not (e.g. Aru and Bachmann, 2013; Bronfman *et al.*, 2014; Lamme, 2004; Li *et al.*, 2002; Tallon-Baudry, 2012; Van Boxtel, Tsuchiya and Koch, 2010; but see Cohen *et al.*, 2012).

Naccache, 2001): a conscious content is encoded by the sustained activity of a fraction of the GNW neurons in the prefronto-parietal network, the long-distance axons of pyramidal cells in these areas allowing for the broadcast of the selected information throughout the cortex. The GNW theory predicts that consciousness correlates with an all-or-none and late (around 300 ms after stimulus onset) activation of the prefronto-parietal network, corresponding to the global broadcast of information. This global broadcast responsible for consciousness also correlates with an event-related potential (the electrophysiological activity in the brain in response to a particular event) appearing only when information becomes conscious, around 300 ms after stimulus onset, called the P3b wave (or just P3b for short). Hence, according to the GNW theory, both the late and all-or-none activation of the prefronto-parietal network and the P3b are NCCs.

At first, there seems to be extensive evidence in favour of a late and all-or-none activation of the prefronto-parietal network being a reliable NCC. Indeed, using contrastive analysis, experimenters observe an all-or-none, late (from 300 ms), and sustained firing in the fronto-parietal network only when subjects are conscious of a stimulus across different experimental paradigms such as stimulus masking (Dehaene and Naccache, 2001; Gaillard *et al.*, 2009; Fisch *et al.*, 2009; Del Cul, Baillet and Dehaene, 2007; Del Cul *et al.*, 2009), attentional blink (Sergent and Dehaene, 2004; Sergent, Baillet and Dehaene, 2005; Williams *et al.*, 2008), binocular rivalry (Sterzer, Kleinschmidt and Rees, 2009), or conscious perception of errors (van Gaal *et al.*, 2011; Charles *et al.*, 2013). Furthermore, these results have been replicated across different modalities, using conscious and subliminal tactile stimuli (Boly *et al.*, 2007) or conscious and subliminal sounds (Sadaghiani, Hesselmann and Kleinschmidt, 2009). The hypothesis that the P3b wave is an NCC seems also to be supported by a wide array of results across paradigms using visual masking, the attentional blink, or conscious perception of errors (Sergent, Baillet and Dehaene, 2005; Del Cul, Baillet and Dehaene, 2007; van Aalderen-Smeets, Oosterweld and Schwarzbach, 2009; Bekinschtein *et al*., 2009; Gaillard *et al.*, 2009; El Karoui *et al.*, 2015). For example, when two target stimuli are presented in close temporal succession, the second target is often invisible to the subject, a phenomenon called ‘attentional blink’. Sergent, Baillet and Dehaene (2005) used the attentional blink to demonstrate that the P3b is observed only in trials in which subjects report perceiving the second

target, and thus conclude that the P3b is an NCC. So far, so good for the GNW theory.

However, all these experiments used both subjective reports of conscious perception and the contrastive analysis method. Hence, these results are all prone to the measurement problem argument. Since the only way to operate the subjective measure is by asking subjects to introspect in order for them to report the presence of a conscious experience, both the activation of the prefronto-parietal network and the P3b wave could result from the use of introspection. In other words, both the activation of the prefronto-parietal network and the P3b could be methodological artefacts (neural prerequisites, correlates, or consequences of introspection) rather than genuine NCCs. I now argue, against the GNW theory, that the activation of the prefrontal cortex and the P3b are methodological artefacts.

I will focus first on the claim that activation of the prefrontal cortex is an NCC. In what follows, I both argue that no-report paradigms indicate that activation of the prefrontal cortex is not an NCC, and that this activation is best explained by the use of introspection in report-based paradigms.

The prefrontal cortex is now well-known for its central role in introspection. Results from Fleming, Huijgen and Dolan (2012) showed both that the activity of the lateral prefrontal cortex is systematically linked to metacognitive accuracy (a subject's capacity to correctly assess her performance on a sensory discrimination task), and that functional connectivity between the prefrontal cortex and visual cortices increases during metacognitive reports. Confirming these results, both Fleming and Dolan (2012) and Baird *et al.* (2013) found that metacognition of perceptual information and memory are subserved respectively by the lateral anterior prefrontal cortex and the medial anterior prefrontal cortex (see also Valk *et al.*, 2016). Interestingly, individual differences in grey matter volume in the anterior prefrontal cortex correlate with differences in metacognitive accuracy (Fleming *et al.*, 2010; McCurdy *et al.*, 2013). Furthermore, a study of seven patients with lesions of the prefrontal cortex showed a specific impairment in metacognitive accuracy on a perceptual discrimination task, despite their objective performance being equivalent to healthy controls (Fleming *et al.*, 2014). Similarly, using theta-burst transcranial magnetic stimulation to depress activity in the lateral prefrontal cortex, Rounis *et al.* (2010) specifically impaired metacognitive accuracy while subjects still had the same perceptual discrimination performance as control subjects. Hence, lesion studies

and studies on the neural underpinnings of introspection suggest that activity of the prefrontal cortex is a neural correlate of introspection. However, lesion studies also seem to indicate that the prefrontal cortex is not required for having conscious experiences, as complete bilateral frontal lobectomy or large bilateral resection do not seem to impair consciousness while affecting executive functions and working memory capacities (Müller and Knight, 2006). In a case study by Markowitsch and Kessler (2000), a young woman showed preserved perceptual abilities and seemed to be perfectly conscious despite extensive bilateral damage to her prefrontal cortex. Hence, it seems that activity of the prefrontal cortex specifically accounts for introspection of conscious experiences, and not for these experiences being conscious in the first place.

Additional evidence comes from the study of sleep and dreams. Activation of the prefrontal cortex does not increase during REM sleep (the sleep stage in which subjects dream vividly) (Nir and Tononi, 2010), suggesting that having a conscious dream experience does not correlate with activations of the prefrontal cortex. Consistent with the hypothesis that the activation of the prefrontal cortex could be a methodological artefact, the prefrontal cortex is activated during lucid dreaming, when subjects are aware of the fact that they are dreaming (Dresler *et al.*, 2012). Filevich *et al.* (2015) recently found shared neural mechanisms between lucid dreaming and metacognition in the prefrontal cortex, concluding that prefrontal areas could be responsible for one's awareness that one is dreaming without being involved in generating the experience of dreaming itself.

A study by Goldberg, Harel and Malach (2006) is also consistent with these results. In this experiment, subjects in an fMRI scanner had to categorize pictures under animal/no-animal categories. The experiment had three conditions: first, a slow categorization condition; second, an introspection condition, in which subjects had to introspect about their emotional responses when presented with the images; third, a fast categorization condition, in which the stimulation rate was three times faster than in the slow categorization condition. While in the introspection condition the prefrontal cortex showed increased activity, it was deactivated below the slow condition during the fast categorization task. The authors conclude that self-related activity in the prefrontal cortex is inhibited during highly demanding sensory tasks, thus revealing potential neural underpinnings of the common phenomenology of 'losing oneself in the act'. This finding is at odds with the predictions of the GNW theory, as the theory predicts that

neural assemblies of the GNW in the prefrontal cortex should display *increased* activity in a sensorily demanding task, and not decreased activity. On the contrary, it confirms that prefrontal areas could be responsible for introspection, rather than for the conscious character of experience.

The prefrontal involvement in binocular rivalry has also recently been challenged in a groundbreaking study by Frässle *et al.* (2014). During binocular rivalry, an image in one eye becomes unconscious because of its competition with a rival and incompatible image presented in the other eye. The result of binocular rivalry is that participants report experiencing temporal alternations between the image presented in one eye and the other. When one image is unconscious, the other becomes conscious, it is then possible to study the transition of a particular content from unconsciousness to consciousness. Crucially, subjective reports (generally by asking the subject to press a button when they become aware of a stimulus) have been considered as the only way for the experimenters to know whether the image in the right or left eye is being consciously experienced by the subject. As noted above, the prefrontal cortex has long been thought to be involved in the switch in the content experienced by the subjects (Sterzer, Kleinschmidt and Rees, 2009), experiments using binocular rivalry are then generally thought as supporting the GNW theory. Frässle *et al.* (2014) hypothesized that activation of the prefrontal cortex during perceptual switch in experiments using binocular rivalry is not responsible for the change in the content of experience but, rather, correlates with introspection. In order to test this hypothesis, they developed an ingenious no-report experimental set-up. They used results from previous experiments showing that a switch in the content experienced by the subject during binocular rivalry can be inferred thanks to reflexes such as pupil dilation and ocular micro-saccades (Einhäuser *et al*., 2008; Naber, Frässle and Einhäuser, 2011). It then became possible for experimenters to infer which image was consciously experienced by the subject from the observation of subtle changes in pupil dilation and ocular micro-saccades. This method was used by Frässle *et al.* in order to determine the NCCs independently of the NCIs by contrasting two conditions: an introspection condition in which subjects had to report the switch from one image to another, and a no-report condition in which they did not.

Crucially, their result is that the prefrontal cortex is activated only in the introspection condition and not in the no-report condition. Frässle and co-workers conclude that ‘frontal areas are associated with active

report and introspection rather than with rivalry per se' (Frässle *et al.*, 2014, p. 1738). Nonetheless, it should be emphasized, first, that some prefrontal areas such as the right superior frontal gyrus or the right inferior frontal gyrus remained active even during the no-report condition (Zaretskaya and Narinyan, 2014). Second, and more importantly, the study by Frässle *et al*. (2014) addresses the neural correlates of the switch in the content of consciousness during binocular rivalry rather than neural correlates of the content of consciousness itself (Naber and Brascamp, 2015). Despite these mitigating factors, and combined with the results by Goldberg, Harel and Malach (2006), it is still reasonable to consider that this study casts doubt on the idea that activation of the prefrontal cortex is a genuine NCC, and supports the hypothesis that it could rather be a methodological artefact.

Now that a case has been made for considering the activation of the prefrontal cortex as a methodological artefact, let me turn to the P3b wave. In attentional blink and visual masking experiments, the P3b wave seems to be a reliable NCC: this event-related potential (ERP) is observed only on trials in which the subject reports being aware of the target stimulus (Sergent, Baillet and Dehaene, 2005; Del Cul, Baillet and Dehaene, 2007). Nonetheless, it should be noted that the P3b is not the only ERP that correlates with visual awareness, other ERPs have been proposed, such as the P1 or the visual awareness negativity (VAN) (for a review, Railo, Koivisto and Revonsuo, 2011; Rutiku, Aru and Bachmann, 2016). I now argue that the P3b is not an NCC but a methodological artefact. On my view, the P3b correlates with a prerequisite of introspection; namely, with working memory encoding, and not with consciousness.

Before we continue, I have to describe briefly what cognitive scientists call 'working memory'. In a nutshell, working memory is a system that underpins many abilities such as reasoning, learning, and comprehension. It enables us to keep particular pieces of information in mind while manipulating them (Baddeley, 2007). Consequently, most of the sustained and high-level processing that is going on in our brains is due to representations being encoded and sustained in working memory.

The neural correlates of working memory do not qualify as proper NCCs: an NCC is supposed to be minimally sufficient for consciousness, but there are cases in which encoding in working memory does not seem to be sufficient for consciousness. Indeed, results indicate that items can be unconsciously encoded and sustained briefly in working memory (Bergström and Eriksson, 2014; Dutta *et al.*, 2014;

Soto and Silvanto, 2014; 2016; Pincham, Bowman and Szucz, 2016), an hypothesis that is now also supported by proponents of the GNW theory themselves (King, Pescetelli and Dehaene, 2016). Now, the crucial point is that encoding information in working memory (or update of the content of working memory) reliably correlates with the P3b (Polich, 2007).

Against the GNW theory, recent evidence suggests that the P3b is neither necessary nor sufficient for consciousness. Complex unconscious processing of a stimulus can evoke the P3b (Silverstein *et al.*, 2015; 2016). Hence, the P3b does not seem to be sufficient for consciousness. Moreover, the P3b does not correlate with conscious perception when subjects consciously see the stimulus but a representation of the target stimulus is already encoded in working memory (Melloni *et al.*, 2011), suggesting that the P3b does not systematically correlate with consciousness but better correlates with working memory encoding. This hypothesis is further supported by the finding that task-irrelevant stimuli (i.e. stimuli that need not be reported or encoded in working memory) do not trigger the P3b wave (Pitts, Martínez and Hillyard, 2012; Shafto and Pitts, 2015) although subjects are aware of them (Pitts, Metzler and Hillyard, 2014; Pitts *et al.*, 2014). The P3b could then correlate with working memory update rather than conscious awareness.

I now argue that working memory update is a prerequisite of introspection. Although the study of the electrophysiology of introspection is still at an early stage of its development, evidence suggests that introspection correlates with the P3 wave (Overgaard, 2006; Desender *et al.*, 2016). It is plausible that it is necessary to encode a representation in working memory for subsequent introspection and report, which would explain the correlation between introspection and the P3 wave. Indeed, if working memory update elicits the P3b, and if working memory encoding is necessary for introspection, then introspection should aways co-occur with the P3b. Although this claim is still speculative, a study by Maniscalco and Lau (2015) addressing the link between working memory and introspection could bring some support for it. In this study, Maniscalco and Lau analysed the effects of both working memory load and manipulation demands on metacognitive accuracy. The results indicate that metacognitive performance was selectively impaired under high working memory manipulation demands. The experimenters conclude that:

> the same processes involved in manipulating and reorganizing working memory contents might also be involved in manipulating and

> reorganizing sensory representations for the purposes of metacognitive evaluation. (*ibid.*, p. 11)

Hence, introspection could draw on the resources of working memory, with working memory encoding thus being a prerequisite of introspection. The P3b seems to correlate with task-relevance and working memory encoding, which may be a prerequisite of introspection, and not with conscious perception itself. Thus, it seems reasonable to conclude that the P3b is more likely to be a methodological artefact than a genuine NCC.

In this section I argued that both activation of the prefrontal cortex and the P3b, considered as NCCs by the GNW theory, are methodological artefacts. On the one hand, activation of the prefrontal cortex is likely to be a neural correlate of introspection rather than a genuine NCC. On the other hand, the P3b wave does not correlate with conscious perception but with working memory encoding, which may be a prerequisite of introspection.

The methodological artefact argument is based on an inference to the best explanation, which, in the case of the GNW, depends on the results reviewed above. These results should be treated with caution: fMRI and EEG recordings are coarse measures of neural activity, and it could be that further enquiry with more sensitive measures would be able to demonstrate frontal activity independent of introspection (although it is difficult to assess whether a subject is introspecting or not without a report) (e.g. Cortese *et al*., 2016). Nonetheless, independently of these experimental worries, I hope that the arguments I provided can serve as a way to illustrate the importance of the measurement problem in consciousness science. I now conclude with some considerations on different possibilities to disentangle NCCs from methodological artefacts.

4. Conclusion: How to Disentangle NCCs from Methodological Artefacts

How can we disentangle NCCs from methodological artefacts? Report-based paradigms alone won't do it, because they are the very source of methodological artefacts in consciousness science. But no-report paradigms alone won't do it neither. Indeed, as noted by Overgaard and Fazekas (2016), nothing rules out that in no-report paradigms subjects are still introspecting or reflecting on their own conscious experiences. Rather, I suggest that a good start would be to study the neural states that underly introspection. Most of our worries

would be dramatically reduced if we had a way of knowing how exactly introspection is modifying conscious experience and the neural states by which introspection is realized. Studying introspection will be necessary for further scientific enquiry on the neural basis of consciousness. One should also combine the study of the neural basis of introspection with a systematic comparison of the results from report-based and no-report paradigms in order to assess their consistency (Tsuchiya *et al.*, 2016). Using results from different sources of knowledge will undoubtedly bring progress in consciousness science (Block *et al.*, 2014; Koch *et al*., 2016). If evidence from report-based and no-report paradigms is inconsistent, it is possible, from our knowledge of the neural prerequisites, correlates, and consequences of introspection, to argue that report-based paradigms uncover methodological artefacts rather than NCCs. This is the strategy I tried to apply here to the NCCs put forward by the GNW theory. Crucially, this methodology could help solve only one of the many problems we face in the quest for NCCs; namely, the problem of confounding NCCs with methodological artefacts. As I argued in the first section, there are many other ways in which the search for NCCs can go wrong, and a great deal of additional work is needed to unravel genuine NCCs from confounding factors.

Acknowledgments

My thanks to Anouk Barberousse, Pascal Ludwig, Émile Thalabard, and two anonymous reviewers for their helpful comments. I also thank Michael Lundie and Josselin Bonnamour for reading and commenting previous versions of this work.

References

Armstrong, D.M. (1968) *A Materialist Theory of the Mind*, London: Routledge.

Aru, J., Bachmann, T., Singer, W. & Melloni, L. (2012) Distilling the neural correlates of consciousness, *Neuroscience and Biobehavioral Reviews*, **36** (2), pp. 737–746.

Aru, J. & Bachmann, T. (2013) Phenomenal awareness can emerge without attention, *Frontiers in Human Neuroscience*, **7**, art. 891.

Aru, J., Bachmann, T., Singer, W. & Melloni, L. (2015) On why the unconscious prerequisites and consequences of consciousness might derail us from unraveling the neural correlates of consciousness, in Miller, S.M. (ed.) *The Constitution of Phenomenal Consciousness: Toward a Science and Theory*, Amsterdam: John Benjamins.

Baars, B.J. (1988) *A Cognitive Theory of Consciousness*, Cambridge: Cambridge University Press.

Bachmann, T. (2009) Finding ERP-signatures of target awareness: Puzzle persists because of experimental co-variation of the objective and subjective variables, *Consciousness and Cognition*, **18** (3), pp. 804–808.

Baddeley, A. (2007) *Working Memory, Thought and Action*, Oxford: Oxford University Press.

Baird, B., Smallwood, J., Gorgolewski, K.J. & Margulies, D.S. (2013) Medial and lateral networks in anterior prefrontal cortex support metacognitive ability for memory and perception, *The Journal of Neuroscience*, **33** (42), pp. 16657–16665.

Bar-On, D. (2004) *Speaking My Mind: Expression and Self-Knowledge*, Oxford: Oxford University Press.

Bekinschtein, T.A., Dehaene, S., Rohaut, B., Tadel, F., Cohen, L. & Naccache, L. (2009) Neural signature of the conscious processing of auditory regularities, *Proceedings of the National Academy of Sciences*, **106** (5), pp. 1672–1677.

Bergström, F. & Eriksson, J. (2014) Maintenance of non-consciously presented information engages the prefrontal cortex, *Frontiers in Human Neuroscience*, **8**, art. 938.

Block, N. (1995) On a confusion about a function of consciousness, *Behavioral and Brain Sciences*, **18** (2), pp. 227–247.

Block, N. (2007) Consciousness, accessibility, and the mesh between psychology and neuroscience, *Behavioral and Brain Sciences*, **30** (5–6), pp. 481–548.

Block, N., Carmel, D., Fleming, S.M., Kentridge, R.W., Koch, C., Lamme, V.A.F. & Rosenthal, D. (2014) Consciousness science: Real progress and lingering misconceptions, *Trends in Cognitive Sciences*, **18** (11), pp. 556–557.

Boly, M., Balteau, E., Schnakers, C., Degueldre, C., Moonen, G., Luxen, A. & Laureys, S. (2007) Baseline brain activity fluctuations predict somatosensory perception in humans, *Proceedings of the National Academy of Sciences USA*, **104** (29), pp. 12187–12192.

Bronfman, Z.Z., Brezis, N., Jacobson, H. & Usher, M. (2014) We see more than we can report: 'Cost free' color phenomenality outside focal attention, *Psychological Science*, **25** (May), pp. 1–10.

Chalmers, D. (2000) What is a neural correlate of consciousness?, in Metzinger. T. (ed.) *Neural Correlates of Consciousness: Empirical and Conceptual Issues*, pp. 1–33, Cambridge, MA: MIT Press.

Charles, L., Van Opstal, F., Marti, S. & Dehaene, S. (2013) Distinct brain mechanisms for conscious versus subliminal error detection, *NeuroImage*, **73**, pp. 80–94.

Cohen, M.A. & Dennett, D.C. (2011) Consciousness cannot be separated from function, *Trends in Cognitive Science*s, **15** (8), pp. 358–364.

Cohen, M.A., Cavanagh, P., Chun, M.M. & Nakayama, K. (2012) The attentional requirements of consciousness, *Trends in Cognitive Sciences*, **16** (8), pp. 411–417.

Cohen, M.A., Dennett, D.C. & Kanwisher, N. (2016) What is the bandwidth of perceptual experience?, *Trends in Cognitive Sciences*, **20** (5), pp. 324–335.

Cortese, A., Amano, K., Koizumi, A., Kawato, M. & Lau, H. (2016) Multivoxel neurofeedback selectively modulates confidence without changing perceptual performance, *Nature Communications*, **7**, p. 13669.

Crick, F. & Koch, C. (1990) Towards a neurobiological theory of consciousness, *Seminars in the Neurosciences*, **2**, pp. 263–275.

de Gardelle, V. & Kouider, S. (2009) Cognitive theories of consciousness, in *Encyclopedia of Consciousness*, **1**, pp. 135–146, Amsterdam: Elsevier.

De Graaf, T.A., Hsieh, P.J. & Sack, A.T. (2012) The 'correlates' in neural correlates of consciousness, *Neuroscience and Biobehavioral Reviews*, **36** (1), pp. 191–197.

Dehaene, S. (2014) *Consciousness and the Brain: Deciphering How the Brain Codes Our Thoughts*, New York: Penguin Books.

Dehaene, S. & Naccache, L. (2001) Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework, *Cognition*, **79** (1), pp. 1–37.

Dehaene, S., Naccache, L., Cohen, L., Bihan, D.L., Mangin, J.F., Poline, J.B. & Rivière, D. (2001) Cerebral mechanisms of word masking and unconscious repetition priming, *Nature Neuroscience*, **4** (7), pp. 752–758.

Dehaene, S. & Changeux, J.P. (2011) Experimental and theoretical approaches to conscious processing, *Neuron*, **70** (2), pp. 200–227.

Del Cul, A., Baillet, S. & Dehaene, S. (2007) Brain dynamics underlying the nonlinear threshold for access to consciousness, *PLoS Biology*, **5** (10), pp. 2408–2423.

Del Cul, A., Dehaene, S., Reyes, P., Bravo, E. & Slachevsky, A. (2009) Causal role of prefrontal cortex in the threshold for access to consciousness, *Brain*, **132**, pp. 2531–2540.

Desender, K., Van Opstal, F., Hughes, G. & Van den Bussche, E. (2016) The temporal dynamics of metacognition: Dissociating task-related activity from later metacognitive processes, *Neuropsychologia*, **82**, pp. 54–64.

Doesburg, S.M., Green, J.J., McDonald, J.J. & Ward, L.M. (2009) Rhythms of consciousness: Binocular rivalry reveals large-scale oscillatory network dynamics mediating visual perception, *PloS One*, **4** (7), e6142.

Dresler, M., Wehrle, R., Spoormaker, V.I., Koch, S.P., Holsboer, F., Steiger, A., Obrig, H., Sämann, P.G. & Czisch, M. (2012) Neural correlates of dream lucidity obtained from contrasting lucid versus non-lucid REM sleep: A combined EEG/fMRI case study, *Sleep*, **35** (7), pp. 1017–1020.

Dutta, A., Shah, K., Silvanto, J. & Soto, D. (2014) Neural basis of non-conscious visual working memory, *NeuroImage*, **91**, pp. 336–343.

Einhäuser, W., Stout, J., Koch, C. & Carter, O. (2008) Pupil dilation reflects perceptual selection and predicts subsequent stability in perceptual rivalry, *Proceedings of the National Academy of Sciences USA*, **105** (5), pp. 1704–1709.

El Karoui, I., King, J.R., Sitt, J., Meyniel, F., Van Gaal, S., Hasboun, D., Adam, C., Navarro, V., Baulac, M., Dehaene, S., Cohen, L. & Naccache, L. (2015) Event-related potential, time-frequency, and functional connectivity facets of local and global auditory novelty processing: An intracranial study in humans, *Cerebral Cortex*, **25** (11), pp. 4203–4212.

Feyerabend, P. (1975) *Against Method*, 4th ed., London: Verso.

Filevich, E., Dresler, M., Brick, T.R. & Kühn, S. (2015) Metacognitive mechanisms underlying lucid dreaming, *Journal of Neuroscience*, **35** (3), pp. 1082–1088.

Fisch, L., Privman, E., Ramot, M., Harel, M., Nir, Y., Kipervasser, S., Andelman, F., Neufeld, M.Y., Kraner, U., Fried, I. & Malach, R. (2009) Neural 'ignition': Enhanced activation linked to perceptual awareness in human ventral stream visual cortex, *Neuron*, **64** (4), pp. 562–574.

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

Fleming, S.M., & Dolan, R.J. (2012) The neural basis of metacognitive ability, *Philosophical Transactions of the Royal Society B: Biological Sciences*, **367**, pp. 1338–1349.

Fleming, S.M., Huijgen, J. & Dolan, R.J. (2012) Prefrontal contributions to metacognition in perceptual decision-making, *Journal of Neuroscience*, **32** (18), pp. 6117–6125.

Fleming, S.M., Ryu, J., Golfinos, J.G. & Blackmon, K.E. (2014) Domain-specific impairment in metacognitive accuracy following anterior prefrontal lesions, *Brain: A Journal of Neurology*, **137** (Pt 10), pp. 2811–2822.

Frässle, S., Sommer, J., Jansen, A., Naber, M. & Einhauser, W. (2014) Binocular rivalry: Frontal activity relates to introspection and action but not to perception, *Journal of Neuroscience*, **34** (5), pp. 1738–1747.

Gaillard, R., Dehaene, S., Adam, C., Clemenceau, S., Hosboun, D., Baulac, M., Cohen, L. & Naccache, L. (2009) Converging intracranial markers of conscious access, *PLoS Biology*, **7** (3), pp. 0472–0492.

Goldberg, I.I., Harel, M. & Malach, R. (2006) When the brain loses its self: Prefrontal inactivation during sensorimotor processing, *Neuron*, **50** (2), pp. 329–339.

Gusnard, D.A., Akbudak, E., Shulman, G.L. & Raichle, M.E. (2001) Medial prefrontal cortex and self-referential mental activity: relation to a default mode of brain function, *Proceedings of the National Academy of Sciences USA*, **98** (7), pp. 4259–4264.

Hohwy, J. & Bayne, T. (2015) The neural correlates of consciousness: Causes, confounds and constituents, in Miller, S.M. (ed.) *The Constitution of Phenomenal Consciousness*, pp. 155–176, Amsterdam: John Benjamins.

Hurlburt, R.T. & Schwitzgebel, E. (2007) *Describing Inner Experience? Proponents Meets Skeptic*, Cambridge, MA: MIT Press.

Irvine, E. (2012) *Consciousness as a Scientific Concept*, New York: Springer.

Kentridge, R.W., Nijboer, T.C.W. & Heywood, C.A. (2008) Attended but unseen: Visual attention is not sufficient for visual awareness, *Neuropsychologia*, **46** (3), pp. 864–869.

Kim, C.Y. & Blake, R. (2005) Psychophysical magic: Rendering the visible 'invisible', *Trends in Cognitive Sciences*, **9** (8), pp. 381–388.

King, J.-R., Pescetelli, N. & Dehaene, S. (2016) Brain mechanisms underlying the brief maintenance of seen and unseen sensory information, *Neuron*, **92** (5), pp. 1122–1134.

Koch, C. (2004) *The Quest for Consciousness*, Englewood, CO: Roberts and Company.

Koch, C., Massimini, M., Boly, M. & Tononi, G. (2016) Neural correlates of consciousness: Progress and problems, *Nature Reviews Neuroscience*, **17** (5), pp. 307–321.

Lamme, V.A.F. (2004) Separate neural definitions of visual consciousness and visual attention; a case for phenomenal awareness, *Neural Networks*, **17** (5–6), pp. 861–872.

Lau, H.C. (2008) A higher order Bayesian decision theory of consciousness, *Progress in Brain Research*, **168**, pp. 35–48.

Li, F.F., VanRullen, R., Koch, C. & Perona, P. (2002) Rapid natural scene categorization in the near absence of attention, *Proceedings of the National Academy of Sciences USA*, **99** (14), pp. 9596–9601.

Lycan, W. (1996) *Consciousness and Experience*, Cambridge, MA: MIT Press.

Lyons, W. (1986) *The Disappearance of Introspection*, Cambridge, MA: MIT Press.

Maniscalco, B. & Lau, H. (2015) Manipulation of working memory contents selectively impairs metacognitive sensitivity in a concurrent visual discrimination task, *Neuroscience of Consciousness*, **2015** (1).

Markowitsch, H.J. & Kessler, J. (2000) Massive impairment in executive functions with partial preservation of other cognitive functions: The case of a young patient with severe degeneration of the prefrontal cortex, *Experimental Brain Research*, **133** (1), pp. 94–102.

McCurdy, L.Y., Maniscalco, B., Metcalfe, J., Liu, K.Y., de Lange, F.P. & Lau, H. (2013) Anatomical coupling between distinct metacognitive systems for memory and visual perception, *Journal of Neuroscience*, **33** (5), pp. 1897–906.

Melloni, L., Schwiedrzik, C.M., Muller, N., Rodriguez, E. & Singer, W. (2011) Expectations change the signatures and timing of electrophysiological correlates of perceptual awareness, *Journal of Neuroscience*, **31** (4), pp. 1386–1396.

Miller, S.M. (2014) Closing in on the constitution of consciousness, *Frontiers in Psychology*, **5** (Nov), pp. 1–18.

Müller, N.G. & Knight, R.T. (2006) The functional neuroanatomy of working memory: Contributions of human brain lesion studies, *Neuroscience*, **139** (1), pp. 51–58.

Naber, M., Frässle, S. & Einhäuser, W. (2011) Perceptual rivalry: Reflexes reveal the gradual nature of visual awareness, *PLoS ONE*, **6** (6).

Naber, M. & Brascamp, J. (2015) Commentary: Is the frontal lobe involved in conscious perception?, *Frontiers in Psychology*, **6** (Nov).

Neisser, J. (2012) Neural correlates of consciousness reconsidered, *Consciousness and Cognition*, **21** (2), pp. 681–690.

Newell, B.R. & Shanks, D.R. (2014) Unconscious influences on decision making: A critical review, *Behavioral and Brain Sciences*, **37** (1), pp. 1–19.

Nir, Y. & Tononi, G. (2010) Dreaming and the brain: From phenomenology to neuphysiology, *Trends in Cognitive Science*s, **14** (2), pp. 1–25.

Norman, L.J., Heywood, C.A. & Kentridge, R.W. (2013) Object-based attention without awareness, *Psychological Science*, **24** (6), pp. 836–843.

Overgaard, M. (2006) Introspection in science, *Consciousness and Cognition*, **15**, pp. 629–633.

Overgaard, M. (2015) The challenge of measuring consciousness, in Overgaard, M. (ed.) *Behavioral Methods in Consciousness Science*, Oxford: Oxford University Press.

Overgaard, M. & Sandberg, K. (2012) Kinds of access: Different methods for report reveal different kinds of metacognitive access, *Philosophical Transactions of the Royal Society B: Biological Sciences*, **367** (1594), pp. 1287–1296.

Overgaard, M. & Fazekas, P. (2016) Can no-report paradigms extract true correlates of consciousness?, *Trends in Cognitive Sciences*, **20** (4), pp. 241–242.

Pincham, H.L., Bowman, H. & Szucz, D. (2016) The experiential blink: Mapping the cost of working memory encoding onto conscious perception in the attentional blink, *Cortex*, **81**, pp. e157–e157.

Pitts, M.A., Martínez, A. & Hillyard, S.A. (2012) Visual processing of contour patterns under conditions of inattentional blindness, *Journal of Cognitive Neuroscience*, **24** (2), pp. 287–303.

Pitts, M.A., Metzler, S. & Hillyard, S.A. (2014) Isolating neural correlates of conscious perception from neural correlates of reporting one's perception, *Frontiers in Psychology*, **5** (Sep), pp. 1–16.

Pitts, M.A., Padwal, J., Fennelly, D., Martinez, A. & Hillyard, S.A. (2014) Gamma band activity and the P3 reflect post-perceptual processes, not visual awareness, *NeuroImage*, **101**, pp. 337–350.

Polich, J. (2007) Updating P300: An integrative theory of P3a and P3b, *Clinical Neurophysiology*, **118** (10), pp. 2128–2148.

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), pp. 972–983.

Rosenthal, D.M. (2005) *Consciousness and Mind*, Oxford: Oxford University Press.

Rounis, E., Maniscalco, B., Rothwell, J.C., Passingham, R.E. & Lau, H. (2010) Theta-burst transcranial magnetic stimulation to the prefrontal cortex impairs metacognitive visual awareness, *Cognitive Neuroscience*, **1** (3), pp. 165–175.

Rutiku, R., Aru, J. & Bachmann, T. (2016) General markers of conscious visual perception and their timing, *Frontiers in Human Neuroscience*, **10** (Feb), pp. 1–15.

Sadaghiani, S., Hesselmann, G. & Kleinschmidt, A. (2009) Distributed and antagonistic contributions of ongoing activity fluctuations to auditory stimulus detection, *Journal of Neuroscience*, **29** (42), pp. 13410–13417.

Schwitzgebel, E. (2008) The unreliability of naive introspection, *Philosophical Review*, **117** (2), pp. 245–273.

Sergent, C. & Dehaene, S. (2004) Is consciousness a gradual phenomenon?, *Psychological Science*, **15** (11), pp. 720–728.

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

Shafto, J.P. & Pitts, M.A. (2015) Neural signatures of conscious face perception in an inattentional blindness paradigm, *Journal of Neuroscience*, **35** (31), pp. 10940–10948.

Shoemaker, S. (1996) *The First-Person Perspective and Other Essays*, Cambridge: Cambridge University Press.

Silverstein, B.H., Snodgrass, M., Shevrin, H. & Kushwaha, R. (2015) P3b, consciousness, and complex unconscious processing, *Cortex*, **73**, pp. 216–227.

Silverstein, B.H., Snodgrass, M., Shevrin, H. & Kushwaha, R. (2016) Unconscious P3b and complex unconscious processing: Reply to Naccache et al., 2016, *Cortex*, **85**, pp. 129–132.

Soto, D. & Silvanto, J. (2014) Reappraising the relationship between working memory and conscious awareness, *Trends in Cognitive Sciences*, **18** (10), pp. 520–525.

Soto, D. & Silvanto, J. (2016) Is conscious awareness needed for all working memory processes?, *Neuroscience of Consciousness*, **2016** (Feb), pp. 1–3.

Stein, T., Kaiser, D. & Hesselmann, G. (2016) Can working memory be non-conscious?, *Neuroscience of Consciousness*, **2016** (1).

Sterzer, P., Kleinschmidt, A. & Rees, G. (2009) The neural bases of multistable perception, *Trends in Cognitive Sciences*, **13** (7), pp. 310–318.
Tallon-Baudry, C. (2012) On the neural mechanisms subserving consciousness and attention, *Frontiers in Psychology*, **3**, art. 397.
Timmermans, B., Sandberg, K., Cleeremans, A. & Overgaard, M. (2010) Partial awareness distinguishes between measuring conscious perception and conscious content: Reply to Dienes and Seth, *Consciousness and Cognition*, **19** (4), pp. 1081–1083.
Timmermans, B. & Cleeremans, A. (2015) How can we measure awareness? An overview of current methods, in Overgaard, M. (ed.) *Behavioral Methods in Consciousness Science*, New York: Oxford University Press.
Tsuchiya, N., Wilke, M., Frässle, S. & Lamme, V.A.F. (2015) No-report paradigms: Extracting the true neural correlates of consciousness, *Trends in Cognitive Sciences*, **19** (12), pp. 757–770.
Tsuchiya, N., Frässle, S., Wilke, M. & Lamme, V. (2016) No-report and report-based paradigms jointly unravel the NCC: Response to Overgaard and Fazekas, *Trends in Cognitive Sciences*, **20** (4), pp. 242–243.
Valk, S.L., Bernhardt, B.C., Böckler, A., Kanske, P. & Singer, T. (2016) Substrates of metacognition on perception and metacognition on higher-order cognition relate to different subsystems of the mentalizing network, *Human Brain Mapping*, **37** (10), pp. 3388–3399.
van Aalderen-Smeets, S.I., Oostenveld, R. & Schwarzbach, J. (2009) Investigating neurophysiological correlates of metacontrast masking with magnetoencephalography, *Advances in Cognitive Psychology*, **2** (1), pp. 21–35.
Van Boxtel, J.J.A., Tsuchiya, N. & Koch, C. (2010) Consciousness and attention: On sufficiency and necessity, *Frontiers in Psychology*, **1**, art 217.
van Gaal, S., Lamme, V.A.F., Fahrenfort, J.J. & Ridderinkhof, K.R. (2011) Dissociable brain mechanisms underlying the conscious and unconscious control of behavior, *Journal of Cognitive Neuroscience*, **23** (1), pp. 91–105.
Williams, M.A., Visser, T.A.W., Cunnington, R. & Mattingley, J.B. (2008) Attenuation of neural responses in primary visual cortex during the attentional blink, *Journal of Neuroscience*, **28** (39), pp. 9890–9894.
Zaretskaya, N. & Narinyan, M. (2014) Introspection, attention or awareness? The role of the frontal lobe in binocular rivalry, *Frontiers in Human Neuroscience*, **8**, art. 527.

Paper received February 2017; revised May 2017.