MARIAN SCHNEIDER & INGO MARQUARDT

The negative neural correlates of consciousness

Essay

For a long time, scientists and philosophers alike have been speculating about the neural correlates of consciousness (NCC). Although definitions differ, usually an NCC denotes neural events whose occurrence correlates with our phenomenal experiences. In this essay, we motivate the introduction of the concept of a 'negative neural correlate of consciousness' (NNCC). The NNCC can be conceptually understood as neural events whose occurrence inversely correlates with a given phenomenal experience. We wish to introduce this concept for two reasons. First, it is an elegant solution to the problem of sufficiency that so far has unduly been neglected in the NCC debate. Secondly, based on the Integrated Information Theory of Consciousness, we argue that only by pairing up each NCC with a corresponding NNCC will scientists ever be able to predict conscious experience from brain data. We discuss empirical implications of and potential objections to the suggested NNCC framework.
Keywords: Consciousness, Neural Correlates of Consciousness, Philosophy of Mind, Cognitive Neuroscience, Integrated Information Theory of Consciousness

Marian Schneider
Ingo Marquardt
Maastricht University, Maastricht, The Netherlands

Marian.schneider@student.maastrichtuniversity.nl &
i.marquardt@student.maastrichtuniversity.nl

INTRODUCTION

In his seminal essay "What is a Neural Correlate of Consciousness?" David Chalmers (2000) discussed important conceptual aspects and caveats of the notion of a neural correlate of consciousness (NCC). He roughly approximated an NCC as a neural system N whose state correlates directly with states of consciousness.

Taking this approximation as a starting point, he then focused his attention on two aspects: what the relevant states of consciousness are and what it means for a state to correlate directly. With respect to the first point—the states of consciousness—he differentiated between states of merely being conscious vs. unconscious, background states of consciousness (awake, asleep, dreaming, under hypnosis), and contents of consciousness. The contents of consciousness are the fine-grained states of subjective experience that one is in at any given time (e.g. experience of a particular visual image, of a particular sound pattern, of a detailed stream of conscious thought). Accordingly, Chalmers defined a content NCC as a neural system N such that the content of N directly correlates with the content of consciousness. Regarding the second point—the nature of the correlation—he concluded that the NCC state should be minimally sufficient for the conscious state. Consequently, he arrived at the following formal definition of an NCC:

> "An NCC is a minimal neural system N such that there is a mapping from states of N to states of consciousness, where a given state of N is sufficient, under conditions C, for the corresponding state of consciousness" (Chalmers, 2000, p.22).

If one is specifically interested in the contents of consciousness, the definition can be put in more specific terms:

> "An NCC (for content) is a minimal neural representational system N such that representation of a content in N is sufficient, under conditions C, for representation of that content in consciousness" (Chalmers, 2000, p.22).

Several authors have pointed out potential shortcomings and blind spots of these definitions. Noe and Thompson (2004) noted that the concept of an NCC rests on an internalist assumption and that this assumption by itself is still a matter of debate within philosophy of mind. Cleeremans (2005) drew attention to the fact that we should not only focus our search on *neural* correlates but also on *computational* correlates of consciousness. Finally, de Graaf, Hsieh and Sack (2012) argued that the term 'correlate' in 'neural correlate of consciousness' is ambiguous and needs to be refined. In their argumentation, de Graaf et al. (2012) pointed out that an NCC may fall into one of three distinct categories: It may be a prerequisite, a consequence, or the actual substrate of a conscious experience. Prerequisites and consequences of a conscious state are likely to be very general and may for example include neural mechanisms that are also associated with arousal and wakefulness. For instance, there are neural systems in the brain stem that are most likely important for general arousal and wakefulness, but these systems are probably not directly relevant for the specific content of conscious experience. Although prerequisites and consequences by definition always co-occur with a conscious experience, they do not directly underlie its subjective phenomenal content—as the neural substrate does. Only the neural substrate is the neural state that is directly underlying the experience and is neither a result nor a precondition for a specific conscious percept. It is the neural substrate that is most relevant to the consciousness debate (de Graaf et al., 2012). Overall, Chalmers' original NCC concept has led to a very productive debate, and several refinements to the original definition have been proposed.

In this essay, we would like to present a further complement to the traditional concept of an NCC. An important aspect that has often been only implicitly assumed (or even neglected) in the past conceptual debate about the NCC is that the concept of correlation implies both positive and negative correlation. This becomes apparent when revisiting Chalmers (2000) definition and reconsidering the mapping of states from a neural system N to states of consciousness. As we argue below, representation of content in N will only be sufficient for the representation of that content in consciousness if, at the same time, the content in systems other than N is prevented from being mapped to consciousness. This additional qualification can be captured conceptually as negative correlation. In this context, a positive correlation means that if a neural event N_1 is present, a corresponding state of conscious experience P_1 is present ($r = 1$). By contrast, a negative correlation can be conceptually understood as the absence of specific neural events N_2 to N_n, while a phenomenal experience P_1 is fully present ($r = -1$). The new specification (that the absence of neural events N_2 to N_n will be as crucial as the presence of event N_1 for conscious content) has not, or only implicitly, been considered in the past conceptual debate. Chalmers' definition, for example, makes no explicit mention of such an additional qualification. We suggest calling neural events N_2 to N_n the negative neural correlate of consciousness (NNCC). By making an implicit assumption explicit, we hope to offer a new exciting conceptual framework, able to motivate new hypotheses and research.

Although the NNCC concept is based on a fairly simple idea we believe that it is not trivial. In fact, we argue that the concept of negative correlation matters for the search for NNCs, for two main reasons. First, as has long been recognized in the philosophy of science, a set of events can only be *sufficient* to produce an effect-event if the set includes an open-ended *ceteris paribus* clause excluding the presence of potential disruptors that could intervene to prevent *E* (Hoefer, 2010; Russell,1912). Thus, if we want a neural event N to be (minimally) sufficient for a corresponding conscious state, we will be forced to include an open-ended list that ensures the absence of potential 'content disrupters'. Content disrupters can be understood as neural events that interfere with the otherwise sufficient mapping of content from representative neural systems to conscious experiences. Here, we suggest that a more elegant solution to this problem is to pair each NCC with a corresponding NNCC, which is defined in global terms. Secondly, there is a strong additional motivation to complement the conventional ('positive') NCC by a corresponding NNCC. We argue that an NCC by itself is not able to account for the content of a corresponding conscious experience. To fully predict the content of a conscious experience, the NCC needs to be paired up with an NNCC. Consequently, an NCC (in the 'positive correlation' sense) will not be minimally sufficient for a corresponding conscious state. This argument is based on the Integrated Information Theory of Consciousness (IITC; Tononi, 2004, 2008, 2012). For these reasons, we think that the concept of a NNCC has important implications for the debate about consciousness and the search for the brain states corresponding to conscious experience.

FIRST REASON: THE SUFFICIENCY PROBLEM

The basic assumption underlying the search for an NCC over the last 20 years has been that it is possible to find minimal neural events that will be sufficient to produce conscious experiences. However, the concept of sufficiency is fraught with problems, as has long been recognized in the philosophy of science (Russell, 1912). To understand why that is the case, imagine that we say that a set of events A, B, C will be sufficient to "make me go to the fridge to get a beer". For illustrative purposes, let us say that these three events are: I am thirsty (A), I want to quench my thirst with a beer (B) and I assume there is beer in the fridge (C).

If the set A, B, C is present, I will go to the fridge and have my beer. No worries here. However, let us consider three alternative scenarios. Imagine that the set A, B, C is present but that just when I am about to go to the fridge my friend tells me that there is no more beer in the fridge. Or, imagine that just when I want to get up from the sofa I suffer a sudden stroke and fall unconscious. And yet in a different case, imagine that when I am already in the kitchen a bird crashes into the kitchen windowpane and because I am much of a bird carer I hurry to its rescue. These scenarios highlight the point that suddenly the set of events A, B, C will not be sufficient anymore for me to go to the fridge and have a beer. Although A, B, and C are conjointly present and on every usual day would have been enough to make me have a beer, in these scenarios they were not.

Confronted with this problem, one could start making ad-hoc modifications. One could say that A, B, and C will still be sufficient to make me have a beer if my friend does *not* tell me that no beer is left, I do *not* suffer a sudden stroke, and *no* bird crashes into the kitchen window. Yet, as becomes easily apparent the list of things that do *not* have to happen for A, B, C to still be sufficient for the effect-event will be endless. We would, for example, also need to exclude that one of my ancestor died before she could participate in the reproduction relevant for my case, and that I did not choke on a pretzel the day before I wanted to get my beer, and so forth. More generally and formally, to achieve the desired sufficiency, we would be forced to accept an open-ended list of negative conditions. Such an open-ended *ceteris paribus* clause is certainly unsatisfactory from a philosophical point of view.

Why do these considerations matter for the concept of a (content) NCC? Let us assume that there is a minimal neural event N1. And let us furthermore assume that N1 is the perfect content NCC for the conscious experience P1. Thus, whenever N1 occurs a corresponding state of conscious experience P1 will also occur. According to Chalmers' (2000) definition, we could say that N1 is minimally sufficient for P1. However, consider the case that simultaneously to N1 another neural event N2 also occurs. This co-occurrence does not necessarily cause the phenomenal experience P1 anymore but might cause a fundamentally different phenomenal experience P2 (see below for a concrete example). Again, we could modify and say that N1 will still be sufficient for P1 unless N2 occurs. Yet—just as in the example with the beer—this would leave us with an open-ended list of potential disrupters that we would need to exclude. We suggest calling events like N2 *content disrupters*, for their occurrence disrupts the otherwise sufficient mapping of content from representative neural systems to conscious experiences like P1. As we hope to have shown in this section, if we wanted a neural event N to be (minimally) sufficient for a corresponding conscious state, this would force us to include an open-ended *ceteris paribus* clause excluding the presence of potential content disruptors that could intervene to prevent P1.

SECOND REASON: THE INFORMATION INTEGRATION THEORY OF CONSCIOUSNESS

In addition to the before mentioned philosophical motivation for introducing an NNCC, we think that the NNCC can be further motivated and sustained by the help of a thought experiment. The thought experiment is inspired by Guilio Tononi's Integrated Information Theory of Consciousness (2004, 2008) as well as by his recent book "Phi" (2012).

THE PHOTODIODE THOUGHT EXPERIMENT

Imagine that you are standing in an empty room, that the light is switched off and that all sounds are absent. The sound of your own breathing can be neglected for now. Assume further that you have regular functioning of your body including your brain and that you are awake and thus fully conscious. Everything is black and you have the conscious experience of seeing black only; the room is perfectly quiet and you are not exposed to any other sensory sensations. Imagine too that there is also a photodiode in the room. A photodiode is an electronic component that senses light. The photodiode in this thought experiment can differentiate between two states: dark or light. If one were to ask you now whether you experience darkness or lightness, you would in all likelihood say 'darkness' since the room is dark. If one were to consult the photodiode about the same question, it would equally signal 'darkness'. Does that mean the photodiode has the same phenomenal experience you have? The intuitive answer is 'no'. We guess that most people would intuitively deny that the photodiode has any conscious experience at all. Yet, we cannot be sure of this since there is no consciousness meter to check, and intuition can trick us into conclusions that are very remote from the truth.

In order to ground answers in a sounder framework, let us refer to Guilio Tononi's (2004, 2008) Integrated Information Theory of Consciousness. According to this theory, consciousness corresponds to the capacity of a system to integrate information. This proposition is based on two phenomenological properties of consciousness: the unity of conscious experience (conscious experience is always integrated) and the availability of a large number of different conscious experiences (differentiation). Information is generated by differentiating between conscious states. Furthermore, the theory suggests measuring the degree of integrated information as the phi value of a system. According to the theory then, both the human and the photodiode in the dark room are conscious. Yet, the human is conscious to a much higher degree than the photodiode is (the human has a much higher phi value). One might wonder how that is possible even though both human and the photodiode are confronted with the same situation—total darkness, total black—and both signal the same responses. The theory claims that the capacity of the human to integrate information is higher than that of the photodiode. The difference between the human and the photodiode in this case does not lie in the integration aspect, which should be similar, but it lies in the information aspect. The human integrates much more information than the photodiode does. In this context, information is understood as the reduction of uncertainty among a number of alternative states. Thus, information is mainly a measure of which phenomenal experiences are excluded by the current phenomenal experience. If the photodiode is in the conscious state of experiencing 'dark', it only excludes one single other state ('light') and therefore does not generate much information. By contrast, if the human is in the conscious state of only experiencing 'dark', this excludes a myriad of other possible phenomenal states. If a human only experiences the blackness of a dark room, this experience excludes the possibility that the human at that moment experiences blue, green, yellow, white, and any other colour. Yet, it does not only exclude the possibility of experiencing colours or seeing anything else (shapes, faces, words). It furthermore excludes the possibility that the human was, at that moment, experiencing any other sensation be it a sound, smell, touch or an itchy right foot. Moreover, the example phenomenal state excludes experiencing any possible combination of the aforementioned sensations. Thus, this one particular phenomenal experience excludes an unlimited number of alternative phenomenal states for the human, while it excludes only a single other states for the photodiode. In different words, the photodiode's repertoire is minimally differentiated, while the one of the human is immensely so. As a result, the human is much more conscious than the photodiode; yet, the photodiode is still conscious with one bit because it knows to differentiate between two states (Tononi, 2004).

IMPLICATIONS OF THE PHOTODIODE THOUGHT EXPERIMENT

What implications do the insights gained above have for the search for an NCC? We think that the thought experiment with the photodiode and the Integrated Information Theory of Consciousness imply the concept of a negative neural correlate of consciousness (NNCC). The total absence of light, as in the thought experiment, is a special case because the absence of light implies the absence of any colour. Therefore, let us now assume that the room is illuminated in blue. In this case, the human in the room would have the phenomenal experience of seeing blue. To predict this first-person perspective of experiencing blue from the third-person perspective of brain data, researchers can proceed in two steps: First, they can try to identify the content NCC for experiencing blue (via bridging principles). Second, if this content NCC is present, they can predict the corresponding phenomenal experience to occur. If the content NCC were absent, they would predict the absence of the phenomenal experience. This is what we call a *positive* content NCC. As science stands at the moment, the occurrence of particular neural events within brain areas V4 and V8 are hot candidates for a positive content NCC of experiencing colours.

Consider the ideal case (for a neuroscientist) that a particular neural event in brain areas V4 and V8 would produce the conscious experience of the colour blue, and all other brain activity that correlates with this experience could be classified as neural prerequisites or neural consequences of consciousness. Hence, there would be a perfect (positive) correlation between the experience of blue and the identified neural events. If we wish to predict the content of experience from brain data, would it be right then to conclude that only this particular content NCC is relevant for our experience of blue? Could we really ignore all other neuronal events in areas V4 and V8 as well as in the rest of the brain? We believe that this is not the case. Those neuronal events in areas V4 and V8 that negatively correlate with the experience of blue (and positively correlate with the experience of colours other than blue) equally contribute to our experience of blue, namely by their absence. This may sound odd at first. However, we just established that what makes the human conscious experience different from the conscious experience of a photodiode, for example, is that being in a particular phenomenal state excludes a myriad of other possible phenomenal states. The phenomenal state of experiencing only blue excludes all other possible states of colour experience. Thus, when we want to predict a given first-person perspective of a human from brain data, we cannot only focus on those neural events that *do* occur (e.g. a certain neural events in areas V4 and V8). Additionally, we need to consider other possible neural events that *do not* occur. Those off-states also contribute to our experience because it is only for them that information is generated and we are conscious the way we are. Only by considering those off-states too, will we be able to fully account for the experience of what it is like to be in a blue room for a human, as opposed, for example, for a photodiode.
There is no reason to restrict the NNCC to the colour domain or even the visual domain. The same reasoning equally applies to other sensory domains such as sound, smell, taste and touch. Moreover, internally generated sensations, thoughts, or states of mind cannot be ignored either and likewise have to be considered in the search for negative content NCCs.

NNCC – ATTEMPT OF A FORMAL DEFINITION

Correlation also implies negative correlation. Thus, for the definition of an NCC it may not only be important which neural events are *present*, but also which ones are *absent*. According to our proposal, the quality and content of phenomenal experiences is not only determined by the presence, but also by the absence of neural events. Hence, the definition of a neural event that is minimally sufficient to produce a particular phenomenal experience would not only have to include the presence of certain neural events, but also the absence of others. This idea of an NNCC can be expressed more formally.

NNCC EQUALS NOMOLOGICAL POSSIBILITY SPACE MINUS NCC OF THE GIVEN EXPERIENCE

Assume that there is a set S of neural events N_1 to N_n. Let us further assume that each of these neural events qualifies as a content NCC, in the 'positive' way it has been traditionally understood. Each of these neural events N_1 to N_n represents content of some sort. Thus, if one of these neural events, for example event N_i, occurred, one may predict the occurrence of a corresponding phenomenal experience P_i. In accordance with Chalmers' (2000) conceptual framework, N_i would be minimally sufficient for the phenomenal state P_i. However, we claim that the quality and content of the phenomenal experience P_i is just as much determined by the absence of the set of neural events $S \neg N_i$. We suggest calling this set $S \neg N_i$ the *negative neural correlate* of the conscious experience P_i since its occurrence relates inversely to P_i.

In fact, as we have argued, the set S should be global and include *all* neural events for which there is a possibility that the event be an NCC. First, this is because not including all those events would force us to make endless additional specifications to each NCC by excluding the occurrence of content disrupters whose presence could intervene to prevent P_i. Secondly, it should include all neural events because it is the exclusion of these events that generates information and thus eventually conscious experience. If we do not account for all of them, we will not be able to predict the difference between the conscious experience of a photodiode and a human when perceiving darkness.

When we say that the set S has to include all neural events that could possibly be NCCs, we are concerned with nomological possibility and not with logical possibility. In other words, the set S has to include all neural events that could be an NCC, given the way an individual brain is structured and functions, and given the laws of nature that determine how brains produce phenomenal experiences. In contrast, the set S does not have to include all neural events that could hypothetically be an NCC in differently structured nervous systems or even under different psychophysical laws, as this would unnecessarily inflate the concept. For these reasons, we propose to call S the nomological possibility space for NCCs. For a given conscious experience, the NNCC will then be defined by the nomological possibility space of NCCs minus the NCC of the given experience itself. Pairing up the NCC with a corresponding NNCC enables us to make sure that the neural correlate we are searching for is truly sufficient for the corresponding phenomenal state. This conceptual modification may bring us a considerable step closer to predicting conscious experience from brain data.

THE NNCC IS CONTENT-SPECIFIC AND SHOULD NOT BE SUBSUMED UNDER CONDITIONS C

A different solution to the two problems outlined above might be to include the absence of neural events S ¬ Ni in the definition of conditions C under which N_i is the neural correlate of P_i. However, we think that this is clearly not satisfactory because conditions C are not supposed to comprise components that are related to the specific content of a phenomenal experience. Instead, conditions C ought to be content-invariant background processes. Therefore, a better solution is to amend our concept of the neural correlate of P_i. Accordingly, the true neural correlate of P_i is the pair of the positive correlate P_i (as defined so far) and the negative correlate S ¬ N_i. If and only if both the positive correlate Ni occurs and the negative correlate S ¬ N_i does not occur, under conditions C, does this lead to the phenomenal experience P_i. Thus, S ¬ N_i cannot be ignored in defining the neural correlate of the phenomenal state P_i. Therefore, negative neural correlates should not be disregarded in the search for a neural correlates of consciousness.

REINTERPRETING EMPIRICAL RESULTS – A CONCRETE EXAMPLE OF THE NNCC FRAMEWORK

Admittedly, given the current limitations of neuroimaging, especially in humans, it will not be possible in the near future, if ever, to map out the nomological possibility space for NCCs. Thus, empirical scientists might turn away disappointedly from our all too theoretical speculations. Yet we think that even for those scientists the NNCC concept offers a new framework of how to think about the results of research. This point can be illustrated by the example of past research on perceiving stationary and moving stimuli.

Several investigators (Beckers & Hoemberg, 1992; Sack, Kohler, Linden, Goebel, & Muckli, 2006; Silvanto, Lavie, & Walsh, 2005) used transcranial magnetic stimulation (TMS) over human brain area V5 to induce transient periods of akinetopsia. Participants were presented with moving random dot patterns and had to identify the direction of movement. During particular time windows, TMS over V5 disrupted motion perception. The time course of this effect (Sack et al., 2006; Silvanto et al., 2005) together with the results from a different TMS paradigm (Pascual-Leone & Walsh, 2001) suggests that in order to perceive the movement of a stimulus, this stimulus first needs to be processed in early visual cortex, followed by activity in V5, and subsequent recurrent activation of early visual cortex. These experiments suggest that the effect of disrupting area V5 is not that perception of a stimulus is obliterated in general, but that the percept loses a certain phenomenal quality—namely that of movement. Hence, neural events in area V5 positively correlate with the perception of movement. Therefore, neural events in area V5 may be part of a negative correlate for the perception of a stationary stimulus. In the following, we are going to further strengthen this claim.

The motion aftereffect is a perceptual illusion that occurs after prolonged exposure to a moving stimulus. For instance, imagine you look at a waterfall for about a minute, without changing your gaze. When you afterwards fixate a different scene without movement, it seems as if part of this new scene is moving. This effect is specific to the part of your visual field where you previously saw the waterfall. This motion aftereffect has been studied with fMRI (Tootell et al., 1995; Culham et al., 1999; He, Cohen, & Hu, 1998), and it was shown that the illusionary movement of stationary stimuli after movement adaptation correlates with activity in area V5. Furthermore, area V5 is not simply active for a particular period of time after adaptation to a moving stimulus. Rather, it only activates when a stimulus is presented in the adapted region of the visual field and the motion aftereffect is actually experienced (Culham et al., 1999; He, Cohen, & Hu, 1998). If a period of darkness—i.e. total absence of visual stimulation—is included after adaptation, no movement is experienced and area V5 is not activated. When a stationary stimulus is subsequently presented, activity in area V5 and illusionary movement perception resume.

Additional evidence comes from an animal study in which electrical microstimulation was applied to direction-sensitive neurons in monkey area V5, also known as area MT (Salzman, Britten, & Newsome, 1990). Monkeys were trained to report the direction of movement of visual stimuli. When V5 neurons that had a particular movement direction preference were stimulated, the monkeys' reports were biased towards this direction. Although no introspective verbal report can be obtained in an animal study, the monkeys' behaviour suggests that the microstimulation probably induced an illusory percept of movement into a particular direction.

The experiments on the motion aftereffect in humans and microstimulation in area V5 in monkeys can plausibly be interpreted in terms of the NNCC framework. It appears that in order to perceive a visual stimulus as stationary, certain neural events in area V5 must *not* occur. Otherwise, the stimulus is perceived as moving—even if in fact it is stationary. If, for instance, after having looked at a waterfall for a minute or so one directs their gaze at a rock face next to the waterfall, a stripe of rock appears to be moving upwards. Hence, we expect the neural substrate of consciously perceiving the same rock face without movement to include a negative correlate—namely the absence of particular neural events in area V5.

POTENTIAL OBJECTIONS TO THE NNCC AMENDMENT

We would like to respond to several anticipated objections to our NNCC framework. First, it could be argued that the NNCC framework is not backed up by empirical observations. Secondly, one could object that if we took the NNCC framework seriously, we would have to include the whole brain as a content NNCC for any given conscious experience. Thirdly, the NNCC framework might be considered irrelevant for human-to-human comparisons. In the following, we will deal with each of these objections in turn.

THE NNCC FRAMEWORK IS CONCEPTUAL, NOT EMPIRICAL

The nature of our argumentation is conceptual, not empirical. As things currently stand, we do not know the exact neural events that cause a particular phenomenal state, although many empirical candidates for an NCC have been proposed, such as 40 Hz oscillations (Crick & Koch, 1990; Engel & Singer, 2001), fast recurrent cortical feedback (Lamme, 2010), and many others. Therefore, we want our framework to be broad enough to leave room for different concrete neuronal correlates, and we refer to these as 'neural events'. In this context, it is important to note that the positive and negative neural correlates of a phenomenal experience can probably not simply be equated with two neural populations that ought to be active and silent, respectively, to cause a particular phenomenal experience. It is an open empirical question to find out what the exact neural correlates of consciousness are, but we would like to consider one example NNC in order to elucidate why the NNCC cannot necessarily be equated with a drop in firing rate. Baars (1988) put forward the theory that content becomes phenomenally conscious if it has access to the 'global workspace'. On the neural level, access to the global workspace could be mediated by a change in the synchronisation of neural populations (Dehaene, Changeux, Naccache, Sackur, & Sergent, 2006; Engel & Singer, 2001). Importantly, in this case the neural correlate of a phenomenal state may include absolutely no change in the overall firing rate of the relevant neural populations, but only a change in the timing (i.e. the phase) of spikes. This point illustrates why we prefer to use the term 'neural events' instead of for example 'neural activity'.

THE NNCC IS GLOBAL IN AN ANATOMICAL SENSE, BUT SPECIFIC IN A FUNCTIONAL SENSE

Another objection to our proposal could be that the NNCC framework suggests including the whole brain as a content NNCC for any given conscious experience. Then, it could be contended that nothing would be gained by doing so. We would like to respond in two ways.

First, let us come back to the above example of experiencing a stationary visual stimulus. Again, the NNCC does not predict a total absence of activity in neural population. Even if a person is experiencing a stationary visual stimulus, motion sensitive cortical areas such as V5 do not necessarily go into hibernation and cease all metabolic and neural activity. The NNCC states that precisely those neural events that would give rise to the conscious experience of motion are not present. For instance, let us assume that fast recurrent cortical backward projections are necessary for any conscious experience. Then we would argue that in order to perceive a given stimulus as stationary, no fast recurrent cortical feedback projections from area V5 to area V1 must occur. Thus, the NNCC does not subsume every kind of neural activity in a relevant brain area or cortical module or between these. Rather, it selectively excludes precisely those neural events that would give rise to conscious experience. Therefore, the NNCC is global in an anatomical sense, since the neural events that would act as a content disrupter for the respective phenomenal state must not occur in a whole lot of brain areas. In contrast, the NNCC is specific in a functional sense, because it only excludes precisely those neural events that are relevant for bringing about consciousness.

Having acknowledged the anatomically global character of the NNCC framework, we would like to voice a second response to the criticism that our proposal will result in too inclusive definitions. Namely, brain systems outside the cortical complex are unlikely to be included in our search. Neural events in the cerebellum, for instance, may neither qualify as a content NCC nor as a NNCC, for neither their presence nor absence contributes to our phenomenal experience. According to all the clinical and experimental evidence that has been gathered so far, the cerebellum does not contribute to our phenomenal experience (Tononi, 2012). This is accounted for by Tononi's (2004, 2008) theory, which assumes that although the cerebellum performs very sophisticated calculations, it is incapable of integrating information and thus does never correspond to conscious experience. Similarly, the processes in the brainstem and other subcortical systems do not need to be included in an NNCC because they are already accounted for as non-content-specific background conditions C in the definition of the neural correlate of consciousness.

A good guideline for deciding whether a neuronal event is an NNCC is to look whether there is any (nomologically) possible situation or case in which this neuronal event could contribute to our phenomenology—independently of whether it contributes at the very moment we measure it. If there is the nomological possibility, then the neuronal event should be considered as an NNCC. If there is no such possible case, like for the cerebellum, then it should not be considered as an NNCC. From all we know at the moment, the thalamocortical system might be a good starting point to look for content NNCCs.

THE NNCC FRAMEWORK ALLOWS ACCOUNTING FOR HUMAN-TO-HUMAN VARIABILITY

Finally, another objection in the vicinity may be that the notion of an NNCC is not all too relevant for human-to-human comparisons. Critics might say that it is reasonable to expect that the nomological possibility space is roughly the same across humans. The NNCCs thus could be kept constant and would not matter too much for human-to-human comparisons. However, making this simplifying assumption would only address the problem resulting from the IITC (second reason) but not the problem of sufficiency (first reason). Furthermore, we think that we ought to be prepared to face some considerable human-to-human variability in the content of phenomenal experiences; maybe we could simply ignore it so far because of the limitations in our current methods of comparison (e.g. verbal reports).

DIRECTIONS FOR FUTURE RESEARCH

A particularly interesting direction for future research would be to investigate the predictions made by the NNCC framework at a microscopic level. Could it for example be that the neural correlate of perceiving the colour red at a given location in the visual field comprises particular neural event in colour-sensitive cells of the cortical module with the corresponding receptive field as a positive correlate, and the absence of the same neural events in other colour-sensitive cells in the same module? A similar logic may apply to the conscious perception of orientation. Another interesting prediction of the NNCC framework concerns the effects of lesions. If for example area V5 is lesioned, does this 'reduce' the total size of the content NNCC of perceiving a stationary stimulus? Does it reduce the overall capacity of visual consciousness? The NNCC framework predicts that the answer to these questions is yes. So what are the consequences of extreme lesions, such as for example severing the two cortical hemispheres like in split-brain patients, for the neural correlates of consciousness? We hope that the NNCC framework will inspire an interesting debate and further empirical investigations.

CONCLUDING REMARKS

We introduced the concept of a negative neural correlate of consciousness (NNCC). An NNCC can be understood as a neural event that contributes to the content of phenomenal experience via its absence. Our framework is motivated by the problem of sufficiency. Furthermore, the framework is inspired by Tononi's (2004, 2008) information integration theory of consciousness, according to which "consciousness corresponds to the capacity of a system to integrate information" (Tononi, 2004, p. 1). In this theory, the unity of phenomenal experience is reflected in the fact that conscious systems *integrate* information, and the complexity of conscious experience is dependent on the ability of the system to *differentiate* between many alternative states. If consciousness is characterised by the integration of information and differentiation of alternative states, we cannot ignore those states that are excluded in our search for the neural correlate of consciousness. Hence, for the definition of a neural correlate of consciousness it may not only be important which neural events are present, but also which ones are absent. After all, the quality and content of phenomenal experiences is not only determined by the presence, but also by the absence of neural events. Therefore, the definition of a neural event that is minimally sufficient to produce a particular phenomenal experience would not only have to include a certain set of neural events that are present, but also such that are absent.

Acknowledgements

We would like to thank David Chalmers and Gulio Tononi for their inspiring writings, as well as our teacher Rob de Vries for his guidance and for igniting the consciousness fire within us.

REFERENCES

Baars, B. J. (1988). *A cognitive theory of consciousness*. Cambridge, MA: Cambridge University Press.

Beckers, G., & Homberg, V. (1992). Cerebral visual motion blindness: transitory akinetopsia induced by transcranial magnetic stimulation of human area V5. *Proceedings of the Royal Society of London*, 249, 173-178.

Chalmers, D. (2000). What is a neural correlate of consciousness? In T. Metzinger (Ed.), *Neural correlates of consciousness: empirical and conceptual questions*. Cambridge, MA: MIT Press.

Cleeremans, A. (2005). Computational correlates of consciousness. *Progress in brain research*, 150, 81-98.

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

Culham, J. C., Dukelow, S. P., Vilis, T., Hassard, F. A., Gati, J. S., Menon, R. S., & Goodale, M. A. (1999). Recovery of fMRI activation in motion area MT following storage of the motion aftereffect. *Journal of Neurophysiology*, 81, 388-393.

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

Dehaene, S., Changeux, J. P., Naccache, L., Sackur, J., & Sergent, C. (2006). Conscious, preconscious, and subliminal processing: a testable taxonomy. Trends in cognitive sciences, 10(5), 204-211.

Engel, A. K., & Singer, W. (2001). Temporal binding and the neural correlates of sensory awareness. *Trends in cognitive sciences*, 5(1), 16-25.

He, S., Cohen, E. R., & Hu, X. (1998). Close correlation between activity in brain area MT/V5 and the perception of a visual motion aftereffect. *Current Biology*, 8(22), 1215–1218.

Lamme, V. A. F. (2010). How neuroscience will change our view on consciousness. *Cognitive neuroscience*, 1(3), 204-240.

Noë, A., & Thompson, E. (2004). Are there neural correlates of consciousness?. *Journal of Consciousness studies*, 11(1), 3-28.

Pascual-Leone, A., & Walsh, V. (2001). Fast backprojections from the motion to the primary visual area necessary for visual awareness. *Science*, 292, 510-512.

Russell, B. (1912/2004). On the notion of cause. *Proceedings of the Aristotelian Society,*13, 1–26.

Sack, A. T., Kohler, A., Linden, D. E. J., Goebel, R., & Muckli, L. (2006). The temporal characteristics of motion processing in hMT/V5+: Combining fMRI and neuronavigated TMS. *Neuroimage*, 29, 1326-1335.

Salzman, C. D., Britten, K. H., & Newsome, W. T. (1990). Cortical microstimulation influences perceptual judgements of motion direction. *Nature*, 346, 174-177.

Silvanto, J., Lavie, N., & Walsh, V. (2005). Double Dissociation of V1 and V5/MT activity in Visual Awareness. *Cerebral Cortex*, 15, 1736-1741.

Tononi, G. (2004). An information integration theory of consciousness. *BMC Neuroscience*, 5(42).

Tononi, G. (2008). Consciousness as integrated information: a provisional manifesto. *The*

Biological Bulletin, 215(3), 216-242.

Tononi, G. (2012). *Phi. A voyage from the brain to the soul*. New York: Pantheon.

Tootell, R. B., Reppas, J. B., Dale, A. M., Look, R. B., Sereno, M. I., Malach, R., Brady, T. J., & Rosen, B. R. (1995). Visual motion aftereffect in human cortical area MT revealed by functional magnetic resonance imaging. *Nature*, 375, 139-141.