Two neural correlates of consciousness

Ned Block

Departments of Philosophy and Psychology, New York University, 100 Washington Square East, New York, NY 10003-6688, USA

Neuroscientists continue to search for ‘the’ neural correlate of consciousness (NCC). In this article, I argue that a framework in which there are at least two distinct NCCs is increasingly making more sense of empirical results than one in which there is a single NCC. I outline the distinction between phenomenal NCC and access NCC, and show how they can be distinguished by experimental approaches, in particular signal-detection theory approaches. Recent findings in cognitive neuroscience provide an empirical case for two different NCCs.

Introduction

I have previously proposed a conceptual distinction between phenomenal consciousness and access consciousness [1–3]. Phenomenally conscious content is what differs between experiences as of red and green, whereas access-conscious content is content information about which is ‘broadcast’ in the global workspace. Some have accepted the distinction but held that phenomenal consciousness and access consciousness coincide in the real world ([4,5] but see [6]). Others have accepted something in the vicinity of the conceptual distinction but argued that only access consciousness can be studied experimentally [7]. Others have denied the conceptual distinction itself [8]. This article argues that the framework of phenomenal consciousness and access consciousness helps to make sense of recent results in cognitive neuroscience; we see a glimmer of an empirical case for thinking that they correspond to different NCCs.

Phenomenal NCC

Christof Koch defines ‘the’ NCC as ‘the minimal set of neuronal events and mechanisms jointly sufficient for a specific conscious percept’ ([9] p. 16). However, since there is more than one concept of consciousness, this definition allows that a given percept may have more than one NCC. In my proposed framework, the Phenomenal NCC is the minimal neural basis of the content of an experience, that which differs between the experience as of red and the experience as of green.

I will start with an example: the neural basis of visual experiences as of motion is likely to be activation of a certain sort in area MT/V5. (Philosophers often use the terminology ‘as of’ motion instead of simply ‘of’ motion, since the experience can and does occur without motion.) The evidence includes:

- Activation of MT/V5 occurs during motion perception [10].
- Microstimulation to monkey MT while the monkey viewed moving dots influenced the monkey’s motion judgements, depending on the directionality of the cortical column stimulated [11].
- Bilateral damage to a region that is likely to include MT/V5 in humans causes akinetopsia, the inability to perceive and to have visual experiences as of motion [12,13].
- The motion after-effect – a moving afterimage – occurs when subjects adapt to a moving pattern and then look at a stationary pattern. These moving afterimages also activate MT/V5 [14].
- Transcranial magnetic stimulation (TMS) applied to MT/V5 disrupts these moving afterimages [15].
- MT/V5 is activated even when subjects view ‘implied motion’ in still photographs, for example, of a discus thrower in mid-throw [16].
- TMS applied to visual cortex in the right circumstances causes phosphenes – brief flashes of light and color [17]. When TMS is applied to MT/V5, it causes subjects to experience moving phosphenes [18].

Mere activation over a certain threshold in MT/V5 might not be enough for the experience as of motion; the activation probably has to be part of a feedback loop – what Lamme [19,20] calls recurrent processing. Pascual-Leone and Walsh [21] applied TMS to both MT/V5 and V1 (the first cortical destination for signals from the eyes) in human subjects, with the pulses placed so that the stationary phosphenes produced by the pulses to V1 and the moving phosphenes from pulses to MT/V5 overlapped in visual space. When the pulse to V1 was applied 5–45 ms later than that to MT/V5, all subjects said that their phosphenes were mostly stationary instead of moving (see [21] for references to single-cell recording in monkeys which comport with these results.) The delays are consonant with the time for feedback between MT/V5 and V1, which suggests that experiencing moving phosphenes depends not only on activation of MT/V5 but also on a recurrent feedback loop to V1 and back to MT/V5, [21].

So recurrent activity in and around MT/V5, in the context of other brain areas functioning normally – exactly which brain areas are required is unknown at present – is a good bet for being the physical basis of visual experience as of motion (but see [22,23] for some data that complicate this conclusion). Corresponding conclusions can be drawn for other types of contents of experience. For example, recurrent activation of the fusiform face area on the ventral surface of the temporal lobe (again in context) may determine experience as of a face [24]. The overall
Box 1. Area MT/V5 in a bottle?
The total Phenomenal NCC for the experience as of motion is a sufficient condition all by itself for the experience. What might that turn out to be? I suggest approaching the question by asking what we could remove from a normal brain and still have that experience. My suggestion is that we might be able to remove – at least – areas responsible for access to experiential contents and still have more or less the same experiential contents. Nakamura and Mishkin [48,49] removed frontal, parietal and superior temporal areas in one hemisphere of monkeys, leaving what is usually considered to be the visual system intact. They also disconnected visual inputs to the undamaged hemisphere. This preparation is sometimes said to cause blindness [13], but Nakamura and Mishkin are careful to say that this is shorthand for behavioral unresponsiveness to visual stimuli (at least temporarily), and should not be taken to show complete lack of visual sensation. One intriguing result is that when the limbic (emotional) system in the damaged hemisphere was left intact, the monkeys showed eye and head movements as if engaged in visual exploration. This contrasts with monkeys in which V1 is ablated who stare fixedly.

The conclusion is that there are different Phenomenal NCCs for different phenomenal contents (cf. Zeki on micro-consciousness [25,26]).

Of course, no one would take recurrent activation of MT/V5+ V1 all by itself in a bottle as sufficient for experience of motion (Box 1). A useful distinction here is that between a ‘core’ and a ‘total’ NCC [27,28]. The total NCC of a conscious state is – all by itself – sufficient for the state. The core NCC is the part of the total NCC that distinguishes one conscious state from another – the rest of the total NCC being considered as the enabling conditions for that conscious experience [9]. In these terms, then, the core Phenomenal NCC for the neural basis of the experience as of motion as opposed to the experience as of red or as of a face is likely to be recurrent activation of MT/V5 (see Figure 1).

Access NCC
We can distinguish between phenomenal contents of experience and access-conscious contents – contents information about which is made available to the brain’s ‘consumer’ systems: systems of memory, perceptual categorization, reasoning, planning, evaluation of alternatives, decision-making, voluntary direction of attention, and more generally, rational control of action. Wide availability motivates the idea that there is a ‘global workspace’ [29], and that information concerning conscious representations is ‘broadcast’ in this global workspace. The neural basis of information being sent to this global workspace can be called the ‘Access NCC’.

Rees et al. [13] note that in studies of the neural correlates of bistable perception, in which there are spontaneous fluctuations in conscious contents, reports of conscious contents correlate with activation in frontal and parietal areas. Dehaene and Changeux [7] suggest that a significant piece of the neural machinery of what they call ‘access to consciousness’ (roughly equivalent to my access-consciousness) is to be found in ‘workspace neurons’, which have long-range excitatory axons that allow, for example, visual areas in the back of the brain to communicate with frontal and parietal areas. Thus it is a good guess that the visual Access NCC, the neural basis of access, is activation of these frontal and parietal areas by occipital and inferior temporal areas (see Figure 2).

As Dehaene and his colleagues [7] have emphasized, there is a winner-take-all competition among representations to be broadcast in the global workspace. This point is crucial to the nature of the Access NCC and the difference between it and the Phenomenal NCC. One item of evidence for winner-take-all processes derives from the attentional blink paradigm, in which the subject is given a string of very brief visual stimuli, most of which are distractors. If there are two targets separated by an appropriate delay, the subject does not report seeing the second one, even though the second one would have been likely to be reported if the subject had not been given the first target. Dehaene et al. [30] used a modified attentional blink paradigm, in which subjects were asked to indicate on a continuous scale the visibility of the second target. The second target was at its peak of invisibility when the targets were separated by 260 ms. The result of interest here is that the subjects almost never used the intermediate cursor positions (at the 260 ms delay); that is, they rated the ‘blinded’ stimulus as either totally unseen or as totally seen almost all the time. Thus, Phenomenal NCC activations compete for dominating the Access NCC. Importantly, it is not the case that the Phenomenal NCC representation that is highest in initial activation will dominate, because domination can be the result of ‘biasing’ factors such as expectations or preferences [20,31].

Although the winning Phenomenal NCC will in general be amplified by the recurrent loop, a losing Phenomenal
Different colors indicate different studies (references in [13]). Activations cluster in superior parietal and dorsolateral prefrontal cortex as indicated by the large light circles. Activity in these frontal and parietal areas fluctuates spontaneously in binocular rivalry and other bistable perception in a way that is time-locked to fluctuation in reported experience. The core Access NCC may be activation of these areas by neural firing in the occipital cortex. Do we count the Phenomenal NCC as part of the Access NCC – in which case this figure pictures the Access NCC minus the Phenomenal NCC? Or do we regard the Access NCC as not including the Phenomenal NCC, in which case this figure pictures the Access NCC? This is a terminological issue: assuming that phenomenal consciousness is the gateway to full-blooded access consciousness, there can be no access consciousness without phenomenal consciousness.

NCC might itself involve recurrent loops to lower areas that will be sufficient for an experiential or phenomenal content. For example, an activation of area MT/V5 might have recurrent interactions with V1, making it the neural basis of an experiential content, but nonetheless lose in the winner-take-all competition and so not be accessed [20]. The general point is that the simplest and most explanatory theory may be one in which recurrent MT/V1 loops are sufficient for an experiential content despite not being accessible when they lose the winner-take-all competition. Thus, the winner-take-all process that is part of the nature of global broadcasting also strongly suggests that the Phenomenal NCC can be instantiated without the Access NCC, so global broadcasting does not encompass all of consciousness. This idea is further bolstered by evidence that there is brief parallel processing of many objects in the ventral visual stream (up to infero-temporal cortex) before zooming in on one or two of them [32].

**But is the phenomenal NCC really the neural basis of a kind of consciousness?**

You may ask, ‘If the Phenomenal NCC can perhaps occur without the Access NCC, how do we know that the Phenomenal NCC is really the neural basis of anything conscious?’ A quick answer is that, since the Phenomenal NCC determines the contents of experience, what it determines is ipso facto a kind of consciousness. The Phenomenal NCC for visual motion determines the experiential content of visual motion – as distinct from, say, the experiential content of seeing something as a face. That content itself is a kind of phenomenology, a kind of consciousness.

But this answer is too quick, since the doubt that motivates the question is a doubt that the Phenomenal NCC really does determine the contents of experience, and since the Phenomenal NCC was defined in terms of the contents of experience, the doubt challenges the evidence presented earlier for a Phenomenal NCC. The doubter may say that without access, there can be no true phenomenal contents but only proto-contents that become contents when globally broadcast. But how does the doubter claim to know that? Some are motivated by a terminological point – that we shouldn’t call something ‘phenomenal’ or ‘conscious’ if it isn’t broadcast for access [24]. The substantive empirical question is: if our evidence always concerns phenomenal contents that are accessed, how can the Phenomenal and Access NCC ever be empirically distinguished?

The answer is that it is not true that our evidence always concerns experiential contents that are accessed. There are a variety of paradigms in which we can use convergent evidence involving varying degrees of access to try to separate the Phenomenal from Access NCC. One such paradigm is signal detection theory.

**Signal detection theory (SDT) approaches**

Suppose a subject is shown a series of stimuli at around threshold level and asked to press one button if a target is seen and another if not. SDT models the subject’s behavior in terms of two factors: the extent to which the subject has an experience of seeing it and the criterion the subject implicitly sets for reporting seeing it. The criterion is famously influenceable by features of the experimental setup that affect the subject’s expectations or motivation – such as the proportion of ‘catch trials’ (where no stimulus is presented) and by rewards for hits and penalties for false alarms. We know from standard SDT analyses that the subject’s reports of whether there was a target or whether he saw it do not only reflect the extent to which the subject did see it, but also the subject’s threshold for reporting. Two experimental setups in which there are the same experiential contents may result in different reports.

A dramatic example is a series of experiments concerning the ‘exclusion’ paradigm [33], in which subjects are instructed to complete a word stem with something other than the end of a masked word just presented to them. If the word ‘reason’ is presented ‘unconsciously’ (for 50 ms), the subject is more likely than chance to choose some other ending (e.g. ‘reader’). This paradigm has impressed many because it appears to yield opposite results for unconscious and conscious stimuli. However, Visser and Merikle [34] showed that changing the motivation of subjects by using a reward structure can change the degree of exclusion. They started subjects with a $15 credit and docked them $1 for each error. Visser and Merikle interpret their result in terms of the effect of reward vs. punishment on increased attention, accepting the idea
that the 50 ms/250 ms difference engenders an unconscious/conscious difference.

But there is an alternative – an SDT interpretation suggested by Snodgrass [35] in which the results in part reflect a criterion shift rather than a difference in consciousness. The idea is that punishment for errors of failing to exclude pushes the criterion for inhibiting the natural response so low that weak conscious perception of ‘reason’ blocks use of ‘-son’ even though the subjects are so unconfident that they think they don’t see the word. That is, their criterion for belief is higher than their criterion for inhibiting a response. The subject’s state of mind when successfully excluding one of the 50 ms stimuli could be articulated – over-articulated, no doubt – as “I probably didn’t see a word but if I did, it was ‘reason’, so I’d better complete the stem with ‘reader’.” [36]. And the SDT interpretation is confirmed by the effect on ‘inclusion’ instructions. With ‘inclusion’ instructions, the subject sees ‘reason’ and then is given ‘rea_’ but is told to complete the stem with the word he saw if possible. In this paradigm, SDT predicts no shift with change in reward or punishment, because there is no issue of a criterion: the subject just uses the first word that comes to mind regardless of level of confidence that it is the word he saw. And the result [34] is just that; the difference in reward/punishment structure makes no difference in the result under ‘inclusion’ instructions.

There is therefore evidence in the ‘exclusion’ case of experiential contents (e.g., as of seeing ‘reason’) without the kind of access required for report, planning, decision-making, evaluation of alternatives, memory and voluntary direction of attention. Some of the 50 ms stimuli are weakly conscious although not broadcast in the global workspace. Thus, SDT gives us reason to think that experiential content – based on the Phenomenal NCC – can be instantiated without the kind of access that is based in the Access NCC.

Neural SDT

In a landmark series of experiments, Super et al. [37] recorded from V1 during a task in which monkeys were rewarded for saccading to a target if there was one or continuing to look at the fixation point if not. Super et al. manipulated whether the locations in V1 corresponded to figure or ground. When the monkey detected the target, there was an increased V1 response for figure as compared with ground (see Figure 3, in which this increased figure response is referred to as ‘modulation’).

Super et al. were able to manipulate the modulation by varying the saliency of the stimulus (i.e., the number of pixels in line segments in the target; Figure 3b) and the proportion of ‘catch trials’ in which there was no target. For high saliency stimuli and small numbers of catch trials, there was a near perfect correlation between modulation and saccades to the target, and in that sense modulation and access to the target corresponded well. But moving the saliency down or the percentage of catch trials up boosted the modulation when the animal did not saccade to the target to the 50% range. That is, with low saliency or a high number of catch trials, the decision criterion was close enough to the visual ‘signal’ that the modulation averaged the same whether the animal saccaded to the target or not. For example, this happened when the pixel count was reduced from 16 to 4, maintaining catch trials at 20%, and also when the pixel count was 16 and the catch trials were raised to 50%. If the pixel count was reduced to 4 but the catch trial percentage was also reduced to zero, then the correlation between modulation and access was restored. These results show that the modulation does not reflect access to the target (since in the low saliency condition it was the same whether the target was or was not accessed). Nor does the modulation reflect the saccade, so it is on the sensory rather than motor side of the decision process. Nor does it reflect attention, since the detected targets can be assumed to draw more attention. The modulation therefore seems to reflect something intermediate between the stimulus and access. In a classic signal detection analysis, Super et al. showed that the modulation is indeed an intermediate level representation that can be disconnected from access either by raising the perceptual decision criterion or by decreasing saliency of the stimulus, lowering the visual ‘signal’ to the range of the decision criterion.

The modulation seen by Super et al. disappears under anesthesia [38] and is probably produced by recurrent processes [39], unlike other V1 representations like direction and orientation tuning. So there is some plausibility to taking it as an indication of, if not directly part of, a Phenomenal NCC for the experiential content of seeing the target (see also [40]).

Can the phenomenal NCC be studied empirically?

Doubts about whether phenomenal consciousness (and hence its neural basis, the Phenomenal NCC) can be studied empirically are common (see also Box 2), and often based on the idea that ultimately, introspective reports, that is, reports about one’s conscious experience, are the fundamental epistemological basis of theories of consciousness, the ‘gold standard’. [7,41,42]. Reports are not supposed to be infallible, but any discounting of reports as reporting too much or too little, will supposedly have to be based solely on other reports. Reports inevitably reflect the Access NCC, not just the Phenomenal NCC: when people tell you about their conscious states, you only hear about the ones that have won the winner-take-all competition. Hence we can only study ‘access to consciousness’ [7], that is, access to experiential content, not experiential content itself. I do not agree with this methodological view for several reasons.

First, observed electrons can provide evidence about electrons that cannot in principle be observed, for example electrons that are too distant in space and time (e.g. outside our light cone) to be observed. Why should we suppose matters are any different for consciousness?

Second, there is no gold standard of evidence, here or in any area of science. We should go for the simplest theory compatible with all the evidence. No evidence is privileged. In particular, it is not true that our theory of consciousness should be completely determined by the introspective reports of subjects. As an analogy, it is trivial to program two computers to yield the same input–output
function via different algorithms. No theory of what goes on in computers based wholly on the computers’ ‘reports’, that is input–output relations, stands a chance of success. Why should we suppose consciousness is any different?

Third, any neuroscientific approach that bases everything on reports about a subject’s own experience will end up finding only the neural basis of ‘higher order thought’ – thought to the effect that I myself have an experience – rather than the neural basis of conscious content or even access to conscious content. To give an introspective report, the subject has to have a higher order thought – so to insist on introspective reportability as the gold standard is to encourage leaving out cases in which subjects have experiences without higher order thoughts.

Finally, even those who assimilate experiential content to its accessibility should not accept introspective reports as a gold standard. Animals have plenty of access to their experiences, but probably little in the way of higher order thought about them of the sort that could be the basis of an introspective report. Cowey and Stoerig [43] showed that monkeys that had been made blindsighted on one side and trained to make a visual discrimination in their sighted field, could make the discrimination in their blind field. However, when given the option, they preferred a third ‘nothing’ response. This is evidence about the monkeys’ perceptual state that does not depend on any introspective reports.

But is the monkeys’ button-pushing just a non-verbal introspective report? Non-human primates that have
learned symbolic systems for communication may not even make spontaneous reports about the world [44,45], so there is little ground for supposing that they are prone to give reports about their own experience. If a human were to push the ‘nothing’ button, we might guess whether there is a thought underlying the response. We might consider two hypotheses: first, the introspective report, ‘I am having no visual experience’ and second, the environmental report, ‘There is nothing on the screen’. If the subject were a child of 3–4 yrs, the introspective report would be unlikely since children have a great deal of difficulty with states of mind about their own mental states [46,47]. Given that the environmental report would be preferable for a child, we can hardly suppose the introspective report would be preferable in the case of a macaque! The take-home message is that you don’t need reports about the subject’s experiences to get good evidence about what the subject is experiencing: indications of what the subject takes to be in front of him will do just fine.

Conclusion

Where are we? I have proposed a distinction between a Phenomenal NCC and an Access NCC. The ‘single NCC’ framework does not do as well in making sense of the empirical data, in particular, signal detection theory data, as an account in which there are two NCCs. Of course both these NCCs are to be firmly distinguished from perceptual representations that are not conscious in any sense (as in the rightmost panel of Figure 3c). More generally, rather than asking ‘What is the direct evidence about the Phenomenal NCC independently of the Access NCC?’ we should instead ask ‘What framework makes the most sense of the data?’

References


