Comments on Joseph Neisser's "Correlates and Causes"
Consciousness Online – The online consciousness conference.
February 2010

Jakob Hohwy
Philosophy, Monash University

Introductory remarks

Neisser's very interesting paper goes to the heart of the debate about the neural correlates of consciousness (NCC). It does this in what I think is a good way, namely by doing some philosophy of (consciousness) science. The flow of the argument is that Chalmers' original formulation of the NCC notion is flawed for reasons having to do with the first-person perspective, which is intrinsic to conscious content. This should turn us towards the NCC for creature rather than just content consciousness. And further, this points to a causality-inspired conception of the content NCC, which is in fact capturing what actually goes on in NCC research. The NCC should stand for the Neural *Causes* of Consciousness.

I agree with much in Neisser's conclusions, in particular that a mechanistic approach to the NCC is desirable. But (a) I think this view can be arrived at without undermining the standard notion of the NCC for conscious content. (b) Whereas I agree that the NCC for creature consciousness is necessary for a satisfactory notion of the neural correlates I think it alone is as unsatisfactory as the NCC for content consciousness is on its own. (c) The NCCs for creature and content consciousness together make up the causal mechanism underlying consciousness; none of them is the cause of consciousness.

I discuss (a)-(c) below. In the interest of minimising reading burden I've set out the core of my view in a box at the beginning of each section. Details can then be found in the further text.

Stage-setting and program

At a central point in the argument Neisser says "[T]he first-person perspective is partly constitutive of content consciousness itself. Thus, we cannot just stipulate, as Chalmers does, that we are studying content consciousness rather than creature consciousness since the latter is an internal structure of the former" (6).

There is much context for this quote (for details please consult the paper itself) - I'll focus on three interesting topics that grow out of it:

(a) Does the first-person perspective block content-matching?
(b) Content NCC vs. Creature NCC
(c) Correlates vs. causation

(a) Does the first-person perspective block content-matching?

- The standard content NCC is meant to identify the minimally sufficient neural conditions, in the form of a neural representational system, for the occurrence of a conscious content.
- But neural representational systems cannot be "intrinsically experiential", so cannot match conscious content.
- I have trouble understanding the notion of content being "intrinsically experiential". Since the notion is most clearly fleshed out in terms of "first-person perspective", I assume that it means that conscious content is not exclusively invariant (i.e., it must be perspective-dependent).
- It is true that some NCC-candidates are invariant, but I set out how it may be possible to identify perspective-dependent neural representational systems, which can match intrinsically first-person perspective conscious content.
- Conscious content's being "intrinsically experiential" does not undermine the content-matching notion of the NCC, which remains on the table for the NCC-search.

Neisser appeals to the idea, promoted by Noë and Thompson, that consciousness is "intrinsically experiential" in making the point that the first-person perspective is partially constitutive of content consciousness.

I have long been puzzled by this notion of being "intrinsically experiential" (see Hohwy and Frith 2004). At a first and slightly superficial glance it is hard to see why anyone would find it controversial at all that conscious experience is intrinsically experiential. There must be something more substantial in play. As far as I can see, it is the idea that being experiential is having a first person perspective (and this is one way Neisser fleshes out the notion). This is also something few would deny but it is indeed something that can easily be overlooked, once the focus is squarely on representational content.[1] Why, then, is the first-person perspective thought to be a problem for the content NCC approach?

Neisser agrees with Noë and Thompson that there is no prospect of matching first-person perspectival conscious content with any neural representational system, such that the match respects the requirement of the latter being minimally sufficient for the former. Here is how I understand the argument. Neural representational states are picked out by virtue of content individuation which is not sensitive to the first-person vs. third-person distinction (perhaps because the receptive fields of neurons don't have first-person perspectives ?). Therefore it cannot match this intrinsic element of the conscious content.

This is a nice challenge so I'll first try to chart a route on which neural content individuation is in fact sensitive to first-person perspective, such that there could be this kind of match. I am going to assume that for content to lack first-person perspective is for it to be invariant (my hope is that this assumption is safe, invariance

[1] It is of course possible that something even more substantial than this is intended with the notion but I think that a thicker notion would quickly raise the spectre of something like the hard problem, which everyone in this debate wants to avoid

here just means perspective-independence). If the content NCC can *only* pick out neural representational systems exclusively devoted to invariant content, then the Neisser-Noë-Thompson line looks good. The single cell findings in MTL during binocular rivalry indeed suggests invariance (especially when we add findings by Quiroga et al (Quiroga, Reddy et al. 2005) that neurons in this area respond in the same way to the same specific objects even when they are presented in different perspectives, or modes of presentation).

But I don't think this is the end of the story for a good case can be made that the content NCC must include more than just neurons that respond invariantly to specific stimuli. In contrast to the single cell studies, fMRI during rivalry reveals a great deal of activity in earlier parts of the visual stream, including V1 and the lateral geniculate nucleus. It seems very likely that neural activity at these earlier stages encode more perspective-dependent sensory attributes (e.g., how this face looks to me right now, rather than the perceptual inference that irrespective of mode of presentation the face is Mother Theresa's). Including these parts of the activation pattern in the core content NCC would thus make it sensitive to perspective-dependence, to the creatures specific perspective on the world.[2] If this is right then the standard content NCC *can* match conscious, intrinsically perspective-dependent content, in the sense that its activation would be minimally sufficient for the conscious experience. This would then obstruct the Neisser-Noë-Thompson line.

Here is what I think fuels this debate. Most studies of rivalry are not sensitive to the invariance vs. perspective-dependence contrast. For example, Logothetis' monkeys were not systematically exposed to sunbursts seen from different perspectives. If we assume, nevertheless, that the target property of this kind of study is perspective-dependent content, then it is understandable that findings of activity that mainly encode invariant content seem irrelevant. However, this doesn't exclude the possibility of rivalry studies that *are* sensitive to the contrast and which could allow us to pick out perspective-dependent neural representational systems. Briefly, we should look for minimally sufficient neural activity that does vary with perspective. Would this kind of content NCC meet the Neisser-Noë-Thompson challenge? If yes, then the content-matching idea seems safe. If no, then what is the notion of the 'first-person perspective' that is in play?

Perhaps my point can be put by appealing to different versions of the standard NCC definition. First-person perspective can be captured if we look for neural *perspective-dependent* representational systems rather than just neural *invariant* representational systems. I don't think acknowledging this distinction really undermines Chalmers' original definition. This differentiation of the definition would be in accordance with Chalmers' idea that the focus on neural representational systems is just one way of systematising the NCC search, that is, just one way, of many possible ones, of going beyond a search for "brute" correlates of any kind of brain activity (Chalmers 2000: Sec. 2(iv)).

Here is a related thing that I would like to put on the table for discussion. Just as we should require the content NCC to be sensitive to first-person perspective we should

[2] There is a theoretical framework for thinking about binocular rivalry that goes with this idea, the gist of which is suggested in Hohwy, J., A. Roepstorff and K. Friston (2008). Predictive coding explains binocular rivalry: An epistemological review. *Cognition* 108(3): 687-701.

require that it be sensitive to invariance. That is, it seems also to be true that conscious experience is *intrinsically invariant*: my experience of Mother Theresa is steeped in but also rises above the particular perspective with which she is represented to me. For this reason I think that the content NCC should combine the two versions of the definition set out above.

Perhaps the idea of conscious content being intrinsically experiential is not well captured by the contrast between invariance and perspective-dependence. As I said, I find it hard to think of a clear alternative. But, perhaps, it is better captured by the notion of pre-reflective self-awareness ("mineness") (Zahavi 2006). Again, this is a notion that is very hard to make explicit. I think it is plausibly connected to the idea of having a first-person perspective or, rather, that it is a natural upshot of activity in the neural perspective-dependent representational systems that I have just discussed (I present such a case in (Hohwy 2007)). Again, this would sit well with the matching-content doctrine part of the NCC definition.

(b) Content NCC vs. Creature NCC

- Neisser argues that the search for the NCC of conscious contentcannot be separated from the notion of creature consciousness because the latter is an internal structure of the former, and is what has any hope of capturing the "intrinsically experiential".
- I think the search for the content NCC is methodologically geared for finding the neural correlates of *selection* of content to consciousness (this has been argued by Searle (2000) and defended by me (2009).
- So, as a methodological point, it is unlikely that the content NCC can stand on its own as the neural correlate of what it is for a content to be conscious.
- This point can be made without appealing to content being "intrinsically experiential. It is just a general matter of what properties are presupposed to be present in these kinds of study design.
- There seems no specific ground for thinking that the NCC for creature consciousness alone will help with notions like first-person perspective: it is just the neural correlate of being conscious rather than unconscious.

Neisser associates respecting the notion of content being "intrinsically experiential" with moving from content consciousness towards incorporating creature consciousness ("[we] cannot just stipulate […] that we are studying content consciousness rather than creature consciousness since the latter is an internal structure of the former" (6)).

I would like to resist this association somewhat. It is true that it is creatures who have first-person perspectives, so I can see the reasoning. But it seems to me a mere appeal to finding the *creature NCC* is not enough. The creature NCC is the neural system that is minimally sufficient for a creature being conscious rather than unconscious. As such it doesn't immediately engage the background idea that creatures are things that have first-person perspectives. Perhaps being creature conscious is having something like a unified conscious field but it seems to me this only begins to attain a first-person perspective when populated (however that happens) with specific contents, which are the kinds of things that can be perspective-dependent. Something else, or

more, than just being conscious rather than unconscious is required to deal with the intrinsically experiential. The question is: what would it be to study creature consciousness, in particular what would it be to study it as an internal structure of content consciousness?

This connects to a point Neisser makes about studies of binocular rivalry. He points out that in this kind of experimental design participants are already conscious when put in the scanner and that therefore "[i]n the binocular rivalry situation, there is plenty of conscious experience going on beyond the changes measured." (7). This point has been pressed by Searle (Searle 2000; Searle 2004) and I have defended it a length (Hohwy 2009) but not on grounds having to do with Noë and Thompson's notion of content being intrinsically experiential. It is mainly an observation about the methodologies in use in this kind of NCC search. As a methodological point I think it does invite us to look more at creature consciousness but not because there is anything inherently wrong with the content NCC search or the definition of it.

The point is that, taken on its own, the content NCC has more to do with whether this rather than that content (whether invariant or perspective-dependent) gets *selected* for consciousness, rather than with what it takes for those contents to *be* conscious rather than unconscious. Perhaps this does imply some kind of refinement of the NCC-definition, namely that the neural representational system should be minimally sufficient for *the selection* of the state of consciousness (rather than for the state *being* conscious). This is not saying that the NCC-definition is flawed, just that we need to refine our conception of what the target property is. And to repeat, I don't think there is a strong case that first-person perspective prevents finding neural systems which are minimally sufficient for matching all the aspects of content that get selected for consciousness.

What I am trying to gesture at here is that I agree there is a very important issue concerning the creature vs. content NCC but that this issue can be brought out without worrying about the matching content idea or appealing to content being intrinsically experiential. The issue is this: a legitimate part of the NCC search is finding the *content NCC* (i.e., the neural systems that are minimally sufficient for selecting this rather than that content into consciousness in a creature who is already conscious). This clearly leaves out another seemingly legitimate part of the NCC search, namely finding the *creature NCC* (i.e., the neural systems that are minimally sufficient for a selected content's being conscious rather than unconscious, in a creature able to select contents for consciousness). I think (and discuss below) that *both* NCCs are needed to arrive at the neural correlates of consciousness.

(c) Correlates vs. causation

- Having sidelined Chalmers' standard definition of the NCC, Neisser interprets NCC research (binocular rivalry studies), in the causal terms of Mackie's *inus* condition, and argues that this NCC research thus reveals the neural *causes* of consciousness.
- There may be some tension in this view. On the one hand, the neural causes of consciousness are *distinct* from their effects, conscious states (of which I assume a neural correlate can be found). On the other hand, creature consciousness is "an internal structure" of content consciousness, so *distinctness* is not obvious.
- Neisser argues something like creature consciousness is an "independent" "baseline" against which these neural causes can be identified (e.g., during binocular rivalry studies). I think it is likely there is no such independence since content cannot be purely inserted against a stable, independent baseline. Instead it is likely that inserting rivalry affects the causal workings of the baseline.
- This means the content and creature NCCs interact and that both are parts of the neural correlates of consciousness.
- This combined neural correlate is, I agree, a causal mechanism but is not itself a cause of consciousness.

- *In sum*: We should keep the standard definition of the content NCC (but use it to look for perspective-dependent neural representational systems). This NCC should then be combined with the NCC for creature consciousness and their causal interaction explored.

Neisser proposes that we swap from Chalmers' proposed NCC definition to an understanding of NCC research in terms of Mackie's *inus* conditions, and more broadly to more mechanistic terms. I have a lot of sympathy for this kind of move, in particular aligning NCC with the notion of mechanism (e.g., Hohwy 2009: 435), so it is good to see it worked out in more concrete detail.

Neisser suggests that, with the *inus* formulation we should understand 'NCC' as short for 'Neural *Causes* of Consciousness'. There are two ways of taking this proposal.

The *first* way of understanding the causality proposal is far-reaching in that it says that the neural activity pinned down in rivalry studies causes the conscious state in question, with the implication that neural and conscious states are distinct existences. I find this unattractive because I think a core motivation for the NCC search has got to be to find the neural 'underpinnings' of consciousness rather than the causes of consciousness. Naturally, there are always causes of consciousness, for example, the photo of Mother Theresa causes in me a conscious experience of it. But it always has to be possible to ask, of that conscious experience itself, what its neural underpinnings are.[3]

[3] There is a metaphysical side to this too: with this understanding of the causality proposal it appears the NCC definition rules out using NCC evidence to support belief in the identity theory of mind (since if it is only causation, then it is not identity).

In general, the relation between a causal mechanism and the phenomenon it gives rise to is not a causal one. It is more like a constitutive relation, or a relation that can be fleshed out in terms of levels, or, what I prefer, a realisation relation (along the dispositional lines of mechanistic categorical bases and the dispositional manifestations they give rise to). Though there has been great progress in figuring out the notion of a mechanism (see works by Bechtel, Craver, Wimsatt and also Woodward) I don't think this issue has been resolved yet.[4] The point here is that these non-causal ways of fleshing out the relation do allow us to ask the questions about the neural underpinning of conscious states.

Now I don't think Neisser's suggestion is that the relation between the neural and the conscious is causal (I think his suggestion is the second one, discussed below). But I want to suggest that the *inus* framework is in fact a good initial way of thinking of this relation precisely because it (as Neisser also observes) is not exclusively causal. It is rather a fairly pragmatic principle for how we pick out difference makers in experimental contexts (choice of background conditions is important, as is the insight that difference-makers are not outright necessary conditions). Only here the difference and the difference-maker stand to each other as different levels, not as cause and effect.

Seen in this light I think that, in fact, Chalmers' definition is very apt and close in spirit to the *inus* story. Chalmers' definition is also pragmatic, and in much the same way as Mackie's – this is reflected in the all important notions of *C*-conditions and of the sufficient conditions being *minimal* (I think Neisser agrees, see his fn10). There is a nice project in figuring out precisely how the two definitions differ, and how, if they do not differ, the *inus* formulation but not the Chalmers formulation is apt for what goes on in rivalry studies.

Leaving aside the best way to characterise the relation between the neural and phenomenal levels, I think it is right that we need more focus on the causal mechanics at the neural level. As Neisser rightly says "the [NCC] research aims at localizing the working parts of an underlying mechanism, and in this respect it is continuous with much of the rest of neuroscience as well as the long tradition of psychophysical research" (11). (What I took issue with in the preceding couple of paragraphs was the notion of causation by 'underlying' causes; what I agree with now is the 'working parts of the mechanism' story).

This leads to the *second* understanding of the proposal to understand NCC as the Neural *Causes* of Consciousness – the one I think Neisser has in mind. The reasoning is like this: If the content NCC search (in the form of rivalry studies, understood in *inus* terms) successfully picks out anything it is neural activity that *causes* the neural underpinnings of the pre-existing unified conscious field to be modulated in a certain way. This concerns causal processes at the neural level and to capture it I suggest a refinement of Neisser's causal reading of the NCC such that rather than Neural Causes of Consciousness it should be CMONCC: Causal Mechanisms Of the Neural Correlates of Consciousness. With that we can keep distinct the ideas that the neural

[4] In Neisser's discussion of this issue the notion of 'causation' drops out and instead there is use of notions like 'constitution', 'exhibition', 'level', 'is' (of identity), 'enacted'; (see p 8-9).

underpins the conscious, and that the NCC search investigates the causal mechanisms of the neural level.

To me, this helps bring an important issue into focus. Let's accept that the rivalry studies track the neural causes of selection of a content for the unified conscious field. Then we ask, are those neural causes and the neural underpinnings of the unified conscious field *independent* of each other? That is, are the neural systems picked out by the rivalry studies part of the underpinnings of consciousness or are they merely causal modulators of the independent neural correlates?

On the one hand, we should expect them to be interdependent and intermixed – as Neisser says "[creature consciousness] is an internal structure of [content consciousness]" (6).

But on the other hand, the Noë-Thompson induced scepticism about the relevance of the content NCC, and the conception of the methodological issues in the NCC rivalry studies (discussed in (b) above) that go with it, indicate that there is a great deal of independence. Neisser (citing Gloor on Berger) says about the general methodology in psychophysics and brain imaging, and in the rivalry studies in particular "[The] experimental condition differs from the baseline condition only with respect to a well defined psychological function. In this way, those changes specifically related to that function can be isolated from those it shares with the baseline condition" (7). Neisser applies this to rivalry studies: "In the experiments conducted by Logothetis et al, the subjectivity (or 'experientiality') of visual consciousness is clearly *baseline*." (7), and makes the Searlian point, noted above, that "in all the experimental conditions, [the] subjective point of view obtains, and is *independent* of the binocular rivalry" (7; my emphases).

I think there is much that is right here but that things are not as clearly *independent* as presented. In particular I suspect that the neural activity that underlies the baseline of being creature conscious is not independent of the neural activity that underlies the rivalry, and vice versa. It may be, that is, that the claim of independence, assumed in general by Gloor/Berger and applied by Neisser to the rivalry studies, relies on an assumption of "pure insertion" that may not hold for the neural mechanisms underpinning consciousness: thus it may be that in reality there are many causal interactions between different kinds of content NCCs, different parts of the creature NCC and other background conditions. Pure insertion is "[t]he idea […] that as a new cognitive (A) component is added to a task, the implementation of the preexisting components (e.g., B) remains unaffected." (Friston, Price et al. 1996: 98). I think it is very likely that A (=content NCC activation during rivalry) affects the effects of B (=creature NCC for the subjective point of view). If this is right then we need both NCCs, harnessed in factorial designs, to get at a useful NCC.

This "NCC-interactionism" would then be a way of respecting the insight that content and creature consciousness are intermixed while still operating with a notion of causal mechanisms underpinning consciousness. Furthermore, the neural mechanisms tracked in the rivalry studies would be part of the neural correlates of consciousness. Finally, if my line of thought under (a) above is right, then what is tracked in the rivalry studies could well be neural representational systems that match content, including the first-person perspective. What is missing from the story told in content-

NCC terms is the causal interactions with another factor, namely the neural systems minimally sufficient for the creature being conscious rather than unconscious.

Summing up

I agree with much of what Neisser proposes. In particular, that the neural correlates of conscousness should be investigated on the assumption that it is a causal mechanism and that the traditional understanding of the NCC is some ways fall short of capturing consciousness. But I prefer different routes to some of these things. (a) I think the content-matching doctrine is still very viable, at least on my preferred understanding of content being "intrinsically experiential". (b) I don't think we need to appeal to content being intrinsically experiential, and to the matching content doctrine being flawed, in order to see the limitations of the dominant content NCC approach and the worth of exploring the creature NCC. And (c) I think we have to be careful in distinguishing the non-causal relation between the neural and the conscious levels from the causal mechanical relations holding between parts of the neural level underpinning the conscious. Concerning those causal mechanics, I don't think the creature NCC is independent of the instantiation of the content NCC. We cannot understand the "intrinsically experiential" without understanding the content NCC.

Chalmers, D. (2000). What is a neural correlate of consciousness? *Neural Correlates of Consciousness: Empirical and Conceptual Issues*. T. Metzinger. Cambridge, Mass., MIT Press.

Friston, K. J., C. J. Price, P. Fletcher, C. Moore, R. S. J. Frackowiak and R. J. Dolan (1996). The Trouble with Cognitive Subtraction. *NeuroImage* 4(2): 97-104.

Hohwy, J. (2007). The sense of self in the phenomenology of agency and perception. *Psyche* 13(1).

Hohwy, J. (2009). The neural correlates of consciousness: New experimental approaches needed? *Consciousness and Cognition* 18(2): 428-438.

Hohwy, J. and C. Frith (2004). The Neural Correlates of Consciousness: Room for improvement, but on the right track. *Journal of Consciousness Studies*.

Hohwy, J., A. Roepstorff and K. Friston (2008). Predictive coding explains binocular rivalry: An epistemological review. *Cognition* 108(3): 687-701.

Quiroga, R. Q., L. Reddy, G. Kreiman, C. Koch and I. Fried (2005). Invariant visual representation by single neurons in the human brain. *Nature* 435(7045): 1102-1107.

Searle, J. R. (2000). Consciousness. *Annual Review of Neuroscience* 23(1): 557-578.

Searle, J. R. (2004). *Mind – A Brief Introduction*. Oxford, Oxford University Press.

Zahavi, D. (2006). *Subjectivity and Selfhood: Investigating the First-Person Perspective*. Cambridge, Mass., MIT Press.