

## **Chapter 11**

### **The Neural Correlates of Consciousness:**

#### **Room for improvement, but on the right track.**

#### **Response to Noë and Thompson**

(Originally published in *Journal of Consciousness Studies* 11(1): 45–51, 2004)

**(with Chris Frith)**

#### *Preview*

In their target article, Noë and Thompson argue that the current approach to identifying the neural correlates of consciousness (NCC) does not, and could not, be successful. Though N&T present interesting observations about the role of the sensorimotor system for consciousness, their pessimistic conclusions about the NCC-approach do not follow: the core distinction in the argument is flawed; neuroscience does not aim at the kind of explanation presupposed by their argument; results from psychophysics support rather than contradict the internalistic nature of the NCC-approach; the experiments on binocular rivalry that N&T criticise in fact integrate very well with other NCC-findings.

## **Chapter 11**

### **The Neural Correlates of Consciousness:**

**Room for improvement, but on the right track.**

**Response to Noë and Thompson**

**(with Chris Frith)**

#### **1. Introduction**

In general, people are pessimistic about finding the neural explanation of consciousness, and relatively more optimistic about finding the neural correlates of consciousness (NCC), and much important work has indeed been done in identifying the NCC. The general atmosphere seems to be that if only we can identify the NCC, we will at least have a possibility of progressing to an explanation of consciousness, though at present we have precious little idea about how that explanation would look.

Noë and Thompson (N&T) offer a welcome respite from the pessimism about the neural explanation of consciousness, and an opportunity to instead evaluate the methods and presuppositions of the current search for the NCC. This is important because, if the search for the NCC is misguided, then its results are not going to be helpful in construing an explanation of consciousness.

N&T argue that the current NCC-approach *in fact* does not establish the existence of NCCs for the contents of conscious states, and that it *could not* establish the existence of content NCCs. They advocate an alternative articulation of the relation between consciousness and the brain where consciousness is not conceived of as ‘an internal occurrence of the mind-brain, but rather as a complex set of capacities of embodied and situated agents’ (22).

Some of their premises rest on interesting and important insights about consciousness, in particular about the role of the sensorimotor system in consciousness, but we shall argue that the pessimistic conclusions about the NCC-approach do not follow.

## **2. The core distinction.**

N&T argue against the idea that the first task of neuroscience is the NCC project of uncovering the neural representational systems whose contents systematically match the contents of consciousness – that is, they argue against the ‘matching-content doctrine’ (1-2). Here they focus especially on studies of binocular rivalry.

The two arguments against the matching-content doctrine rely on a distinction concerning the notion of ‘matching content’. A neural structure may *match* a conscious content, or it may merely *agree* with a conscious content in some respect. N&T illustrate the distinction with an example (11-12): a photograph of some birds flying across a sunny blue sky and a verbal report that some birds are flying across a sunny blue sky *agree* in content, but they don’t *match* in content because the verbal report does not say anything about the shade of blue or the position of the birds. They elaborate this by saying that when two contents match, then they are the same, whereas when they agree there is only a strong correlation (13).

The two arguments that follow try to establish that binocular rivalry studies only provide agreement in content, not match. Therefore something (some crucial aspect of content) does currently, and must, elude NCC research fashioned according to the matching-content doctrine.

But the distinction on which these arguments rests seems problematic. To begin with, it is hard to see why the NCC-approach must establish exact match of content. The requirement that a correlation approach must establish such identity seem too stringent, and does not appear to be a requirement in any other area of scientific explanation. Clearly, the better the correlation, the better the chances of a scientific explanation, but this doesn't imply that an as yet incomplete correlation undermines the validity of the entire approach.

The requirement is also puzzling at somewhat deeper level. When we correlate two sets of properties of things with a view to explaining the constitution of one in terms of the other, we do not require that there is identity in the sense that the thing that is doing the explaining *has* the properties of the thing that is getting explained. For example, when we explain the states of matter (e.g., some water having the property of being frozen) in terms of certain bonds between atoms and molecules (e.g., the particular lattice formation of H<sub>2</sub>O molecules when kinetic energy decreases) we do not want to say that the bonds between atoms and molecules have the property of being in the state in question (e.g., we do not want to say that the lattices of H<sub>2</sub>O assemblies have the property of being frozen). Similarly, we should not require that the neural substrates of conscious states *have* the properties of the conscious states they constitute.

Furthermore, the example that is meant to illustrate the distinction doesn't make it clear whether the distinction is qualitative, or only a matter of degree. It seems one could simply expand the verbal report by adding detail about colours and positions of the birds, and so on, and thereby come to match the content of the photograph better. It seems an analogous move is available to the matching-content NCC strategy: do more studies, add more correlates, and arrive at better matches of content. This possibility is not ruled out by the way the distinction is set out. Put another way, in what way would a complete agreement of all aspects fall short of being a match?

The characterisation of matching in terms of identity has some further problems. The contrast between photographs and verbal reports suggest that identity means 'all of the very same content in the very same representational format' (only two identical photos, or two identical verbal reports, can match in content). This would make most kinds of correlational studies impossible. For example, it would be impossible to say that a speech and transcription of the speech match in content. On the other hand, if it means 'all of the very same content, never mind the representational format', then photos and long sentences can match in content. But this leaves it a possibility that the photo and the sentence may have a partial match in content—they may be identical in certain respects—and the corresponding possibility, with respect to binocular rivalry studies, would be enough to ensure that these studies are not misguided, only incomplete. The point here is that as long as the distinction is not clear, the arguments that rest on it will not be convincing.

The problems we have highlighted suggest that N&T's conceptualisation of the NCC-approach may be flawed. It is true that the NCC-approach is committed to identifying the neural correlates of conscious content, that is, identifying neural representational systems such that representation of a content in the system is sufficient, under conditions *C*, for representation of that content in

consciousness (6). It is also true that this requires some kind of isomorphism between the neural representation and the conscious representation (4). N&T sum this up by saying that the NCC-approach is committed to the *matching*-content doctrine (in their sense of ‘match’ as identity) (6). But sufficiency and isomorphism doesn’t entail identity in any stronger sense than complete agreement (or correlation) of all aspects of content, and that is something the defender of the NCC-approach can live with.

### **3. The argument concerning empirical evidence.**

N&T provide the following argument for their claim that the neural representational systems identified in binocular rivalry studies fail to match conscious content. Activity in neural representational systems just covaries with the presence in the receptive field of certain stimuli, such as a vertical line. Perceptual experience is more than that. For example, it is the representation of the line ‘as against a background, and as occupying a certain position in egocentric space’ (13), relative to the embodied perceiver. Since receptive field content and perceptual content are not the same, there is no match, and so the experiments fail to identify a content NCC. The immediate objection to this is that the conclusion doesn’t follow because the experimenters may be happy with a partial match, and then work at filling in the picture.

N&T are aware of this kind of objection and respond by citing studies that show that the receptive field properties of single neurons depend on what the animal as a whole is doing (behavioural factors, auditory stimulation etc.) (14-15). This is thought to exclude the idea that a complete match of content can be obtained by looking for activity elsewhere in the brain. But, firstly, this doesn’t really address the objection concerning partial match: one might still identify partially matching content NCCs by studying the brain. Secondly, the definition of the content NCC specifies that we

are looking for the representation in a neural system which, *under conditions C*, is sufficient for conscious representation (6). Those conditions may specify some of these environmental factors, for example, those obtaining in the scanner. The specification may have to be the result of a trade-off between our pre-scientific conception of consciousness and the limitations of the experimental settings and techniques. But it is hard to believe that the limitations set by the context of scanners will be fatal to the NCC project, people do after all have conscious experiences in scanners too. Thirdly, there is the important possibility that, insofar as such environmental factors have an influence on consciousness, they make a difference to neural activity, and as such they can be subject to orthodox NCC research.

Of course, N&T are right to the extent that we have to guard against thinking that one kind of study easily can establish the existence of a content NCC. It will be necessary to measure brain activity under various experimental conditions, with a range of measurement techniques, and across multiple conditions (Frith et al 1999). There is a long way to go, but N&T haven't shown that we are going to wrong way.

#### **4. The phenomenological argument.**

N&T's second argument concludes that neural structure content *could* not match conscious content. It begins with the important phenomenological observation that perceptual content is structurally coherent, intrinsically experiential, and active and attentional (16). The idea is that neural systems cannot have contents that match these aspects of perceptual content (though they might 'agree' (17)). For example, perceptual content is intrinsically experiential in the sense that the content is always experienced from a point of view (an experience of a butterfly is always experienced as of a butterfly standing in a certain egocentric spatial relation to the experiencer). But, as they observe, it

is false to say that neurons experience things from a point of view (18). They arrive at the conclusion that therefore there can be no match. (Similar arguments are mounted for the other phenomenological aspects).

This argument brings the problems about the match-agree distinction to the fore. As mentioned above, it seems too strong to require that when an underlying state constitute a more superficial state, then the underlying state must itself possess the properties of the superficial state. It also seems that one could argue in a parallel manner that a verbal report about a sunny blue sky cannot match a photograph of a sunny blue sky in content because words do not have colours. This seems to show either that the match-agree distinction is used to beg the question against the NCC defenders, or that the distinction is misconceived for addressing the issues at hand.

N&T also use such considerations about embodied experience to broaden their target to the very idea that perceptual experience has a minimal neural substrate: 'if perceptual content depends on the skilful activity of the whole animal or person, making use of its capacities for eye, head, and whole body movements, and for directed attention, then it becomes questionable whether there is any such thing as a minimal neural substrate sufficient to produce experience' (20). They argue that this supports the view that the content of neural systems may 'causally enable' perceptual content, without being constitutive of it (22). On this view, content NCCs cannot be identified because content is not neurally constituted.

This position on the mind-brain issue may be a kind of materialism, since it could be the view that consciousness supervenes on the material state of the whole body and its environment. It is an interesting question how such a broader supervenience base would itself avoid the problems raised

by N&T. And again, it is worth pointing to a different possibility that is closer to current orthodoxy. It may be that embodiment has an influence on conscious content, but that differences in embodiment are correlated with neural differences that go into the constitution of the conscious contents. N&T in fact acknowledge this possibility (30) but do not provide an argument against it.

Lastly, it is not entirely clear what it is to 'causally enable' something. Presumably it is something that falls between causing and constituting, but it is unclear what that may be. A clear example would be very helpful here. In the absence of such an example it just looks like the idea that neural systems form part of the supervenience base for consciousness, and that would again bring the partial matching notion into play.

##### **5. The internalist presuppositions of the content-NCC program.**

The NCC approach is committed to the internalist idea that whenever the content NCC for butterflies is activated, then there will be a conscious experience as of a butterfly irrespective of whether there really is a butterfly or not. This is internalist in the sense that veridicality in principle can be irrelevant for phenomenology (24-6).

N&T suggest that internalism of this kind is controversial. But the argument they mention seems weak. It builds on the notion that there is great difference 'between something's looking a certain way to one, and its merely seeming to one as if something looks a certain way to one. In the first case, one's experience involves an object in the world. In the second, it does not [...] so we can reject the idea that we are aware of one and the same thing when we perceive and when we hallucinate' (27). There might be this kind of difference in some cases, but not in very many real life cases of hallucinating subjects. They mostly do not experience things merely *as if* someone is,

say, shouting at them, they experience someone shouting at them. This includes schizophrenia, musical hallucinations associated with deafness, and normal observers doing difficult psychophysics experiments. The experiencer tests the hallucination to *discover* if it is real or not (e.g., it moves with your eyes), and some hallucinations are more difficult to test than others. For example, Hammeke et al. (1983) report how a lady, on first experiencing vivid musical hallucinations searched her flat for the source. This doesn't rule out that there is a phenomenological difference between perception and hallucination, but that point cannot be argued by claiming that hallucinations are 'as if'.

#### **6. Are binocular rivalry studies suited for identifying content NCCs?**

N&T also doubt whether binocular rivalry studies are suitable for identifying the correlates of the rivalling contents. Their first argument is that, being committed to internalism, the NCC approach must say that it is possible for there to be no difference between experiencing binocular rivalry and experiencing non-rivalling alternating stimuli. They claim that this is phenomenologically implausible because binocular rivalry has a distinct feel (30). This seems like an empirical claim, and though the studies do not seem to have been done yet, we are told that it is in fact very hard to emulate binocular rivalry, perhaps because subjects have such a feel when undergoing rivalry, but not alternating stimuli (David Leopold, personal communication). However, even if there is such a difference, this doesn't automatically disqualify binocular rivalry studies, because there might still be a partial match between the rivalling (minus the feel) and the non-rivalling content.

Their second argument against binocular rivalry studies anticipates this objection about the possibility of partial matching. It begins with the observation that binocular rivalry studies presuppose that the experience of binocular rivalry is like the conjunction of two 'atomic'

alternating experiences (first an experience of a sunburst, then of a butterfly, and so on). Without this presupposition one could hardly claim that these studies are identifying the NCC for the conscious content of one of the rivalling contents, for example, a butterfly. N&T report that the experience of binocular rivalry is not like the conjunction of two atomic experiences. Rather, it is one bistable experience that is not decomposable and that is unlike the normal experience of first seeing a butterfly, then seeing a sunburst, and so on. If that is right, then binocular rivalry studies are not identifying the NCCs for the rivalling contents (30-31).

However, there is some evidence that binocular rivalry studies are identifying the right kind of thing. Kanwisher and her colleagues identified the fusiform face area (FFA) and the parahippocampal place area (PPA) as part NCCs for faces and places, respectively (Kanwisher et al 1997, Tong et al 2000, Epstein and Kanwisher 1998). Importantly, they did not rely on binocular rivalry studies. In a separate study (Tong et al 1998), they compared binocular rivalry of faces and places with non-rivalling alternating presentation of faces and places. The study showed a striking similarity in activation of PPA and FFA, also over time, of the rivalry and non-rivalry scans. This is exactly what one would expect if binocular rivalry studies were apt for identifying the NCCs for the rivalling contents. It might be that there is a phenomenological difference, but these studies give strong evidence that there is also considerable overlap.

As in their requirement for matching rather than just agreement, N&T expect results from NCC studies of binocular rivalry that go far beyond the aspirations of neuroscientists. N&T require a neural correlate for the totality of the rivalrous experience that extends over minutes.

Neuroscientists, on the other hand, are pleased to find a neural correlate for a small component of the contents of consciousness (a micro-consciousness, Zeki & Bartels, 1999) such as the awareness

of a face in a particular position in the visual field that lasts for milliseconds. N&T's arguments do not establish that this aspiration is misguided.

To sum up, Noë and Thompson highlight some important and often ignored aspects of consciousness, but their arguments do not establish the conclusions that the NCC approach (and in particular binocular rivalry studies) has failed in identifying NCCs for conscious content, nor that that such approaches must always fail.

## References

- Epstein, R., Kanwisher, N. 1998. A cortical representation of the local visual environment. *Nature* 392: 598-601.
- Frith, C., Perry, R., Lumer, E. 1999. The neural correlates of conscious experience: an experimental framework. *Trends in Cognitive Sciences* 3: 105-114.
- Hammeke, T.A., McQuillan, M.P. & Cohen, B.A. 1983 Musical hallucinations associated with acquired deafness. *Journal of Neurology, Neurosurgery, and Psychiatry* 46: 570-572.
- Kanwisher, N., McDermott, J., Chun, M. M. 1997. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *The Journal of Neuroscience* 17(11): 4302-4311.
- Tong, F., Nakayama, K., Thomas Vaughan, J., Kanwisher, N. 1998. Binocular rivalry and visual awareness in human extrastriate cortex. *Neuron* 21: 753-759.
- Tong, F., Nakayama, K., Moscovitch, M., Weinrib, O., Kanwisher, N. 2000. Response properties of the human fusiform face area. *Cognitive Neuropsychology* 17: 257-279.
- Zeki, S. & Bartels, A. (1999). Toward a Theory of Visual Consciousness. *Consciousness & Cognition* 8: 225-259