

Noë, A. and Thompson, E. (2003), 'Are there neural correlates of consciousness?', *Journal of Consciousness Studies*, this issue. Peacocke, C. (1983), *Sense and Content: Experience, Thought, and Their Relations* (Oxford: OUP). Putnam, H. (1999), *The Threefold Cord: Mind, Body, and World* (New York: Columbia Univ. Press). Raffman, D. (1995), 'On the persistence of phenomenology', in *Conscious Experience*, ed. Thomas Metzinger (Exeter: Imprint Academic). Tye, M. (2000), *Consciousness, Color, and Content* (Cambridge, MA: MIT Press/Bradford).

APPEARANCE IS NOT KNOWLEDGE: THE INCOHERENT STRAW MAN, CONTENT-CONTENT CONFUSIONS AND MINDLESS CONSCIOUS SUBJECTS

*Thomas Metzinger**

The value of Noë and Thompson's contribution consists in highlighting a series of important methodological and conceptual issues associated with current research in the cognitive neuroscience of consciousness. Frequently, these are not seen by empirical scientists attempting to delineate the neural correlate for a given type of phenomenal experience. More importantly, they draw attention to a deeper theoretical problem that may soon become even more pressing: What precisely are the *identity criteria* we can employ in making reference to specific forms of conscious content, both introspectively, when generating the firstperson experience (or even judgment) of *sameness*, as well as conceptually, when making explanatory claims from the third-person perspective? However, the paper has a number of shortcomings too. Let us briefly look at them first.

The first problem with Noë and Thompson's strategy is that they persistently equivocate between the two basic notions of 'intentional content' and 'phenomenal content'. Thus they create a straw man (by misdescribing their opponent's claim) and their own argument becomes incoherent (by committing a fallacy of equivocation). As a consequence they also misrepresent what actually constitutes their opponent's epistemic goal. What is *intentional* content? It is the *representational* content, which, for instance, a certain neural system in our brain may possess. In accordance with a naturalized version of Franz Brentano's original definition (Brentano, 1874, II, 1: §5) we may say that the representational or intentional content (IC) of this neural system is what its activity is *directed at*. It is the *object* of this activity. Very often this object will be external to the brain, and quite frequently it may not even exist at all. Therefore the most important property of our neural system — namely, if it is a misrepresentation or not, if it constitutes a potential source of knowledge for the organism or not — is often determined by properties external to the brain. What many of us today call 'a' content is an abstract property, and one and the same content may be given or represented in a number of different ways. What is *phenomenal* content? Phenomenal content (PC) is first-person content, the way a certain state feels from the inside, *what it is like* to be in that state (as Thomas Nagel put it, exactly one century after Brentano [but cf. Farrell 1950, p. 183]). It is 'phenomenal', because it determines the way reality *appears* to you. The implicit, but seldom argued assumption behind searching for a neural correlate of consciousness (NCC) is

*I want to thank Jürgen Fell and Bettina Walde for helpful comments and critical discussion.

68

that neural states can be *carriers* of PC. As a large majority of researchers in philosophy of mind currently agrees, PC can be characterized by the principle of local supervenience:

(PLS): By nomological necessity, once all internal and contemporaneous properties of an organism's nervous system are fixed, the PC of its subjective experience is fixed as well.

Very obviously, the same is not true of IC. The PC, on the other hand, is something a veridical perception and a hallucination may share. As I wrote in the general introduction to *Neural Correlates of Consciousness*: 'The phenomenal content of your mental representations is that aspect which, being independent of their veridicality, is available for conscious experience from the first-person perspective, while simultaneously being determined by inclusively internal properties of your brain'

(Metzinger, 2000, p. 3). Consciousness research is all about PC. This does not exclude the possibility that we may one day discover how PC actually is a specific *kind* of IC, how conscious content actually is a specific, neurally realized type of intentional content satisfying an as yet unrevealed set of additional constraints (for a first example of ten such constraints, see Metzinger, 2003, chapter 3). It is conceivable how one and the same intentional content may be given consciously and unconsciously. As a matter of fact, a whole range of philosophers today accept what has been dubbed the ‘hegemony of representation’ by William Lycan (cf. Lycan, 1996, p. 11; for typical examples see Dretske, 1995; Lycan, 1996; Tye, 1995; 2000). This strategy consists in a weak variant of Franz Brentano’s intentionalism: The explanatory base for *all* mental properties is formed by a certain, exhaustive set of functional and representational properties of the system in question.

Now we can look at some examples where Noë and Thompson equivocate between IC and PC. When describing the matching content doctrine in their introductory paragraph, they immediately equivocate on IC (the *representational* content of certain neural systems introduced as *representational* systems) and PC (the contents of *consciousness* or *conscious visual experiences*). In section I they describe the target of their criticism as the doctrine of a match for *E-content* and *N-content*, referring to Chalmers (2000), but not giving a page number or referring to any of the specific definitions Chalmers developed. The question arises if there is any proponent of *this* claim in the real world and, more fundamentally, if there are any conceivable experiments that could at all support *this* claim. How would one empirically show that a given, internally determined kind of PC (*‘E-content’*) correlates with a specific, externalistically individuated IC (*‘N-content’*) in all cases forming the target domain, e.g., during hallucinatory episodes as well? The idea itself is incoherent, because IC depends on the state of the external world.¹⁴ Put crudely, it depends on if there *is* anything out there that

[14] I am skipping a minor complication here: Of course internal IC may exist as well, for instance in mentally represented self-knowledge, cognitive self-reference, and the like. The properties determining the IC of such self-representations however, will in almost all cases be external to the minimally sufficient neural correlate underlying the *phenomenal* self-representation, which could potentially go

our neural representational system can internally portray for the system as a whole. The ‘minimal substrate thesis’ developed from this equivocation is not only a straw man (because nobody really defends it) but also logically incoherent, because it makes no sense to assume that a given state solely individuated by its IC could be nomically and minimally sufficient for a specific target PC.

Further instances of the equivocation can be found in the notions of ‘perceptual states’, ‘perceptual experience’, ‘contents of consciousness’ (all section I), ‘the intentional content of visual *experience*’ (italics TM, p. 7. fn 7), of *a* or *the* ‘content NCC’, of ‘perceptual content of the subject’s visual experience’ (all section II), and so on. In section III the authors interestingly introduce the additional notions of ‘personal-level content’ and ‘subpersonal-level content’, but without giving defining characteristics. However, persons are not something we simply find out there in the world: Personhood is something that is generated in societies, through the development of a shared normative context. Are all animals unconscious, because they are not persons? Can I suddenly go unconscious by being transposed into another social context, simply because I now do not count as a person any more? ‘Going normative’ is not a good antireductionist strategy as long as one is unable to specify a deep and precise conceptual link between normativity/personhood and phenomenality. As the underlying conceptual distinction is unclear, so becomes the charge of equivocation against Chalmers in section V. As it is unclear what ‘the’ content-matching doctrine between subpersonal and personal levels actually amounts to, a second straw man emerges: Who was it who made or argued for *this* specific claim? It could be argued that Noë and Thompson inherit their problem from Chalmers’ overall definition (Chalmers, 2000, p. 31), which does not explicitly distinguish between locally versus regionally/globally supervening kinds of content, as well as between PC and IC. But it is clear that the minimal substrate thesis aims at PC only.

A second problem with their strategy for framing the problem is their assumption that states of consciousness causally depend on states of the brain (see introductory paragraph). This excludes the case of identity for neural and phenomenal states, which in turn is the implicit assumption governing most

empirical research programmes. Identity is a relation in which everything stands to itself; the relations of causal relations, however, are distinct events. No event causes itself. If the event is a phenomenal property-exemplification, contributing to the set of currently actualized properties in a specific domain which we call a 'state of consciousness', then it follows that it must be distinct from any physical event causing it. This begs the question against the reductionist right at the outset. Teller and Pugh's notion of a 'bridge locus' unfortunately adds to this confusion. First, Chalmers certainly never wanted to say that bridge locus neurons are *necessary* for the activation of a particular type of PC. Second, if, for each phenomenal property, there is a nomologically coextensive property in the physiological

along with such forms of being representationally directed at oneself. It is of course an intriguing thought experiment to conceive of states phenomenally as well as intentionally representing exactly the same set of neurobiological properties, namely their own, shared minimal correlate — but I will leave the discussion of such scenarios for a later stage. For more on consciousness and cognitive self-reference, see Metzinger 2003b.

basis set (pp. 4–5, fn 4) then local supervenience collapses into token-identity, therefore, again, one may not want to assume 'causal dependence' of consciousness on the brain at all.

What I like about Noë and Thompson's piece is that it brings out what a limited part of the more general project of developing a theory of mind developing a theory of consciousness actually is. Appearance truly is an internal phenomenon, determined by functional roles exclusively realized by internal states of our brains. This is why consciousness is *only* appearance: As such, it has nothing to do with epistemic relations and intentional content, with what it means to have a *mind* in terms of being contextually situated in the world through knowledge relations, as rational agents and as empathizing social subjects. Of course, in our own case, and in standard neurophenomenological configurations, consciousness beautifully and elegantly complements all these more complex features, namely by elevating them to higher levels of flexibility, selectivity and globalized context-sensitivity. But being a philosopher who subscribes to the notion of a single sufficient neural correlate for every specific activation of PC, I also have to assume the possibility of a mindless subject of conscious experience: In a radicalized version of the original brain-in-a-vat thought experiment, it is obviously plausible to assume that we could mathematically model the activity of the minimally sufficient neural correlate of my reader's current phenomenal state on just the right level of functional granularity, a level that excludes any kind of ambiguity in the underlying isomorphism. We could in principle implement this mathematical model — the 'conceptual essence' of my reader's consciousness — on a single machine, or even have it wander through the internet like a softbot, a flexibly embodied phenomenal nomad. At least in the very first moment, during the "snapshot of creation", we would then have created mindless consciousness. Everything Noë and Thompson say, equivocatingly, about 'experience' necessarily being integrated into normative social contexts, personal-level descriptions, about the hidden intentionality of RF-content or regional supervenience in terms of true minds dynamically transgressing 'brain–body–world divisions', everything they may want to point out about embodiment, agency and empathy, could then constitute a convincing criticism against anybody who claimed that we had created subjectivity in any stronger sense of term— or even a mind. At least in the beginning, our isolated phenomenal pattern of activity would have none of these features. Conscious experience as such is not knowledge, and this may be one reason why traditional philosophers of mind have always been more interested in knowledge than in appearance. Finding the NCC will be a difficult task and one of the great achievements of humanity, but it will be only *one* building block on our way towards a comprehensive theory of mind.

Potentially the greatest merit of Noë and Thompson's target article is that it raises the issue of identity criteria for phenomenal states (see p. 23, fn 21, and Putnam, 1999, pp. 128–32 and 153–4 in particular). Do we possess transtemporal identity criteria that allow us, in introspection, reliably to judge that two phenomenal states really were *qualitatively identical*? There are good empirical reasons to believe that for many types of PC, e.g., for conscious colour perception, this is not true (see Raffman, 1995; Metzinger, 1997; 2003: section 2.4). Hilary Putnam has pointed out that the phenomenal experience of sameness may functionally be based on introspective indistinguishability, and not on a reliable form of identifying reference. In his own words

‘indistinguishability in appearance is not a transitive relation, but being in the same state ... is a transitive relation’ (p. 130), and he even goes as far as saying that the idea of an identity condition for phenomenal states as such is simply absurd (p. 132, see also n. 41). Unfortunately I here lack the space to point out how serious this difficulty is for anyone interested in introducing phenomenal content as a solid theoretical entity. To make a long story short: Public sameness-judgments for PC — which will obviously play a central role for every NCC-related research strategy — are anchored in private, introspective sameness-experiences, but, as we have just seen, *as such* these are just another form of PC. They are sameness-*appearances*. I think one way to solve this truly fundamental problem may lie in strengthening the holism-constraint for PC, also discussed by Noë and Thompson. Here is the idea: Phenomenal states are not isolated first-order atoms, at which a higher-order process of introspection is directed. Rather, they are integrated into a complex and dynamic world-model, containing the simulation of, e.g., a *present* and of a *self in the act of introspecting*. For every single time slice of this highest-order phenomenal whole it will be true that there is one single neural correlate with which it is token-identical and one functional analysis at an appropriate level of abstraction, which supports a non-ambiguous mapping of its experiential content. My own proposal would be that we may initially use the singularity of its holism as a target property in correlation studies (there is only *one* phenomenal whole of this type at any point in time, and so the indistinguishability problem does not arise — or so my own argument would have to run). The remaining question is if identifying first-person reference to *subglobal* units of phenomenal content via introspective access could ever be epistemically justified, and if not, if *this* point would not pose a major methodological obstacle to the NCC research strategy. From my perspective, it is a second strength of Noë and Thompson’s contribution to have once again moved this issue into the focus of attention.

References

- Brentano, F. (1874/1973), *Psychology from an Empirical Standpoint*, ed. Oskar Kraus, English edition ed. Linda McAlister, trans. Antos, C. Rancurello, D.B. Terrell and Linda McAlister (London: Routledge & Kegan Paul / New York: Humanities Press).
- Chalmers, D.J. (2000), ‘What is a neural correlate of consciousness?’, in Metzinger (2000).
- Dretske, F. (1995), *Naturalizing the Mind* (Cambridge, MA: MIT Press).
- Farrell, B. (1950), ‘Experience’, *Mind*, **50**, pp. 170–98.
- Lycan, W.G. (1996), *Consciousness and Experience* (Cambridge, MA: MIT Press).
- Metzinger, T. (1997), ‘Präsentationaler Gehalt’, in *Bewusstsein und Repräsentation*, ed. F. Esken and H.-D. Heckmann (Paderborn: mentis).
- Metzinger, T. (ed. 2000), *Neural Correlates of Consciousness: Empirical and Conceptual Questions* (Cambridge, MA: MIT Press).
- Metzinger, T. (2003a), *BeingNoOne* (Cambridge, MA: MIT Press).
- Metzinger, T. (2003b). Phenomenal transparency and cognitive self-reference. *Phenomenology and the Cognitive Sciences*, **2**, 353-393.
- Putnam, H. (1999), *The Threefold Cord: Mind, Body, and World* (New York: Columbia University Press).
- Raffman, D. (1995), ‘On the persistence of phenomenology’, in *Conscious Experience*, ed. T. Metzinger (Exeter: Imprint Academic / Paderborn: mentis).
- Tye, M. (1995), *Ten Problems of Consciousness* (Cambridge, MA: MIT Press).
- Tye, M. (2000), *Consciousness, Color, and Content* (Cambridge, MA: MIT Press).