PhiMiSci
Philosophy and the Mind Sciences

A double anniversary for the neural correlates of consciousness

Editorial introduction

Sascha Benjamin Fink[a] (sfink@ovgu.de)

This article is part of a special issue on "The Neural Correlates of Consciousness", edited by Sascha Benjamin Fink.

At the heart of the contemporary science of consciousness is the search for neural correlates of consciousness (NCCs). NCCs and the research programme associated with it rose to prominence 30 years ago with the article *Towards a neurobiological theory of consciousness* by Francis Crick and Christof Koch (1990).

The notion of a *neural correlate of consciousness* itself was around before their seminal paper. Crick and Koch, for example, cite Logothethis and Schall's *Neuronal correlates of subjective visual perception* (1989), which arguably points in the same direction. That paper reports of single neuron recordings of the subjective perception of movement. It is therefore far less ambitious than Crick and Koch's, which asks generally about the where, when, how, and why of neural activation correlating with consciousness. But during the reign of cognitivism, where consciousness was largely ignored, its overall usage was sparse and mainly in passing, e.g. when the term is also used by C. L. Hardin (1988, p. 130) in *Color for Philosophers: Unweaving the Rainbow* where he talks of the chromatic cells in the visual cortex as "the hypothetical neural correlate of 'consciousness' " – arguably shying away from such a thorny subject as consciousness by putting it in scare quotes. Earlier, we find the term in J. S. Griffith's *A View of the Brain* (1967, p. 25) as part of a more mathematical model of the brain's workings. However, before the advent of behaviorism, NCCs where popular enough to be mentioned in the entry on *Psychology* in the *Encyclopedia Britannica*, written by James Ward (1911).

[a]Philosophy-Neurosciences-Cognition Otto-von-Guericke Universität Magdeburg

Fink, S. B. (2020). A double anniversary for the neural correlates of consciousness: Editorial introduction. *Philosophy and the Mind Sciences*, *1*(II), 1.
https://doi.org/10.33735/phimisci.2020.II.85
 https://philosophymindscience.org ISSN: 2699-0369

This indicated that the idea that conscious processes have neural correlates was not uncommon in 19th century psychology, e.g. when Edmund Gurney (1881, p. 156) writes that "the movements of mind are correlated with nervous processes" or when Mary Whiton Calkins (1892, p. 399) speaks of "cerebral correlates of [...] intellectual processes".

So while Crick and Koch may not have originated the term "neural correlate of consciousness", their article was crucial in pushing it onto the scientific agenda by sketching a general research programme guided by the question "Where in the brain are the neural correlates of consciousness?" – which still occupies researchers today, 30 years later. The paper therefore deserves to be the most cited source and the central root of our contemporary neuroscience of consciousness.

Spelling the goal of this investigation out in terms of neural *correlates* (rather than something else) had several advantages and allowed researchers to get an empirical investigation off the ground. First, methods for capturing correlations (unlike more philosophical notions like supervenience, grounding, co-extensionality, identity and so on) is something already within the standard toolbox of empirical science. At its core, it measures the strength of dependence between two variables. Bernard Baars (1997) made this point explicit: In the NCC-framework, scientists treat consciousness as a variable in experimental set-ups. The goal is to see which neural variables co-vary with it. Second, and more importantly, while correlation may not be fully metaphysically neutral, it is not too restrictive either. Unlike *neural substrates, constituents, reduction bases*, or other notions that are more metaphysically loaded, neural correlates are compatible with a bunch of different but mutually excluding metaphysical views on how our neural goings-on relate to consciousness. Even in Leibnizian parallelistic dualism, there can be a neural correlate of a conscious experience, as Ward (1911, p. 600ff) already pointed out. But the same holds for monistic identity theory, where we might think of the correlates as ways in which one and the same thing is given to us or picked out, introspectively or neuroscientifically. Only a denial of the existence of matter or consciousness (i.e., eliminativism or irrealism) precludes a search for correlates because there is nothing to correlate with. Or one must claim that conscious goings-on lack even a statistical relation to neural goings-on, which is rare even among proponents of enactive, embodied, or extended approaches. Beyond such extreme positions, most metaphysics will allow for something neural correlating with consciousness. This metaphysical promiscuity allowed researchers to bracket ancient but ongoing philosophical debates on the mind-body problem and simply start their empirical work. If coined in terms of correlation, the results will be compatible with most of what philosophers (eventually or never) converge on. Empirical researchers could also bracket their own varying metaphysical temperaments – some like Popper & Eccles (1977) have dualistic, others more reductionistic tendencies. Such divergences need not hinder fruitful collaboration in a coordinated attempt to find out: "Where in the brain are the neural correlates of consciousness?"

While elegant in how it opens up avenues for empirical work, the NCC research programme also raised conceptual, methodological, and empirical questions, some

Fink, S. B. (2020). A double anniversary for the neural correlates of consciousness: Editorial introduction. *Philosophy and the Mind Sciences*, *1*(II), 1.
https://doi.org/10.33735/phimisci.2020.II.85

of which are still hotly debated. Consider, for example, that correlation is relatively cheap as any two variables correlate to some degree. In order to prevent the search for NCCs from becoming trivial, we must ask: Which further constraints on correlation should be imposed to distinguish mere statistical correlates from core correlates that hint at a foothold of consciousness in the material world? How can we screen off core correlates from enabling conditions or causal factors? How can we best access, categorise or measure the conscious side of the correlation, which seems elusively subjective – especially if it might be tied to observable behaviour in only a loose sense? How do neural correlates relate to theories of consciousness? How can candidates for core NCCs be tested? How do correlates relate to theories of consciousness? How do correlates of specific experiences (like *seeing vermillion* or *smelling burnt toast*) relate to correlates of broader conscious states like (like *dreaming* or *waking*)? How many different core NCCs should we expect for a given conscious state? How does this research programme relate to embodiment and extended views of the mind or to research on artificial consciousness? The list continues.

In approaching questions like these, the anthology *The Neural Correlates of Consciousness: Empirical and Conceptual Questions* (Metzinger, 2000), edited by Thomas Metzinger, stands out. Metzinger, who later also served as president for the *Association for the Scientific Study of Consciousness* ten years ago (2009/2010), brought together a wide range of empirical researchers and philosophers and thereby illustrated that the self-conception of the budding science of consciousness includes a continuous reflection about the philosophical foundations of its empirical side – but also a demand for deriving testable implications from philosophical views on consciousness in order to keep them accountable to empirical findings. This anthology also featured a detailed analysis by David Chalmers (2000) of what a neural correlate of consciousness is supposed to be: An NCC of an experience C is the neural state N that is *minimally sufficient* for that experience C. Being merely sufficient allows for several neural systems to give rise to a conscious experience. For example, before a stroke, neural area N_1 may have brought about experiences as of seeing red, feeling one's hand, or comprehending a word. After recovering from the stroke, these experiences may be brought about by N_2. Neither N_1 nor N_2 is therefore necessary for that experience. But searching for only sufficient criteria might be trivial: The state of my whole brain is sufficient for my seeing red now. Thanks to the minimality criterion, we can reject such trivial answers and ask again: "Where in the brain are the neural correlates of consciousness?" Chalmers' definition has become the gold standard, the one most often proclaimed to be what NCC-researchers aim to find.

This special issue of *Philosophy and the Mind Sciences* celebrates this double anniversary: 30 years since the research programme for finding neural correlates of consciousness was brought back onto the scientific agenda and 20 years since the standard definition of an NCC was published in a seminal anthology. Here, we

Fink, S. B. (2020). A double anniversary for the neural correlates of consciousness: Editorial introduction. *Philosophy and the Mind Sciences*, *1*(II), 1.
https://doi.org/10.33735/phimisci.2020.II.85

want to take up some of the conceptual, methodological, and empirical questions raised and take stock of how the field has developed since then. In this issue, seven original peer-reviewed research papers take up these threads.

Talis Bachmann, Mototaka Suzuki, and Jaan Aru in *Dendritic Integration Theory: A thalamocortical theory of state and content of consciousness* (2020) improve on the prominent view of thalamo-cortical loops as NCCs. They start from ten observations motivated by phenomenology and empirical data. A neuroscientific account of consciousness must explain all ten. Their *Dendritic Integration Theory* focuses specifically on L5p neurones that receive input from both cortex and thalamus and thereby can mitigate between the important cortico-cortical processing (relevant for the specificity of what one is conscious of) and thalamo-cortical processing (relevant for the integration of different conscious features into a unified experience). This allows for an interplay between contents of consciousness and larger background states being thalamically modulated. They argue that this also accounts for the ability for unconscious cortical processing.

Jakob Hohwy and Anil Seth present predictive processing as a framework for progressing the search for NCCs in their *Predictive processing as a systematic basis for identifying the neural correlates of consciousness* (2020). The problem they raise is that there is a discrepancy between theories of consciousness (which are not always cashed out in neural terms), and empirical findings about neural correlates of consciousness. They stress that we must distinguish between *arbitrary* correlates – statistical coincidences not predicted by any hypothesis or theory – and *systematic* correlates, which are directly implicated by a theory. The goal of the neuroscience of consciousness is to understand, explain and predict mechanisms facilitating consciousness. For this, we require a theory. But so far, there is no convergence onto one. We may facilitate convergence by comparing different theories neuroscientifically. In order to cash out theories of consciousness in neural terms and make them comparable with each other and neuroscientific data, they suggest *predictive processing* as a suitable mitigator. Predictive processing as a framework therefore promises a unification of the field without a commitment to a specific theory. This go-between, however, also has the potential to become a theory of consciousness in its own rights.

Colin Klein, Tim Bayne, and Jakob Hohwy question that we should search for neural correlates of consciousness at all in their *Explanation in the science of consciousness: From the neural correlates of consciousness (NCCs) to the difference makers of consciousness* (2020). They make the point that the search for NCCs rests on some controversial assumptions, which culminate in the *Lite-Brite-Model*: (i) that a neural event brings about a specific experience autonomously and independently of what else goes on in consciousness or the brain, (ii) that an experience is uniquely dependent on a single neural element, and (iii) that global changes in consciousness uniformly affect all occurrent experiences. They contrast this with the difference maker framework. Here, we ask which neural events make a difference to consciousness. Not only is this research programme logically more complex, it

Fink, S. B. (2020). A double anniversary for the neural correlates of consciousness: Editorial introduction. *Philosophy and the Mind Sciences*, *1*(II), 1.
https://doi.org/10.33735/phimisci.2020.II.85

is also tied to a rich tradition of analysing causal interactions, experimental set-ups, and has facilitated great progress in other fields like genetics. The days of searching for NCCs might be numbered.

Tomáš Marvan and Michal Polák in their *Generality and content-specificity in the study of the neural correlates of perceptual consciousness* (2020) pick up the question of what makes any content *conscious*. This is to them the pivotal question. According to them, just searching for NCCs is then too broad: The notions must be split up into *neural correlates of content* (NCc, with a lower case c) and a *general neural correlate of consciousness* (gNCC), which is the mechanism shared by all conscious contents. They assume that when we compare the neural correlates of conscious events, we will find the general neural correlate of consciousness as a stable part, but combined with varying neural correlates of content (which can, in principle, occur outside the NCC as well). So it is this broad mechanism that, in interaction, turns the content of a neural representation conscious. They map out which theories do or don't allow for such a distinction – which in turn lends itself to evaluating such theories if the distinction is accepted.

Matan Mazor and Stephen M. Fleming in their *Distinguishing absence of awareness from awareness of absence* (2020) focus on one of the most important paradigms in the search for NCCs, namely the contrastive analysis of trials where a subject is aware of a stimulus and those where a subject is unaware of a stimulus. They point out that this falls short of what is actually needed: Subjects are not only aware of stimuli, they sometimes are also aware *of the absence* of a stimulus. Or, vice versa, they can be unaware of the presence of a stimulus but also unaware of the absence of a stimulus. In both cases, the awareness of absence and the awareness of a stimulus, the neural correlate of awareness might be present. Similarly, in both the unawareness of a stimulus and the awareness of a stimulus, a neural correlate of the stimulus may be present. If we do not capture this difference, we might either discard or fail to see relevant neural data. Hence, they argue for a two-dimensional report scheme. In concordance with Marvan and Polák, they suggest that this allows for distinguishing between content-specific and content-invariant NCCs.

Andy McKilliam (2020) asks the question *What is a global state of consciousness?* This is a central question to NCC research, specifically regarding the distinction between content-specific and content-invariant correlates. A general mechanism for consciousness might also vary when global states vary from wakefulness to drowsiness or to dreaming and so on. Still, such global states might be content-invariant. In order to research global states of consciousness, we need to know what it is. McKilliam finds all explications currently at play wanting, be they as levels, phenomenal states in their own right, total states constituted by all occurrent experiences, phenomenal signatures, or as phenomenal structures. As a contrast, he presents the capacities account of global states, where global states simply govern what a creature is capable of experiencing and how it is capable of accessing these experiences while in that global state. But I need not occurrently exercise my capacities in order to have them, e.g. I have the capacity to ride a bike

Fink, S. B. (2020). A double anniversary for the neural correlates of consciousness: Editorial introduction. *Philosophy and the Mind Sciences*, *1*(II), 1.
https://doi.org/10.33735/phimisci.2020.II.85

even though I am not currently riding a bike. The controversial upshot is that we may miss NCCs for global states if we only look at neural data correlating with occurrent experiences. This suggests a totally new approach for investigating the neural correlates of waking, dreaming, being under anaesthesia and so on.

Matthias Michel and Hakwan Lau report *On the dangers of conflating strong and weak versions of a theory of consciousness* (2020). They emphasise that we have to distinguish *markers* from *constituents* and both from *correlates* of consciousness. Neural markers are events that we can exploit and use epistemically as indicators of consciousness. In that role, they differ from someone telling us that they are conscious mainly in that they are neural, not verbal. Constituents, in contrast, lend themselves to reduction because they are what conscious events can actually be identified with. Both are neural correlates, but markers are correlates only in a statistical sense, while constituents go beyond Chalmers' conception because they are also necessary in some sense. In that, neural constitution is more metaphysically demanding than correlation – and much harder to prove empirically. Failing to make the trifold distinction between neural markers, constituents and correlates of consciousness might lead to overestimating the scope of a theory. They illustrate this with a critique of the *Integrated Information Theory*: If integration is understood as a marker of consciousness, then the theory is empirically testable and maybe even supported – but the theory becomes untestable if integration is understood as a constituent of consciousness. Integrated Information Theory in this fundamental reading goes far beyond finding correlates. Instead, it detaches itself from empirical findings by claiming something untestable.

These seven original research papers clearly illustrate on the one hand the impact which NCCs still have on the science of consciousness and on the other hand the controversies this notion still raises. We still make progress and amend both the research programme inaugurated by Crick and Koch 30 years ago as well as the definition of an NCC presented by Chalmers 20 years ago, all in the interdisciplinary, empirically grounded and philosophically open spirit exhibited in the anthology by Metzinger. *Philosophy and the Mind Sciences* will continue this discussion in a second issue in 2021 with several more articles. These, like this issue as a whole, have been severely affected and hindered by the COVID-19 pandemic. So I am especially thankful to all the authors and reviewers for their work during these hard times. My gratitude goes to them as well as our supporters, but also to Wanja Wiese and Jennifer Windt as well as Fabian Fuchs for their help in getting this first part of the special issue out in 2020. Only thanks to this communal effort was it possible to celebrate this double anniversary for the science of consciousness this year.

Fink, S. B. (2020). A double anniversary for the neural correlates of consciousness: Editorial introduction. *Philosophy and the Mind Sciences*, *1*(II), 1.
https://doi.org/10.33735/phimisci.2020.II.85

References

Baars, B. J. (1997). *In the theater of consciousness.* Oxford University Press.

Bachmann, T., Aru, J., & Suzuki, M. (2020). Dendritic integration theory: A thalamocortical theory of state and content of consciousness. *Philosophy and the Mind Sciences*, *1*(II: Special Issue on the Neural Correlates of Consciousness). https://doi.org/10.33735/phimisci.2020.II.52

Calkins, M. W. (1892). A suggested classification of cases of association. *The Philosophical Review*, *1*(4), 389–402. http://www.jstor.org/stable/2175939

Chalmers, D. J. (2000). What is a neural correlate of consciousness? In T. Metzinger (Ed.), *Neural correlates of consciousness: Empirical and conceptual questions* (pp. 17–39). MIT Press.

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

Griffith, J. S. (1967). *A view of the brain.* Oxford University Press.

Gurney, E. (1881). Monism. *Mind*, *VI*(22), 153–173. https://doi.org/10.1093/mind/os-VI.22.153

Hardin, C. L. (1988). *Color for philosophers: Unweaving the rainbow.* Hackett Publishing Company. http://gen.lib.rus.ec/book/index.php?md5=3B87964D6381925DC1468BE87B5B077E

Hohwy, J., & Seth, A. (2020). Predictive processing as a systematic basis for identifying the neural correlates of consciousness. *Philosophy and the Mind Sciences*, *1*(II: Special Issue on the Neural Correlates of Consciousness). https://doi.org/10.33735/phimisci.2020.II.64

Klein, C., Bayne, T., & Hohwy, J. (2020). Explanation in the science of consciousness: From the neural correlates of consciousness (NCCs) to the difference makers of consciousness (DMCs). *Philosophy and the Mind Sciences*, *1*(II: Special Issue on the Neural Correlates of Consciousness). https://doi.org/10.33735/phimisci.2020.II.60

Logothetis, N. K., & Schall, J. D. (1989). Neuronal correlates of subjective visual perception. *Science*, *245*(4919), 761–763. https://doi.org/10.1126/science.2772635

Marvan, T., & Polák, M. (2020). Generality and content-specificity in the study of the neural correlates of perceptual consciousness. *Philosophy and the Mind Sciences*, *1*(II: Special Issue on the Neural Correlates of Consciousness). https://doi.org/10.33735/phimisci.2020.II.61

Mazor, M., & Fleming, S. (2020). Distinguishing absence of awareness from awareness of absence. *Philosophy and the Mind Sciences*, *1*(II: Special Issue on the Neural Correlates of Consciousness). https://doi.org/10.33735/phimisci.2020.II.69

McKilliam, A. (2020). What is a global state of consciousness? *Philosophy and the Mind Sciences*, *1*(II: Special Issue on the Neural Correlates of Consciousness). https://doi.org/10.33735/phimisci.2020.II.58

Metzinger, T. (Ed.). (2000). *Neural correlates of consciousness: Empirical and conceptual questions.* MIT Press.

Michel, M., & Lau, H. (2020). On the dangers of conflating strong and weak versions of a theory of consciousness. *Philosophy and the Mind Sciences*, *1*(II: Special Issue on the Neural Correlates of Consciousness). https://doi.org/10.33735/phimisci.2020.II.54

Popper, K. R., & Eccles, J. (1977). *The self and its brain.* Springer.

Ward, J. (1911). Psychology. In *Encyclopedia britannica: Vol. XXII* (11th ed., pp. 547–604). Cambridge University Press.

Fink, S. B. (2020). A double anniversary for the neural correlates of consciousness: Editorial introduction. *Philosophy and the Mind Sciences*, *1*(II), 1.
https://doi.org/10.33735/phimisci.2020.II.85