Professor Georg Northoff is a philosopher, neuroscientist and psychiatrist, he works at the University of Ottawa. In his research, he focuses on problems about neuronal basics of subjective phenomena like self and consciousness.

SOLVING THE WORLD-BRAIN PROBLEM

INTERVIEW WITH GEORG NORTHOFF

Filip Stawski, Mateusz Tofilski

Piąte Piętro: *We would like to start by thanking you for the acceptance of our invitation and by asking you a question about your research practice, where you combine the perspective of a philosopher, a psychiatrist and a neuroscientist. How do you think these spheres intertwine and how they can build a reliable dialogue with each other?*

Georg Northoff: Thank you for giving me the opportunity for this interview which I very much appreciate. When you look into the history of philosophy combining philosophy with science was rather the norm than the exception. Moreover, physics in the 20th century and nowadays strongly touches upon both ontological and epistemological issues. Only some parts in philosophy of the 20th century denting strongly into our time conceived a split between science and philosophy along the divide of the conceptual-logical vs the observational-empirical dimensions. I consider myself to stand on the shoulders of my predecessors in both the earlier philosophy and 20th century physics rather than on those of the 20th century dichotomy of philosophy vs science. What is currently lacking is a systematic and valid method for linking conceptual-logical and observational-empirical dimension and thus, more generally, philosophy and science. This is indeed one of my major aims. You can see such methodology, e.g., non-reductive, already developed in my earlier book *Philosophy of the Brain* (2004) and more elaborated in my textbook *Minding the Brain* (2014) (especially chapter 4). In short, I claim for a methodological strategy I designate as "concept-fact iterativity" as a continuous methodological and iterative movement between philosophical concepts and empirical data/facts. Historically, such concept-fact iterativity stands on the shoulders of Kant who argued that "concepts without intuitions are empty and intuitions without concepts are blind" – one can conceive my method of concept-fact iterativity as development towards a systematic relationship between concepts and facts/intuitions in methodological regard. And obviously, you can see that method applied in my various neurophilosophical writings especially in *Unlocking the brain* (Vol II, 2014) and *The Spontaneous Brain* (2018). Hence, to answer your question, I conceive a solid and valid method a first and indispensable step for a sound interdisciplinary dialogue something which our predecessors in philosophy and physics did more on an intuitive (and not always systematic) basis. And, even more, important and reaching beyond methodological issues, I conceive it almost a necessity to tackle the basic question in ontology, epistemology and ethics in the terms of such interdisciplinary methodology. Following the motto of Kant, our epistemological limitations will only yield to shortcomings and further limitations when we conceive these questions in only empirical or conceptual terms. My empirical background may thus serve, so I hope, to make me a better philosopher in the same way my philosophy, as I think, makes me a better neuroscientist.

PP: *You have already mentioned the term "neurophilosophy". It was coined in 80's by Patricia Churchland, since then it has evolved and has been used in various contexts. In our latest interview, Henrik Walter even said that "the golden age of neurophilosophy is over" (Walter, 2018). Meanwhile, you are running your studies under this sign (for example using the division into reductive and non-reductive neurophilosophy). How would you define this discipline?*

GN: I am really disappointed that a person as intelligent as Henrik Walter follows the myopic trends of the anglo-american mainstream which often is more interested in public relation than philosophical development. Let us be more sophisticated. The concept of neurophilosophy can be understood in different ways. It can be understood in a reductive vs non-reductive way – that is a methodological characterization about the relation between concepts and facts. As it is clear from my concept-fact iterativity, I opt for a non-reductive approach. Then one can distinguish narrow vs wide neurophilosophy: this is about the view and model of the brain. Narrow neurophilosophy conceives the brain in a purely empirical way, in the same way, neuroscience conceives it. Wide neurophilosophy, in contrast, conceives the brain is not only an empirical but also ontological and epistemological context – this amounts to what I call *Philosophy of the Brain* in my early book (2004). I agree that Churchland's reductive narrow neurophilosophy is a dead end; but that is something I already said ten or twenty years ago, for that one does not require much philosophical insight.

Reductive neurophilosophy favors a reduction of philosophical, e.g., ontological and epistemological and ethical concepts to the empirical facts of the brain. This amounts to concept-fact reduction or even stronger concept-fact elimination. Conversely, concepts and facts may stand in a parallel relationship – concept-fact parallelism - as it espoused by Bennett and Hacker and major lines in current philosophy of mind. Non-reductive neurophilosophy opts against both concept-fact reduction and concept-fact parallelism by providing methodological tools for their systematic investigation in dependence and mutual constraint – this amounts to concept-fact iterative as a non-reductive methodological strategy.

So, standing on the shoulders of both the history of philosophy and neuroscience, wide non-reductive neurophilosophy has, as I would say, a rather bright future as it allows to raise novel questions and problems like the world-brain problem replacing traditional ones like the mind-body problem. Again, developing and applying such wider non-reductive neurophilosophy, I am standing on the shoulders of my philosophical predecessors like Schopenhauer and Bergson who, being neurophilosophers 'avant le mot', conceived the brain in an epistemological and ontological context.

PP: *How this neurophilosophy differs from the philosophy of mind or philosophy in cognitive science?*

GN: For me, philosophy of mind is about the conceptual, ontological, metaphysical, and epistemological investigation of the mind. Philosophy of mind in this sense can be complemented

WHEN YOU LOOK INTO THE HISTORY OF PHILOSOPHY COMBINING PHILOSOPHY WITH SCIENCE WAS RATHER THE NORM THAN THE EXCEPTION

by wide non-reductive neurophilosophy – this, as I claim in my 2018 book, can lead to the replacement of the mind-body problem by the world-brain problem. Philosophy of cognitive science is for me like the application of philosophy of science to cognitive science. As such philosophy of cognitive science concerns, among other issues, the models of mind and cognition presupposed in cognitive science. I would say that we desperately need a proper philosophy of neuroscience (as a branch of philosophy of science) that, for instance, discusses the sometimes rather naive models of the brain we most often tacitly presuppose in both neuroscience and philosophy of mind.

PP: *Then, speaking of the complementing of the philosophy of the mind with a neurophilosophical perspective, you argue that the traditional mind-body problem should be replaced by the world-brain problem. Actually, you also write about it in your last book The Spontaneous Brain. Could you describe the differences between these two perspectives?*

GN: Thank you for asking this question. I do not intend to provide an answer to the mind-body problem. Instead, I aim to question and, even stronger, dissolve (rather than answering and solving) the mind-body problem by questioning its tacit presuppositions; this resembles the kind of methodological approach Kant designated as transcendental. One such presupposition consists in the possible existence and reality of mind: only if one presupposes the mind, one can raise the question for its possible relationship to the body, the mind-body problem. If, in contrast, one no longer presupposes the mind, the question for its relationship to the body, the mind-body problem, becomes non-sensical as one cannot raise the question for a relation of something, e.g., the body, to something that remains impossible, e.g., the mind. Yet another usually tacit presupposition is that the mind exhibits necessary relationship to mental features like consciousness – that connection is necessary as the mind is supposed to account for the ontological substrate of mental features. This, by itself, presupposes distinction of mind and mental features: mental features are consciousness, self, etc. while the mind is their supposedly underlying ontological (or metaphysical) substrate. That distinction entails that something else other than the mind may provide the ontological substrate for mental features. Moreover, the distinction entails that the rejection of mind (as ontological substrate) does not entail the rejection of mental features – absence of mind is well compatible with the presence of mental features.

Based on various lines of empirical (chapter 1-8), ontological (chapter 9-12), and epistemological (chapter 12-14) evidence, I argue that the world-brain relation (taken in an ontological rather than empirical sense) is necessarily related or connected to mental features like consciousness and can therefore serve as their underlying ontological substrate. The world-brain relation can thus takeover the ontological role the mind is traditionally supposed to play for mental features. Now, given the fact that the mind is the presupposition of the possible mind-body problem, replacing the mind in its role for mental features by world-brain relation entails that the mind-body problem becomes non-sensical and can thereby be dissolved – it simply becomes nonsensical and meaningless to even raise the question for the mind's relationship to the body as there is no mind anymore. One can then replace the mind-body problem by the world-brain problem as the more plausible (on ontological, conceptual, and empirical grounds) problem to address the question for the existence and reality, e.g., the ontological substrate of mental features. Accordingly, rather than providing an answer to the mind-body problem, I replace its presupposition of mind by the one of world-brain relation as the more plausible ontological substrate of mental features. One can then speak of a world-brain problem that, as I hope, provides a novel ontological framework for discussing mental features like consciousness (as I discuss in the 2018 book) and others like self and personal identity (on which I am currently working).

PP: *So you suggest, that consciousness is not in the head but it is a relational feature between the brain and the world. Such a sentence immediately brings to mind associations with the context of 4E cognition (embodied, embedded, enactive, extended). How would you define your position in relation to this current?*

GN: The answer to this question leads me to yet two further tacit assumptions we usually presuppose in our account of mind. Descartes postulated the existence and reality of mind as distinguished from the world we live in and observe. The concept of mind is thus necessarily isolated from the world as otherwise, he could have no longer distinguished mind and world. Since mental features like consciousness are assumed to be necessarily dependent upon the mind, they inherit the isolation from the world of the concept of mind. To account for the relation of mind to the world on the basis of the assumption of such world-isolated mind, one can then pursue different strategies. One such strategy is suggested by John McDowell who conceive the mind's conceptual capacities and rationality as "second nature of man" and integrates it in what one can describe as "conceptually-extended logical space of nature". Yet another strategy to account for the relation of mind and world are the 4E's you mention. Now the primarily world-isolated mind is related to the body and the world. While this is a very laudable attempt, it nevertheless cannot overcome its primary birth-defect, as I say, the necessity of the isolation of mind from the world by means of which the mind is defined. Put simply. One first excludes the world from the mind (when defining it) and then tries to bring back world into mind. However, as we all know, that what is excluded once on a necessary basis, e.g., the world

ACCORDINGLY, RATHER THAN PROVIDING AN ANSWER TO THE MIND-BODY PROBLEM, I REPLACE ITS PRESUPPOSITION OF MIND BY THE ONE OF WORLD--BRAIN RELATION AS THE MORE PLAUSIBLE ONTOLOGICAL SUBSTRATE OF MENTAL FEATURES

In other words, we need something more radical than the 4E's to overcome the birth deficit in our definition of mind, that is, the necessary exclusion of the world

in the definition of mind, cannot be brought in through the backdoor as that cannot remedy the initial birth deficit. In other words, we need something more radical than the 4E's to overcome the birth deficit in our definition of mind, that is, the necessary exclusion of the world. Importantly, when I speak of world, I do not mean the world of our higher-order cognition as in rational conceptions or the world of consciousness as in phenomenological approaches. I mean the world as it is by itself independent of our rationality, consciousness, or otherwise, that is, in a mind-independent way. Recruiting Kant, Husserl, and others, one may now be inclined to argue that such mind-independent concept of the world may remain impossible; the only way, we can approach and conceive the world is in relation to us. That is the moment where the brain comes in: our brain aligns us to the world by means of which we become part of the world – this amounts to what I describe as world-brain relation (as distinguished from brain-world relation where the brain imposes itself upon the world (rather than aligning and adapting itself to the world). Together with my repudiation of the first two presuppositions (see above), I argue that the world-brain relation, framed in an ontological sense, can overcome the birth deficit of the exclusion of world in our definition of mental features. To fully understand my point, one need to conceive yet another a fourth tacit presupposition in our account of mind. We usually presuppose properties like mental or physical properties as the basic units of existence and reality – this amounts to what I describe as "element-based ontology" (Northoff 2018). The concept of mind is based very much on such element-based ontology as it is accounted for by substance (Descartes) or properties (nowadays). However, this is to neglect alternative forms of ontology that have been more developed on the side-lines than the mainstream of philosophy. Such alternative ontologies emphasize the priority of a process, relation, and transformation over elements and properties – these lines of ontology can be traced to Cassirer, Whitehead, Bergson, and nowadays structural realism. I presuppose such relation-based ontology, that is, structural realism, in my account of the world-brain relation – it is the relation of world and brain as an integration or alignment of a part, e.g., the brain, to the whole, e.g., the world, that constitutes and provides the necessary ontological condition of possible mental features like consciousness. Hence, I conceive mental features ontologically to be relation- rather than property-based as they can be traced to the world-brain relation. This, as I hope, makes it clear that I am much more radical than the 4E's; mental features are necessarily and thereby intrinsically relational and thereby neuro-ecological.

PP: *The paradigm of the embodied mind is also mentioned in the context of psychiatry, so do you think that it can change our understanding of mental illness? We ask this question primarily because you are also a promoter of the neuro-ecological model of psychiatric disorders.*

GN: The impact of the 4E approach in psychiatry is largely theoretical in explaining these disorders. This differs from my neuro-ecological approach which, albeit theoretical, guides our empirical research including both experimental designs and data analysis. Let me briefly explain how the neuro-ecological approach conceives mental disorders. Mental disorders are world-brain disorders in both ontological and empirical regard. The brain shows abnormalities in these patients but, equally, one can see the strong dependence of the brain's neuronal function on the contextual changes, e.g., the environmental context as for instance in terms of life events and neuro-developmental changes. The same kind of world-dependence holds for the symptoms themselves. Psychiatric patients do not experience their symptoms in their head, they experience them as part of a wider world of which they and their symptoms are part. You can see that I here dwell on yet another historical line in philosophy, phenomenology, which provides an excellent account of the structure of our experience of ourselves, time and space, body, and most importantly, the world. Hence, we see the necessary dependence on the world in both brain and symptoms in psychiatric disorder – they are thus disorders of world-brain relation and, more succinct, neuro-ecological disorders (rather than neuronal or neuro-cognitive disorders).

PP: *One more, general question about the relationship between psychiatry and philosophy. Please tell us whether your philosophical education and experience help you in your medical practice?*

GN: The exposure to psychiatric disorders left a deep impression upon me. As George Canguilhem developed so nicely, we can learn from the pathological for our understanding of the normal. Extremes can tell us about the average. Hence, one strong source of inspiration of my ontology and epistemology in philosophy is psychiatry; it provides validation and plausibility for any ontological assumption: if the latter is not in accordance with the former, my ontological assumption may be conceptually plausible but not empirically consistent. That means, to go back to the drawing board and develop a more empirically yet also conceptually consistent ontology of mental features. That is all I try to do when I put forward relation-based ontology and world-brain relation as ontological predispositions of mental features. Finally, on a more practical basis, psychiatric patients often raise questions about their standing the world and their own self as their tacit common sense assumptions are shattered by the disease. That is a moment where my philosophical background can be very helpful for them to find meaning and explanations. And finally, my strong phenomenological background and inclinations are extremely helpful in trying to understand the experiences of these patients – it is, after all, the experiences rather than the symptoms why these patients come to psychiatrists. They want to understand what is going on; the symptoms are secondary attempts to compensate for a more basic underlying disturbance in the experience of oneself and the world. Taken in this way, psychiatric disorders are disturbances of consciousness which, as I say, can be traced to an altered world-brain relation.

PP: *As far as psychiatry is concerned, we would also like to ask about the process of its "neuronalization" (visible, for example, in the discussion on DSM and RDoC categories). What do you think about this discussion?*

GN: My answer to this is clear. Compare this problem with genetics. Twenty years ago genetics was treated as the big breakthrough for psychiatric disorders. What has been delivered? No clear genetic pathways. What instead has been found that the genetic expression is strongly dependent upon the respective genetic and, even more important, environmental context – this led to epigenetics. It is now equally naïve to assume "neuronalization" of these disorders. This may be hip and trendy now, yield careers and money, but scientifically it is already clear now that it is a dead-end. What we need to understand is the necessary and thereby intrinsically neuro-ecological basis of the brain's neuronal activity. For instance, when you listen to music, you unconsciously tap your feet in the rhythm of the music – this means that your brain must somehow automatically, e.g., by default, align its own neuronal activity to the ongoing activity of the music. Since it is based on shared temporal (and spatial) features, such alignment can be described as "temporo-spatial alignment". We have not yet well understood the mechanisms underlying our brain's temporo-spatial alignment to its environmental contexts. That is important, as it is here where, as I postulate, the main change in mental disorders can be found. To answer your question, we need a neuro-ecological rather than neuronal classification of mental disorders.

PP: *Finally, you declare that your purpose and vision is to reveal the key mechanisms that underlie the generation of mental states and the self. Are you an optimist regarding of further development of knowledge about the*

Extremes can tell us about the average. Hence, one strong source of inspiration of my ontology and epistemology in philosophy is psychiatry; it provides validation and plausibility for any ontological assumption: if the latter is not in accordance with the former, my ontological assumption may be conceptually plausible but not empirically consistent

I CONCEIVE THE MENTAL FEATURES LIKE SELF AND CONSCIOUSNESS TO BE PARADIGMATIC INSTANCES OF A MORE BASIC AND BROADER SENSE OF SUBJECTIVITY

nature of subjective phenomena like self and consciousness? In one of your article, you discuss the conception of First-Person Neuroscience, there are several similar projects (e.g. heterophenomenology, neurophenomenology etc.). Can the fusion of the first-person perspective and neuroscientific data provide us a consistent conception of consciousness?

GN: Yes, that is the project of my research if not of my life. In one sentence. My aim is to find out where and how subjectivity comes from in our world. Hence, I conceive the mental features like self and consciousness to be paradigmatic instances of a more basic and broader sense of subjectivity. Given my ontological framing, it is then only natural to raise the question how processes and relations in the world relate to the brain, e.g., the world-brain relation, in such way that the brain's neuronal activity can transform into mental activity. Hence, I conceive the world-brain relation as a necessary ontological (and empirical) condition of possible neuro-mental transformation. This may sound strange in your ears as the question for mental features and their neuronal basis is usually not framed in this way. Rather than asking for neuro-mental transformation, we usually asked for the neural correlates of mental features like in the neural correlates of consciousness (NCC). This presupposes directionality from mental to neuronal which I reverse as I prefer starting from the neuronal (and ultimately the ecological) to the mental – only when presupposing the latter directionality one can raise the question for neuro-mental transformation whereas it is nonsensical in case one presupposes directionality from mental to neuronal-like in the NCC. This may all sound even more strange. However, consider other disciplines like biology, physics, and chemistry. They focus on processes of transformation, that is, how state A transforms into state B. In order for such transformation from A to B possible, A and B must share some common features as without that A could not transition to and thus transform into B. This is what I call in a recent paper "Common currency" of neuronal and mental features as it must underlie their transformation, e.g., neuro-mental transformation. In my more optimistic moods, I would claim that the identification of the "common currency" of neuronal and mental features should provide one central piece in the puzzle of our search for the neuronal (and ontological) basis of mental features and, more generally, subjectivity. Now the answer to your question. I hypothesize that temporo-spatial dynamics provides the "common currency" of neuronal and mental features: neuronal features show temporo-spatial dynamics which, as I postulate, is transformed into and manifest in the subjective experience of time and space in consciousness, e.g., "spatiotemporality' like William James' 'stream of consciousness' and Husserl's conception of protention, presentation, and retention. To grasp and account for such spatiotemporality on the mental level of consciousness and self (and other mental features), we require a first-person account. To link that first-person account of spatiotemporality to the brain's temporo-spatial dynamics, we require first-person neuroscience – the latter (and related conceptions as you indicate) is thus a methodological tool to investigate my hypothesis of temporo-spatial dynamics providing the "common currency" of neuronal and mental features. And, as I postulate that finding out and identifying the "common currency" will provide a consistent or, as I say, a plausible conception of consciousness in both ontological and empirical (and also epistemological) terms.

Bibliography:

Craver C. (2009), *Explaining the Brain*, Oxford University Press, Oxford.

Northoff G. (2004). *Philosophy of the Brain: The Brain Problem*. Advances in Consciousness Research.

Northoff G. (2013). *Unlocking the Brain: Volume 2: Consciousness*. Oxford University Press.

Northoff G. (2014). *Minding The Brain: A Guide to Philosophy and Neuroscience*. Red Globe Press.

Northoff G. (2018). *The Spontaneous Brain From the Mind–Body to the World–Brain Problem*. The MIT Press

Northoff G., Heinzel A. (2006). First-Person Neuroscience: A new methodological approach for linking mental and neuronal states. *Philosophy Ethics and Humanities in Medicine:* 1(1).

Walter (2018). Trzecia fala biologicznej psychiatrii. *Piąte Piętro Bydgoskie Czasopismo Filozoficzne* 3/2018: 13-31.