Synthese (2021) 199:2701–2725
https://doi.org/10.1007/s11229-020-02907-6

What is neurophilosophy: Do we need a non-reductive form?

Philipp Klar[1,2]

Received: 24 April 2020 / Accepted: 8 October 2020 / Published online: 17 October 2020

Abstract

Neurophilosophy is a controversial scientific discipline lacking a broadly accepted definition and especially a well-elaborated methodology. Views about what neurophilosophy entails and how it can combine neuroscience with philosophy, as in their branches (e.g. metaphysics, epistemology, ethics) and methodologies, diverge widely. This article, first of all, presents a brief insight into the naturalization of philosophy regarding neurophilosophy and three resulting distinguishable forms of how neuroscience and philosophy may or may not be connected in part 1, namely reductive neurophilosophy, the parallelism between neuroscience and philosophy which keeps both disciplines rather strictly separated and lastly, non-reductive neurophilosophy which aims for a bidirectional connection of both disciplines. Part 2 presents a paradigmatic example of how these three forms of neuroscience and philosophy approach the problem of self, mainly concerning its ontological status (existence and reality). This allows me to compare all three neurophilosophical approaches with each other and to highlight the benefits of a non-reductive form of neurophilosophy. I conclude that especially non-reductive neurophilosophy can give full justice to the complementary position of neurophilosophy right at the intersection between neuroscience, philosophy, and psychology.

Keywords Neurophilosophy · Philosophy of mind · Philosophy of neuroscience · Consciousness · Self

✉ Philipp Klar
Philipp.Klar@hhu.de

1 Cécile and Oskar Vogt Institute of Brain Research, Heinrich Heine University Düsseldorf, Düsseldorf, Germany

2 Mönchengladbach, Germany

1 Introduction and an overview of the distinct forms of neurophilosophy

Neurophilosophy is a scientific discipline connecting neuroscience and philosophy and that intends to research former genuine philosophical topics, such as the ancient and major topics of consciousness, the self, and free will. These philosophical topics faced the enormous development of imaging-methods (neuroimaging) in the last past 35–40 years, hence resulting in an increasing interest of neuroscience in them which allows different kinds of interaction between both disciplines today. To chronologically introduce the development of each form of neurophilosophy, a threefold differentiation between reductive neurophilosophy, parallelism between neuroscience and philosophy, and non-reductive neurophilosophy will be defined in the perspective of the following main principles concerning the possible connection of neuroscience with philosophy:

– Naturalization of philosophy;
– Branches of philosophy and linkage to empirical sciences;
– Philosophical and empirical methodology; and
– Stance towards the brain and mind or consciousness.

Since naturalization of philosophy is a prerequisite for the connection of neuroscience with philosophy and therefore neurophilosophy, it shall be highlighted in the following part 1.1, while the threefold differentiation between reductive neurophilosophy (part 1.2), parallelism between neuroscience and philosophy (part 1.3), and non-reductive neurophilosophy (part 1.4) follow subsequently.

1.1 Naturalization of philosophy as a prerequisite for neurophilosophy

Naturalization of philosophy stands as a prerequisite to enable the connection of empirical science, namely neuroscience, with philosophy. In a first instance, the differentiation between empirical science and philosophy in a classical sense is necessary so that consequently it becomes more comprehensible how the strict classical dissociation of both disciplines is in principle dissolvable via the naturalization of philosophy.

Philosophy, in a classical sense, qualifies as an a priori analytic science that mainly operates on the rational-argumentative basis of linguistic concepts which are primarily focused on logical conditions within imaginable possible worlds. The main branches of philosophy among others are metaphysics, epistemology, ethics, and phenomenology. Linguistic concepts are used to explain philosophical topics, problems, and the therewith connected approach. On the other hand, empirical sciences are classified as a posteriori and synthetic; that is, they are based on the observational-experimental methodology and investigation in third-person-perspective.[1] This scientific methodology of empirical sciences focuses primarily on processes and mechanisms that underly phenomena within the natural and real world. In other words, it focuses more on the *how* in the sense of functionality instead of on the *what* in the sense of ontology (existence

[1] The terms a priori analytic and a posteriori synthetic used here are based on the definitions of Kant's transcendental philosophy (1781/1996).

and reality) as usually pursued in philosophy. The observational-experimental investigation may then provide certain possible inferences to the underlying processes and mechanisms of phenomena. Altogether in a classical perspective, empirical sciences and philosophy differ completely from each other as diametral confronted extremes regarding their branches and methodologies. As long as this categorical distinction is maintained, neurophilosophy does not become an option.

However, in the middle of the twentieth century, American philosopher Willard van Orman Quine (1908–2000) elaborated a possible naturalization of philosophy and stated that the principal distinction between empirical sciences and philosophy is not reasonable (Quine 1951, 1969). Quine (1951, 1969) further argued that philosophical linguistic concepts and the rational-argumentative methodology within logical conditions and reasoning can be seen as an *abstraction* of scientific results and its observational-experimental methodology. This allows a *mutual quantitative continuum* between empirical sciences and philosophy to open up, and in which genuine empirical sciences and genuine philosophy are represented solely by their respective endpoint on that continuum (Fig. 1). Based on this reasoning, Quine (1951, 1969) argued on the following three fundamental levels:

1. There is a continuum between analytic and synthetic sentences;
2. There is a continuum between a priori and a posteriori reasoned knowledge; and
3. Consequently, a mutual continuum exists between empirical sciences and philosophy.

Following the introduction of the naturalization of philosophy, the categorical distinction between analytic and synthetic sentences as previously defined by Kant (1781/1996), was replaced.[2] Rather than a categorial distinction between analytic and synthetic sentences, Quine (1951, 1969) preferred a quantitative continuum between both, which he explained in detail within his writings.

Naturalization of philosophy can be achieved in two different major ways that touch the implementation of neurophilosophy: (1) *replacement naturalism* and (2) *cooperative naturalism*. In replacement naturalism, the empirical observational-experimental methodology strongly dominates or even overrules the conceptual-logical methodology of classic philosophy. This gives rise to reductive neurophilosophy (part 1.2), which is common in Anglo-American countries today, as represented by the Churchlands (1981, 1985, 1989, 2013b) among others. Therefore, philosophy, including its branches and methodology, is ultimately reduced to the empirical realm. Cooperative naturalism, on the other hand, avoids this reduction by allowing a bidirectional interaction between the branches and methodologies of empirical sciences and philosophy, which results in a non-reductive neurophilosophy, and thus a truthful interdisciplinary interaction. Hence cooperative naturalism is a necessary prerequisite for non-reductive neurophilosophy (part 1.4), which may allow for a more comprehensive perspective

[2] According to Kant (1781/1996), an *analytic sentence* already contains the predicate (P) within the subject (S), and therefore the predicate does not offer any further information: S = P, whereas a *synthetic sentence* is empirical: it is based on experience. He further elaborated that *synthetic sentences* a priori are possible and are especially highlighted as a prerequisite fundament concerning a reinterpretation of metaphysics (Kant 1781/1996). Kant's examples for synthetic a priori sentences refer to mathematics and physics by Newton among others (Kant 1781/1996; Kutschera 2006).

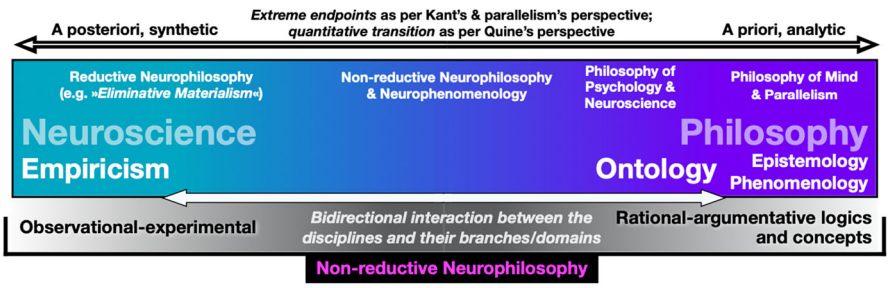

Fig. 1 Empirical sciences and naturalized philosophy are located on the same *mutual continuum*. The branches/domains and methodologies (observational-experimental vs. rational-argumentative) of both disciplines are thus faced with a possible union. Naturalized philosophy consequently allows for an interdisciplinary and systematical bidirectional interaction between neuroscience and philosophy, as represented by non-reductive neurophilosophy (part 1.4), to become possible

on the concrete phenomena of investigation. The naturalization of philosophy, including the main principles concerning the possible connections between neuroscience and philosophy, is summarized in Table 1. These main principles will be discussed subsequently within parts 1.2–1.4 concerning the threefold differentiation of possible connections between neuroscience and philosophy.

1.2 Reductive neurophilosophy

The term neurophilosophy was explicitly shaped for the very first time in the year 1986 by Canadian philosopher Patricia S. Churchland (1943–) in her eponymous book»*Neurophilosophy*« (Churchland 1989). Together with her husband Paul Churchland (1942–), who is also a philosopher, she developed a strong reductive neurophilosophy (»*Eliminative materialism*«) which states that philosophy, including its branches consisting of metaphysics, ontology, epistemology, ethics, phenomenology, etc., is ultimately reduced to the observational-experimental methodology and empirical research of neuroscience, and that philosophy, as well as folk psychology, will be reduced more and more into neuroscience as the latter advances in its scientific research (Bickle 2006; Churchland 1981, 2013a). Reductive neurophilosophy is especially common in the Anglo-American countries based on their understanding of neurophilosophy as a discipline today (Bickle 2003, 2009, 2019).

The abstract principles and practical methodological approach of the Churchlands (1989, 2002a, b) are heavily shaped by the application of the observational-experimental methodology within the empirical level of neuroscience to former genuine philosophical topics, and consequently reflect the *replacement naturalism* form of the naturalization of philosophy. This approach ultimately concludes that this form of neurophilosophy can, therefore, be considered as reductive, as similarly to neuroscience, it also takes a *brain-reductive stance*: the person and her or his (self-)conscious phenomenal first-person-perspective's experience with its aspects like point-of-view, intentionality, sense-of-self, sense-of-agency, that is especially phenomenology itself, are reduced to the neuronal activity of the brain. The mind or

Table 1 Reductive and non-reductive neurophilosophy, as well as genuine philosophy, are distinguishable by main principles concerning naturalization of philosophy, hence possible forms of neurophilosophy

Discipline	Reductive neurophilosophy	Non-reductive neurophilosophy	Philosophy (of mind)
Naturalization of philosophy	Replacement naturalism	Cooperative naturalism	–
Branches of philosophy	Reduction of philosophical branches to the empirical domain of neuroscience	Bidirectional interaction of philosophical branches with empirical neuroscience	Focus on genuine philosophical branches
Methodology	Classical philosophical methodology, i.e., rational-argumentative concerning linguistic concepts within natural conditions of imaginable worlds, is reduced to the observational-experimental one of neuroscience	A bidirectional interaction of empirical facts concerning the natural and real world with logical concepts; natural conditions and plausibility are weighted more important than mere logical ones	A priori concepts and logics concerning conceivable possible worlds; a primary focus on logical conditions and plausibility
Stance towards the brain and mind or conscious-ness	*Brain-reductive* E.g. consciousness and the self exhibit no ontological status; mental features are either reducible to the neuronal activity or even eliminated in favor of the former	*Brain-based* Neurophenomenal linkage: correspondence between neuronal activity and pheno-menology; e.g. consciousness is based on the brain and its relation to the body and world	Mind-based appro-aches are still very common; e.g. a metaphysical mind is presupposed as the vantage point for investigations
The notion of the concept of the self	Phenomenal experience of the self is caused and ultimately entirely reducible to the brain's neuronal activity. The self inhibits no ontological status	The self's existence and reality are provided by a relational structure between the brain, body, and the environment in opposite to entity or property based ontologies	Principally different but frequently metaphysical accounts of the self are discussed

A *fluent overlap*, that is a quantitative continuum rather than a categorical difference, regarding the main principles above between the three disciplines, exists. Consequently, the table reflects an ideal differentiation between the presented three forms. It especially has to be pointed out that empirical neuroscience reaches out more and more towards philosophical aspects, as in cognitive neuroscience, while on the other side, philosophy of mind today focuses more on the brain and its functionality since neuroscience heavily advanced the past 20–30 years. As a result, certain quantitative overlaps between the three disciplines of neuroscience, neurophilosophy, and philosophy of mind exist today

consciousness does not correspond to the neuronal level of the brain, but instead, it is considered to be reducible to it. Moreover, according to Churchland's (2002a) metaphysics of eliminative materialism, the mind (e.g. consciousness) does not have an ontological status (existence and reality). Instead, consciousness, the self, and mental features are scientifically *eliminated* and are replaced by a complete focus on the neuronal level of empirical neuroscience in favor of an isolated observation of only the brain. This brain-reductive stance is also well reflected in the denial of the existence and reality of the self as stated by the German philosopher Thomas Metzinger (2004) in his book»*Being No One*«, just as Patricia Churchland (2013b) considers the self illusionary and to be nothing but the brain.

Therefore, reductive neurophilosophy applies a *unidirectional inference* from mere empirical data and findings to philosophical concepts which concludes that the empirical level of neuroscience strongly overrules philosophical branches and their respective concepts, since concepts are unidirectionally created and adapted to empirical data and interpretations. This unidirectional inference shall be demonstrated with an example: since classic philosophical concepts of the self, such as a substance in form of a mental entity (Descartes 1641/1993), cannot be found as a physical entity within the brain, philosophers like Metzinger (2004) and Churchland (2013b) consequently consider the self to be non-existent. In other terms, the self has no ontological status. According to the two philosophers, the self does not exist because inside the *mere empirical realm of neuroscience*, the philosophical concept of a self as a substance is not directly in itself findable, and hence its ontological status needs to be eliminated.[3] Instead of adapting or creating entirely new concepts of the self in accordance and matching correspondence with empirical data, reductive neurophilosophy primarily infers from sole empirical data and facts to the existence and reality of philosophical concepts. The foundation of concepts thus exclusively relies on their empirically neuroscientific plausibility within natural conditions, while conceptual plausibility concerning logical conditions is either neglected or even ignored. Finally, reductive neurophilosophy also dismisses philosophically generated concepts *as input* for scientific investigations and research; by contrast, it creates concepts unidirectionally as mere outputs from empirical data and facts, so that philosophical concepts are left as *entire empirical induced outputs*.

Additionally, today's philosophy of neuroscience can be subsumed under the umbrella of reductive neurophilosophy as it discusses principles and methodological aspects of plain neuroscience (Bechtel et al. 2001), similarly as the philosophy of psychology critically discusses the methodology of psychology (Bermúdez 2005). The reductive form of neurophilosophy as discussed above is also subsumed under the broader umbrella of the philosophy of neuroscience by American philosopher John Bickle (2019). According to Bickle (2019), in contrast to the philosophy of neuroscience, neurophilosophy has "fallen" from its initial vision and aims of revolutionizing philosophy by explicitly introducing neuroscientific research and its implications to philosophy as a discipline. However, famous approaches that avoid a rather complete reduction of philosophical concepts, branches, and its methodology, such as neurophe-

[3] Whereas other concepts of the self, for example as relational constituted structure between the brain, body, and environment (Northoff 2013b, 2014c, 2016a, 2018b), or as the mind, i.e., conscious experience itself (Vacariu 2016), can be per with neuroscientific data.

nomenology or non-reductive neurophilosophy (part 1.4), are neither considered nor even mentioned within Bickle's criticism. The formerly elucidated reductive notion of neurophilosophy is now so prominent in the Anglo-American countries, that other forms of neurophilosophy which avoid today's reductionism seem to be non-existent among the corresponding academical philosophy of science's circles.

Furthermore, neurophilosophy has to be separated from the philosophy of mind, which especially asks about the existence and reality of the mind and the latter's relationship to the matter. The topics of philosophy of mind are rather basically of an analytic-metaphysical nature, i.e., they involve mainly the mind–body problem and the consequences which result in different positions regarding the latter (Brüntrup 2018; Kutschera 2006, 2009; Newen 2013). After introducing the widely known reductive neurophilosophy, a strict parallelism between neuroscience and philosophy, which denies any form of bidirectional interaction between both, hence denying the possibility of neurophilosophy at all, will be presented.

1.3 Parallelism between neuroscience and philosophy

Besides the possible forms of neurophilosophy, there is also the conviction that strict *parallelism* between the empirical realm of neuroscience and philosophy is required (Bennett and Hacker 2003). Maxwell Bennett (1939–), an Australian neuroscientist, and Peter Hacker (1939–), an English philosopher (Philosophy of Language, Philosophy of Mind, and an expert for the philosophy of Wittgenstein), published the book»*Philosophical Foundations of Neuroscience*« together in 2003. The book covers comprehensive analysis and criticism of *cognitive* neuroscience, with particular reference to how cognitive neuroscientists accidentally mislead themselves. First of all, Bennett and Hacker (2003) argue that a wrong and *confusing usage of terms and concepts* concerning empirical investigation is very common especially in cognitive neuroscience. Furthermore, there are conceptually confused interpretations of findings since a conceptual-theoretical confused input will lead to even more confusing investigations and results (Bennett and Hacker 2003). For example, Bennett and Hacker (2003) state that this was the case for the first generation of modern neuroscientists in the twentieth century, when they either explicitly argued in favor of ontological substance dualism between the brain and mind, like neurophysiologist Charles S. Sherrington (1857–1952) and neurosurgeon Wilder Penfield (1891–1976), or they demonstrated implicitly an unintentionally induced substance dualism by conceptual confusion, as for example by neuroscientist Edgar D. Adrian (1889–1977).

According to Bennett and Hacker (2003), a far more subtle yet erroneous neo-Cartesianism lives on in cognitive neuroscience today. In other terms, (self-)consciousness, free will, and mental features or psychological attributes like attention, memory, knowledge or sense-of-agency, are considered as exclusive brain functions in form of distinct entities or processes, which in turn shall be reducible to the brain's neuronal activity in the neuroscientist's perspective. Bennett and Hacker (2003) instead argue that the mind in general and the distinct mental features in particular, are nothing but *capabilities and behavioral executions* of the organism as a whole and not of the brain. The attribution of mental features and psychological attributes to the brain

would represent the especially pointed out *mereological fallacy*, which is a part-whole confusion. The reification of the above capabilities, i.e., as mind or mental features, is simply wrong and instead, the latter has to be seen merely as a linguistic expression of these capabilities.

According to Bennett and Hacker (2003), it is especially and fundamentally important to thereby distinguish between *scientific empirical* and *conceptual questions*. Concerning neuroscience and philosophy, this means that neuroscience is constrained to research the brain in a strict empirical manner throughout empirical scientific questions, while philosophy and its respective branches focus on genuine conceptual questions concerning the mind, e.g. (self-)consciousness, mental features or psychological attributes. Consequently, Bennett and Hacker (2003) refrain from merging a bidirectional interaction between neuroscience and philosophy since neuroscience must concentrate solely on the empirical realm and its observational-experimental methodology, while philosophy should focus exclusively on the definition of concepts, terms, and categories including their elaboration in distance to the empirical realm. Hence, any form of neurophilosophy is simply not an option. On the contrary, Bennett and Hacker (2003) argue in favor of a classical branch and methodological monism regarding each discipline. Furthermore, philosophy is not able to generate real new knowledge as it is the case in empirical sciences, but instead, philosophy is principally and most widely a linguistic-logic based analytic science which allows for precise verification and reflection concerning human knowledge; e.g. what knowledge was obtained through empirical sciences, after confused concepts, terms and categories were revealed and revised (Bennett and Hacker 2003; Hacker 2010).

However, philosophy is nevertheless allowed to at least suggest exactly defined concepts and terms, which are not to be confused with possible topics of investigation for the empirical research in neuroscience, as well as to provide interpretations and verifications concerning the question if neuroscientists interpret their findings and empirical data wrong, particularly when applying these empirical findings to philosophical concepts. An example is linguistic confusions and conceptual fallacies, like the already mentioned *mereological fallacy* that was widely pointed out to be very present in today's neuroscience, precisely when psychological predicates are attributed to the brain, instead of to the organism as a whole. This is the limit that philosophy can offer to neuroscience (Bennett and Hacker 2003). On the contrary, real and genuine philosophical problems and hence therewith connected topics, e.g. in connection to neuroscience, do not exist according to Bennett and Hacker (2003) but are especially induced by linguistic confusions. According to Hacker (2010), this is also particularly the case for consciousness studies.

This strict separation of neuroscience and philosophy is also partially present in today's philosophy of mind whenever the focus of the investigation is extensively laid on genuine philosophical branches like metaphysics, while empirical data and findings are not significantly included. This is especially the case when philosophical investigations focus on concepts concerning conceivable possible worlds and their inherent logical plausibility (instead of their empirical plausibility of natural conditions regarding the real and natural world). This is well reflected by the common presupposition of a metaphysical mind, which consequently leads to the question of how the mind is related to the matter, hence maintaining the mind–body problem.

Instead of challenging the question if a metaphysical mind exists at all, philosophy of mind typically takes the mind for granted and then starts its investigation upon it. In summary, it is now possible that philosophy of mind is partially subsumed under the umbrella of parallelism between neuroscience and philosophy, since the philosophy of mind can be considered more on the side of genuine philosophy in comparison with neurophilosophy when considering a quantitative continuum (Fig. 1) in Quine's perspective (1951, 1969), between empirical sciences and philosophy.

1.4 Non-reductive neurophilosophy

While Churchland (1989) introduced the term neurophilosophy with a reductive-eliminative imprint, she was not the first person in the history of philosophy to practice neurophilosophy. Non-reductive neurophilosophy originates further back in the nineteenth century. In the year 1818, 29 years old Arthur Schopenhauer (1788–1860) finished his main work»*Die Welt als Wille und Vorstellung*« (The World as Will and Representation) which was published in 1819, and in which he took the vantage point of Kant's philosophy by interpreting his a priori categories and forms of intuition as brain functions; i.e., not a mental entity like the mind shall be responsible for the subjective-phenomenal experience of the first-person-perspective, but the brain (Schopenhauer 1819/2011). Using the above, Schopenhauer introduced the brain explicit into the philosophical investigation which led to a *brain-based* approach instead of a *mind-based* approach (as it is still common in genuine philosophy today) to neurophilosophy. Hence, he can be considered to be the very first neurophilosopher ever (Northoff 2018a; Göhmann 2018). It took over a hundred years more, particularly until the middle of the twentieth century, before non-reductive neurophilosophy was realized implicitly once again. The French phenomenological philosopher Maurice Merleau-Ponty (1908–1961) can also be considered as an early neurophilosopher who likewise introduced the brain to philosophy whilst connecting the brain and the body (accounting for embodiment) to perception and phenomenology. Based on their respective arguments and approaches, it can be implicitly considered that both Schopenhauer and Merleau-Ponty were against a reductive-eliminative approach as put forward by the Churchlands (Merleau-Ponty 1945/2013; Schopenhauer 1819/2011). Subsequently, in the second half of the twentieth century, more precisely in the year 1977, Australian neuroscientist John C. Eccles (1903–1997) and Austrian-British philosopher Karl R. Popper (1902–1994) came up with a different approach in their famous book»*The Self and Its Brain*« (Popper and Eccles 1985). Popper and Eccles (1985) argued in favor of ontological substance dualism between the mind and brain, more specifically trialism, whose explanation is beyond the aim of this article.[4]

As a first modern approach to combine neuroscience with philosophy in bidirectional interaction using the heavily increasing development of neuroscience, Chilean neuroscientist and philosopher Francisco Varela (1946–2001) found-

[4] It is worthwhile to mention that most of the famous modern neuroscientists of the first generation in the twentieth century, like Charles S. Sherrington (1857–1952) or the neurosurgeon Wilder Penfield (1891–1976), explicitly or implicitly argued in favor of ontological substance dualism (Bennett and Hacker 2003, 2012; Penfield 1975).

ed»*Neurophenomenology*« in the 1990s, which can be considered as non-reductive neurophilosophy. Neurophenomenology presents a methodological strategy that takes a vantage point from phenomenology, that is the conscious experience of the first-person-perspective, to especially research consciousness and its connection to the neuronal level of the third-person-perspective of the observational-experimental empirical research of neuroscience (Khachouf et al. 2013; Lutz and Thompson 2003; Varela 1996). Neurophenomenology by Varela (1996) seriously considers subjective-phenomenal experience, for example, shaped by the aspects of intentionality, self, point-of-view, and sense-of-self/agency, as non-illusionary and real so that phenomenology is not considered to be simply reducible to neuronal activity in the brain. Varela (1996) especially contemplated *embodiment* regarding consciousness. In embodiment, the brain's sensorimotor functions in direct connection to the body, and its linkage to the environment is viewed as a major constituting factor for consciousness. Furthermore, at the end of the 1990s, more precisely in the year 1998, the investigation into the topic of free will by German physician and philosopher Henrik Walter (1962–) in his book»*Neurophilosophy of Free Will*« (Walter 1998/2009) can be emphasized as a truthful neurophilosophical approach towards an original philosophical topic.

Another modern and more advanced approach for a *bidirectional interaction* between neuroscience and philosophy, consequently leading to non-reductive neurophilosophy which neither aims for a reductionist engulf of philosophy to neuroscience nor aims for an absolute distinction without interaction between both disciplines, stems from the German physician (psychiatrist), neuroscientist and philosopher Georg Northoff (1963–). Northoff (2012) points out that the brain is *undisciplined*, i.e. the borders between the distinctive scientific disciplines of philosophy, neuroscience, psychology, psychiatry, etc. are ultimately artifacts of the human mind that shall be overcome by interdisciplinary scientific research.

Considering the four main principles relevant to the connection of neuroscience with philosophy as listed in part 1 and Table 1, non-reductive neurophilosophy, first of all, represents the *cooperative naturalism* type of the naturalization of philosophy; i.e., philosophical branches and their respective concepts are not reduced to the empirical realm of neuroscience. Northoff's (2004, 2014a, b, c, 2018b, 2019a) approach applies a bidirectional connection of both sciences. Philosophical concepts require empirical evidence; i.e., concepts need to be established on the empirical level within natural conditions. Their *empirical plausibility,* as Northoff (2004, 2014a) terms it, is primarily weighted over their mere logical plausibility within the borders of genuine philosophy and its logical conditions concerning conceivable possible words. In conclusion, a *domain and methodological pluralism* (Northoff 2014a) becomes possible and hence introduces truthful neurophilosophical investigations.

Referring to the bidirectional interaction between neuroscience and philosophy, *concept-fact iterativity* (Northoff 2014a) stands out as the main principle of non-reductive neurophilosophy which shall now be presented and further elaborated. As described above in part 1.2, reductive neurophilosophy unidirectionally infers concepts solely from empirical data and evidence as mere output while non-reductive neurophilosophy starts the investigation of a certain topic with its prior philosophical concept. First of all, philosophical concepts offer an *input* for empirical research which

may be followed by *empirically plausible modification* of a priori established philosophical concepts, while on the other side there is also the possibility to put the results of empirically conceptualized and operationalized concepts as well as their resulting neurophilosophical investigation back as *output* into philosophy to evaluate their conceptual plausibility in a second step. Proceeding from this empirical-theoretical interaction, a modified neurophilosophical investigation may follow so that concepts and their bidirectional connected empirical facts and modification according to the latter pass through the research-loop of concept-fact iterativity, thus allowing for a converging interdisciplinary approach to reality, as the empirically modified concepts are in return reviewed for their logical plausibility as well.

Northoff (2014a) mentions that in a genuine philosophical perspective the principle of concept-fact iterativity may reflect a category error. According to classical philosophy, empirical facts cannot be connected with logical argumentation, and respectively, natural and logical conditions require a strict separation. However, neurophilosophy primarily strives for empirical plausibility of concepts, which is then connected with logical plausibility. According to Northoff (2014a), while empirical plausibility is valued to be fundamentally important, logical plausibility is *not* neglected. However, at the same time, it is not possible to infer unidirectionally from mere empirical data and facts to ontological postulations. Such unidirectional inferences would correspond to an *empirical-ontological fallacy* as Kant (1781/1996) pointed out concerning British physician and philosopher John Locke (1690/1996). Instead, a matching-process between the empirical and philosophical realm, particularly between empirical facts and corresponding ontological assumptions, is required. Concept-fact iterativity thus reflects a principle of branch pluralism between empirical sciences and philosophy: the branches of metaphysics, epistemology, ethics, etc. are systematically connected with empirical facts, instead of merely investigating into either only empirical or logical plausibility.

Concerning the brain, non-reductive neurophilosophy takes a *brain-based* stance: even though consciousness, the self, and mental features are based on the brain, the latter is only a necessity but is not a sufficient condition for these. Taking the temporo-spatial theory of consciousness (TTC) (Northoff 2013a, 2014b, c, 2016a, b, 2018b; Northoff and Huang 2017) as an example, consciousness, the self, and mental features are based on a relational structure between the brain, body, and environment conceptualized as *empirical-ontological*»*World-brain relation*« which entails embodiment and embeddedness. Without going into the details of the empirical and philosophical aspects of this theory, it ought to be mentioned that consciousness and the self are here considered to have an ontological status. They are existent and real but they correspond neither to a physical entity nor to a mental entity as it is the case in common property-based ontologies. Instead, consciousness and the self, including mental features, are fundamentally based on empirical-ontological relations between the brain, body, and world, hence forming a balanced *structure,* which is then for example altered in neuropsychiatric disorders such as depression, schizophrenia or mania. These neuropsychiatric disorders are situated on the more extreme ends of a hybrid relational continuum, than being located in the more healthy and functioning centered areas (Northoff 2014c, 2016a, b, 2018b; Northoff and Tumati 2019).

2 Forms of neuroscience and philosophy: distinct approaches to the topic of the self as a paradigmatic presentation

After briefly introducing three forms of how or how not to connect neuroscience and philosophy in part 1, it is now possible to paradigmatically conceptualize and operationalize their respective approaches to the original and formerly genuine philosophical topic of the self regarding their presented main principles (Table 1). These divergent methodologies consequently entail different notions and arising concepts of the self to both the empirical and/or the philosophical realm.

Philosophy discusses the topic of the self, e.g. regarding its existence and reality, for centuries. It is this very ontological status of the self, more precisely the question if the self is real at all, which represents the hereby chosen example of how forms of neuroscience and philosophy practically differ in their approaches to genuine philosophical concepts when facing empirical sciences. The sequence of these presentational approaches is equal to part 1: reductive neurophilosophy's approach to the self sets the beginning (2.1); which is then followed by the parallelism between neuroscience and philosophy (2.2); and finally, non-reductive neurophilosophy concludes the paradigmatically approaches (2.3).

2.1 Reductive neurophilosophy and its approach to the self

> Nobody ever *was* or *had* a self. All that ever existed were conscious self-models that could not be recognized *as* models. […] subjective experience of *being someone* emerges if a conscious information-processing system operates under a transparent self-model (Metzinger 2004, p.1).

As elaborated in part 1, one main principle of reductive neurophilosophy is the reduction of philosophical concepts, which arise from the rational-argumentative method of analytic reasoning within logical conditions, to the observational-experimental methodology of empirical neuroscience. Already at the onset of the investigation, reductive neurophilosophy is defined by a specific methodological step, particularly by taking a *vantage point from within the empirical realm of neuroscience* (Churchland 1989, 2002a). More precisely, none of the many established philosophical concepts of the self is chosen as heuristical input, instead, the investigation starts with a specific and only empirical phenomenon. For example, abnormalities of (self-)consciousness in the neuropsychiatric disorders of depression and schizophrenia, that is their altered or diminished first-person-perspective's subjective sense-of-self, may serve as a starting point to investigate into the concept in question. Since a genuine philosophical concept as input is missing, the investigation into the topic of the self fully shifts from the formerly metaphysical, ontological, epistemological and phenomenological realms to only the empirical realm of neuroscience.

lready with this initiating step, the self is transformed into a matter of empirical research that reflects a rather complete replacement naturalism of the relationship between neuroscience and philosophy, more precisely with a strong focus on general biological functions of the brain's neuronal activity. Any possible and resulting concept of the self is therefore *unidirectionally inferred from plain empirical facts and*

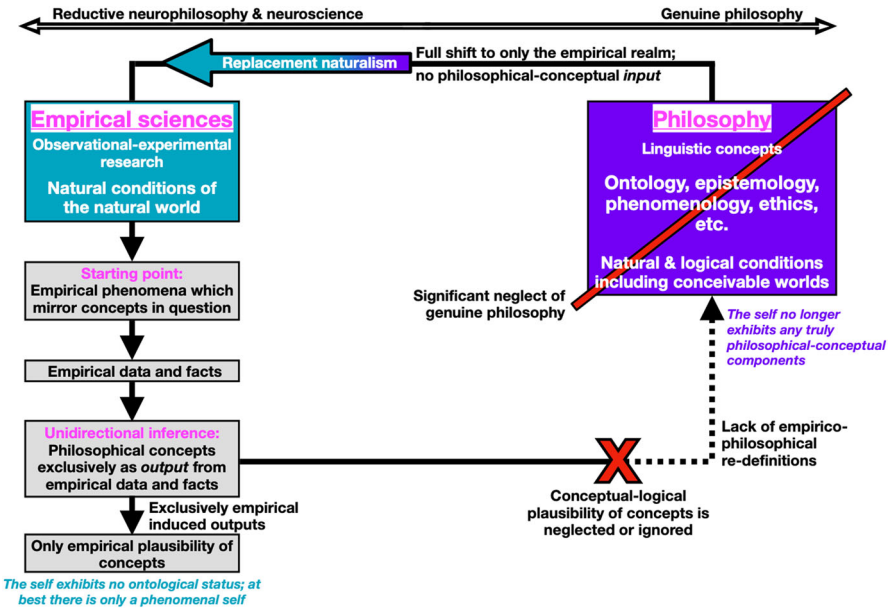

Fig. 2 Reductive neurophilosophy shifts its investigation fully into the empirical realm of neuroscience. Firstly, no significant input by philosophy of the concept in question is given for empirical research. Consequently, empirical sciences already represent the starting point of the investigation. Secondly, empirical phenomena that are believed to mirror the philosophical concept in question are investigated by only the observational-experimental methodology of neuroscience. Lastly, philosophical concepts are accordingly and unidirectionally inferred from mere empirical data without further and sufficient philosophical consideration including critical reflection, that is regarding their conceptual-logical plausibility and implications within the branches of ontology, epistemology, phenomenology, etc.

data, which are obtained by the observational-experimental methodology of neuroscience. Beyond that, there is no further sufficient philosophical consideration of the empirically induced concept, neither regarding its conceptual-logical plausibility nor regarding its philosophical implications, e.g. concerning its ontological or epistemological significance (Fig. 2).

However, scientific research now faces a major "problem" concerning the self: it is not possible to find a distinct self such as an "object", "core", substance, or entity within the mere empirical realm of neuroscience. This problem derives from the *brain-reductive stance*, i.e., reductive neurophilosophy's complete and isolated focus on only the brain. Within the brain, only its neuronal activity, which is electrical action-potentials and biochemical substances between chemical synapses (neurotransmitters), be it on the molecular, cellular, or the area and network level are detectable. Even though neuroscience offers various empirical concepts and effects such as self-referential reflection (D'Argembeau et al. 2005), self-reference effects (SRE) (Klein 2012) or self-referential processing (D'Argembeau 2013; Liu et al. 2014; Knyazev 2013) among many others, a self defined as a traditional physical or even a mental entity or property is simply neither detectable nor deductible by using only the

empirical methodology when investigating exclusively into the isolated brain and its biological functionality.

Reductive neurophilosophers like Thomas Metzinger (1999, 2004, 2009) correspondingly infer that any former traditional philosophical concept of the self, especially defined as a mental entity, is simply a false inference from the phenomenal experience of the self, i.e., originating from (self-)consciousness as in sense-of-self or sense-of-agency, to the ontological and underlying reality. According to Metzinger's (1999, 2001, 2004, 2009) representational theory and naturalization of (self-)consciousness, phenomenal so-called»*self-models*« developed over phylogenesis and are caused by neuronal activity. Metzinger (1999, 2001, 2004, 2009) does not deny the immediate phenomenal experience of the self but instead denies any underlying ontological reality or status of the self, more precisely because the only ontological reality is only the brain including its body. The brain causes a self-model that is "transparent" to us–we principally cannot experience the fact that the self *is* a model within our phenomenological naive realism. Hence, real selves do not exist and this is culminated in eliminating all ontological characterizations of the self and the title of his famous book»*Being No One*« (Metzinger 2004). On the grounds of the above, formerly philosophical concepts of the self are then eliminated in favor of the empirical reality of only the brain (Churchland 2002b, 2013b; Metzinger 2004, 2009). Consequently, new conceptual definitions of the self (no-self theories) originate solely as output from the mere empirical realm.

However, this reductive approach makes it obvious that *concepts still implicitly frame the empirical starting-point into data and facts*. Firstly, the self is a philosophical concept which was implicitly given as a conceptual frame; and secondly, only a specific concept of the self was implicitly presupposed and is then rejected on the grounds of the obtained empirical data and facts which are not in accordance to the self as this particular entity. In reductive neurophilosophy's conclusion, no reality of the self, neither as a mental entity nor as a physical entity or property, truly exists–ontologically conceived, the self is considered to be an illusion.

2.2 Parallelism between neuroscience and philosophy and its approach to the self

> It should be evident that the philosophical conception of self-consciousness not only deviates from the common or garden notions but is also a product of philosophical confusions rooted in the notion of apperception transmitted from Locke to Leibniz and from Wolf to Kant (Hacker 2013, p. 57).

In the perspective of a rather strict parallelism between neuroscience and philosophy, neither replacement nor cooperative naturalism between empirical sciences and philosophy becomes an option in the light of truthful interdisciplinary collaboration. When taking the example of the self into reflection, this clear-cut stance of parallelism is equivalent to the notion that the self, both concerning its ontological status as well as towards its phenomenological aspects, is only a philosophical topic that cannot be investigated by neuroscience in principal. The investigation of the self within the empirical realm of neuroscience would reflect nothing but a category error. Firstly, there would be confusion between scientific empirical and conceptual questions in

general. Secondly and more precisely, a confusion of capabilities and behaviors of the organism and person as a whole with empirical data and facts of specific brain functions would occur.

Most fundamentally in the perspective of parallelism, problems which require both empirical (neuro-)sciences and philosophy to be solved do not even exist, more accurately because topics and problems which allegedly span across the disciplines are nothing but errors which are *initially induced by conceptual confusions already inherent in the philosophical realm* and then transferred to empirical sciences. Bennett and Hacker (2003) exemplify that philosophical misconceptions also involve the case of (self-)consciousness, especially proceeding from the notion of the self as an entity by Descartes (1641/1993) to the self as a psychological feature which is supposed to be accessible via introspection by Locke (1690/1996), over to the corresponding notion of a phenomenal self in present-days cognitive neuroscience and its relation to specific brain regions and networks (Damasio 1999, 2000, 2010; Frewen et al. 2020; Gazzaniga 2000, 2005; LeDoux 2003; Panksepp 1998, 2003; Turk et al. 2003; Wolff et al. 2018). These postulations which account for any additional self within consciousness and related questions concerning the underlying ontological status of such a self are fallacious and meaningless (Bennett and Hacker 2003; Hacker 2007, 2013).

It is a misconception, specifically a mereological fallacy, to ask how the pure physical brain can have a state, i.e., in form of a distinct entity, of (self-)consciousness or how the latter can arise from the brain's neuronal activity (Bennett and Hacker 2003; Hacker 2013). This is so precisely because it is the living being as a whole which exists and which is conscious. Consciousness is a capability that is inherent within the living being and that the physical brain lacks on the contrary. (Self-)consciousness is a linguistically expressed capability of the human being (Bennett and Hacker 2003; Hacker 2007, 2013). Consequently, the search for neuronal correlates of (self-)consciousness is simply meaningless because based on the grounds of the above, the self is a linguistically induced concept of which no immediate correspondence within the physical brain exists, as a result of the fact that the self does not exist. Conceptual confusions about the self's existence and reality, including its phenomenological aspects, need to be eliminated right from the onset of research. More precisely, misconceptions need to be detected and eliminated already within the philosophical realm. Otherwise, the neuroscientist will investigate topics and problems whose implicit or explicit presupposition as input is already erroneous. Consequently, any following interpretation of empirically induced outputs will be nothing but a result of misguided research and faulty interpretations of empirical data and facts which are not related to the real concepts or phenomena in question.

Since the enterprise of neurophilosophy combines neuroscience and philosophy, its approach is doomed right from the beginning in the parallelism's perspective. Therefore, a neurophilosophy of the self cannot have a stand. In the framework of parallelism, philosophy does not create genuine new knowledge concerning the self or any empirical facts, instead, philosophy provides a better *understanding* of already established knowledge and experience concerning the way humans think about themselves and the world (Bennett and Hacker 2003; Hacker 2010). For example, a better understanding of (self-)conscious experience, including its grammatical and linguistic expressions, may shed light on how the latter relates to conceptual confusions towards

the self, e.g. its existence and reality or the "self's" phenomenological aspects (which are in fact experiences of a conscious living being and only linguistically mediated expressions of a self).

Most fundamentally, in the perspective of Bennett and Hacker (2003), real philosophical problems do not exist. Philosophical problems solely occur based on confusions within specifically presupposed conceptual schemes. Hence, the enterprise of neurophilosophy, including its investigation into neuroscience to advance questions, e.g. concerning the self, is not just erroneously, that is it would require correction and could then be properly investigated, but it is completely meaningless right from the onset. No matter how promising and complex these enterprises appear and how special their resulting outcomes seem to be, such misconceptions, e.g. about the self, ultimately become obvious once their erroneous presuppositions are carefully analyzed and revealed. Regarding the parallelism between neuroscience and philosophy, the paradigmatically chosen question if the self is real at all is a good example of such a misdirected enterprise. All that is required is to avoid and dissolve initial misconceptions so that erroneous investigations, both in neuroscience and philosophy, are prevented. This is one possible contribution of philosophy to neuroscience. As a result, neuroscience and philosophy cannot operate in any immediate and merging interaction, instead, they require a strict separation from each other (Fig. 3). In conclusion, the ontological characterization of the self is erroneous. The self's ontological characterization is a misconception that already arose in the traditional philosophical realm and now resurfaces in neuroscience as well as in neurophilosophy (Bennett and Hacker 2003; Hacker 2007, 2013).

2.3 Non-reductive neurophilosophy and its approach to the self

> […] such concept of self as structure and organization is embodied, e.g., intrinsically linked to the body, and embedded, e.g., intrinsically linked to the environment. Hence, the virtual structure of the self spans across the brain, body, and environment with the brain's midline structure activity being a neural predisposition for its constitution, while at the same time being dependent upon the respective environmental context (Northoff 2013a, b, p. 11).

A most fundamental principle of non-reductive neurophilosophy is reflected by its pluralism of branches which consequently entails methodological pluralism between neuroscience and philosophy. Firstly, right at the onset of the investigation, non-reductive neurophilosophy's starting-point is defined by considering philosophical concepts, e.g. of the self, concerning their ontological determination. On one side, this distinguishes non-reductive neurophilosophy from its reductive variant, as the reductive approach conceives the empirical realm, including its data and facts, as a starting point. On the other side, philosophical concepts as input distinguish non-reductive neurophilosophy from the parallelism between neuroscience and philosophy, since the parallelism considers philosophical concepts not as starting-point, but as a realm by itself which is completely separate from neuroscience.

Accordingly, non-reductive neurophilosophy chooses *specific and genuine philosophical concepts (e.g. of the self) as input*. Consequently, a specific philosophical

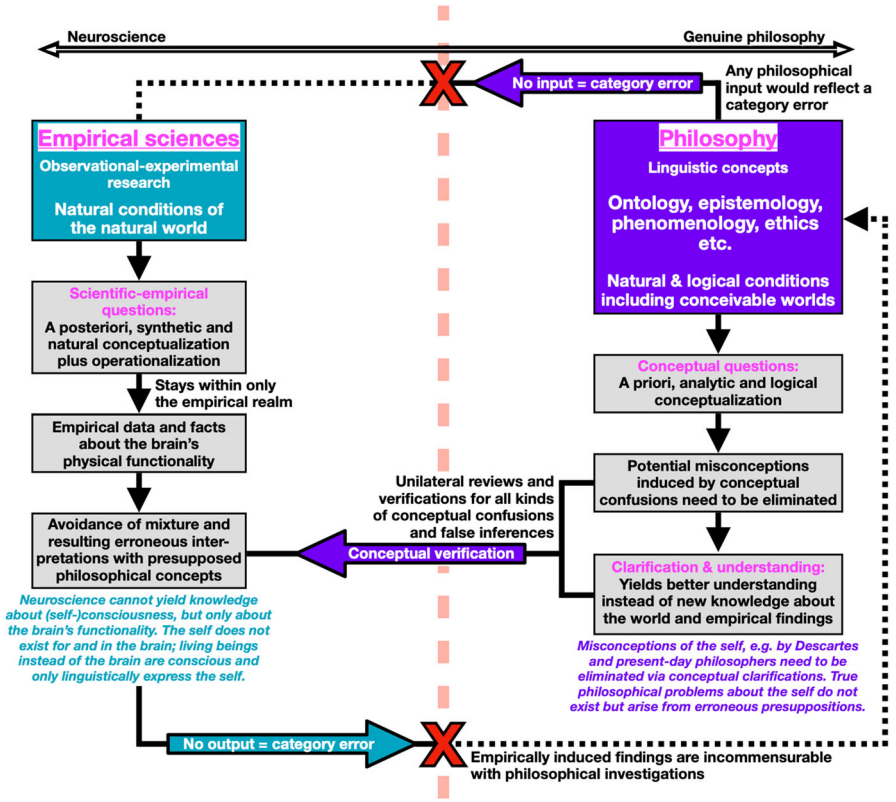

Fig. 3 In the perspective of Bennett's and Hacker's (2003) parallelism, neuroscience and philosophy require a principal and most basic separation as individual disciplines. It is not just their methodology which completely differs (observational-experimental vs. conceptual-linguistic), but the categories in which the topics of investigation fall. Neuroscience, as part of the empirical realm, researches the brain's functionality, which is bio-physiological processes within the brain and body. On the contrary, philosophy's aim is the creation of precise concepts and clarification of already established knowledge about ourselves and the world. Furthermore, philosophy offers clarification of findings that originate from empirical sciences. Hence, philosophy, unlike science, does not create new knowledge, but better understanding. Neuroscience cannot contribute to philosophical knowledge because true philosophical problems do not exist. However, philosophy can unilaterally correct misconceptions, such as wrong interpretations of empirical data and facts, made by neuroscientists. That is, not the data and facts themselves change, but the corresponding flawed concepts which served as input and/or erroneous interpretations and which represent the output of empirical research

concept of the self has to be empirically conceptualized and/or operationalized to the extent that its empirical observational-experimental investigation becomes possible. Therefore and within this step of research, a strong focus on empirical data and facts subsequently allows for validation concerning the empirical plausibility of the philosophical concept (e.g. the self as entity or structure) in question. The empirically obtained data and facts can then be taken into a matching-process with the specifically chosen philosophical concept, e.g. of the self.

In this further step of matching-process, the specific concept of the self eventually requires a re-definition, which is per empirical data and facts, hence approving and ensuring its empirical plausibility. However, precisely this step is likewise provided by reductive neurophilosophy, specifically when philosophical concepts are unilaterally adapted to empirical facts as Searle (1999, 2004) favors in a weak version of reductive neurophilosophy, or if concepts are completely deduced from the empirical realm of neuroscience without philosophical input as per Churchland's (1989, 2013b) strong reductive approach.

On the contrary, non-reductive neurophilosophy, as developed by Northoff (2014a), goes two steps further. (1) The now re-defined empirically plausible and formerly genuine philosophical concept is put back into the philosophical realm where it additionally faces its validation in respect to its logical-conceptual plausibility. Furthermore, implications regarding the distinct branches of philosophy, i.e., the re-defined concept's implications for the ontological, epistemological, or phenomenological realm, allow for wide-ranging philosophical considerations up to investigations. (2) The finally resulting empirico-philosophical re-defined concept of the self can now be taken as another starting-point for renewed and advanced deepening research. It re-enters a loop of bidirectional empirical-conceptual investigation between neuroscience and philosophy, hence reflecting cooperate naturalism in general and especially what Northoff (2014a) labels as *concept-fact iterativity* in particular. Such concept-fact iterativity represents truthful neurophilosophical research and resulting in interdisciplinary developed concepts.

From the initial onset to the preliminary end of the investigation, concepts include both neuroscientific research as well as a philosophical reflection *both as input and output into both directions*. (1) Firstly, there is an initiating philosophical-conceptual input for neuroscience; (2) neuroscience then returns an empirically plausible concept as output; (3) this output serves as input for neurophilosophical re-definition and investigation; and finally (4) the interdisciplinary re-defined neurophilosophical concept is taken as the vantage point for further investigation within new research-loops. This concept-fact iterativity consequently guarantees that neither the self stays remains as a sole philosophical concept lacking empirical data and facts, which match and correspond to a specific re-defined concept of the self, nor that philosophical concepts are completely reduced to neuroscience, as a consequence of only using the plain observational-experimental methodology of neuroscience. Altogether, there is philosophical input of the self to the empirical realm, which consequently ensures an empirico-philosophical output of interdisciplinary re-defined concepts of the self, as well as a constant loop of freshly developed concepts into a further and deepening investigation regarding future research (Fig. 4).

While it is beyond the scope of this article to present a full-blown elaboration of the non-reductive approach to the self, nevertheless, an overview shall be provided in comparison to its reductive sibling which is nowadays rejected by Bickle (2019). When empirically investigating the brain's neuronal functionality, the self as a physical, or even a mental entity, is certainly not traceable. If the investigation would stop at this point, since the presupposed *narrow* framework of reductive neurophilosophy does only consider the isolated brain, the conclusion would be correct that there is no empirical (neuronal) mirror of the self's hereby implicitly chosen ontological deter-

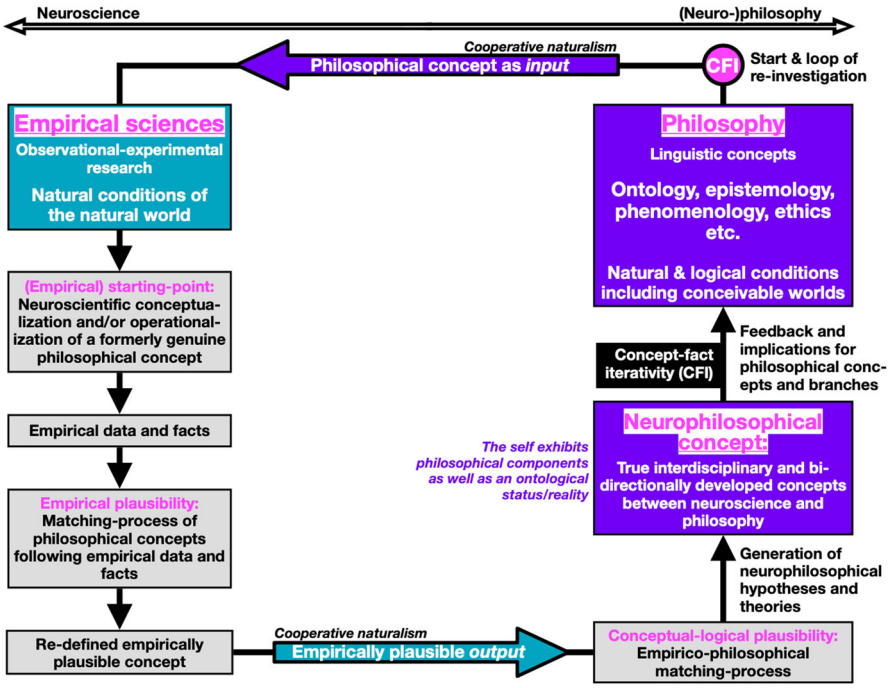

Fig. 4 Non-reductive neurophilosophy aims for a bidirectional, and hence a truthful interdisciplinary connection of neuroscience and philosophy. Firstly, genuine philosophical concepts serve as input for subsequent empirical conceptualization and/or operationalization. Secondly, the empirical plausibility of philosophical concepts is verified within a matching-process to empirical data and facts. Consequently, a re-defined empirically plausible concept arises, which is then given as output and is additionally investigated concerning its philosophical plausibility, which is its conceptual-logical one. Thus real neurophilosophical concepts become possible, which are lastly placed back into the philosophical realm and its respective branches such as ontology or epistemology. These completely re-defined interdisciplinary concepts can then serve as a starting point for further investigations. This research loop allows us to increasingly determine concepts concerning both empirical and conceptual plausibility and hence within a broader framework that non-reductive neurophilosophy covers

mination as an entity or property-based concept. So far, reductive neurophilosophers like Metzinger (2004, 2009) and Churchland (2013b) are correct insofar that the self's ontological determination as a physical or as a mental entity seems absent in the brain's neuronal activity. Nevertheless, unlike reductive approaches, non-reductive neurophilosophy does not rely on the straightforward elimination of the self's ontological status in favor of unidirectional inferences from plain empirical data and facts of an isolated observed brain to the denial and rejection of philosophical concepts.

Consequently, other concepts of the self need to be firstly provided as philosophical input for the empirical realm of neuroscience, and secondly developed within a bidirectional enterprise, thus resulting in a broader framework consisting of both neuroscientific investigation and philosophical reflection. Instead of the common entity and property-based ontologies, structure- or process-based ontologies may better reflect the empirical reality of the brain's functionality. Recent neuroscien-

tific findings speak in favor of a *neuro-ecological structure* of the brain's empirical reality, which accounts for the concept of»*World-brain relation*« (Northoff 2018b, 2019a).[5] Most basically, the brain's temporo-spatial structure of its spontaneous activity's dynamics has to align itself to the wider temporo-spatial context of the world on adaptational grounds. Such constant alignment of the brain to the world virtually spans the temporo-spatial structure across the brain, body, and environment, hence reflecting a neuro-ecological structure (Northoff 2013a, 2018b, 2019a; Northoff and Huang 2017). The brain's neuro-ecological structure and world-brain relation require relational based ontologies instead of entity-based ontologies. This could amount to structural realism (SR), more precisely moderate ontic structural realism (OSR) (Esfeld and Lam 2008, 2011) which assumes that relations and structures are ontologically more fundamental than relata/elements. OSR is also favored by the non-reductive neurophilosophical approach to the brain and (self-)consciousness (Northoff 2018b).

Conceiving the brain in this *broader* framework (compared to the narrow and reductive framework of an isolated brain) may then opens the door for the possibility of the self's ontological determination. In other terms, these obtained empirical data and facts then serve as output for further philosophical reflection and implications concerning the self within philosophical branches. Following empirical findings and philosophical reflection, both mind based as well as reductive-eliminative concepts of the self is rejected and replaced by a structural determination of the self. Ultimately, such a truthful neurophilosophical concept is chosen as the starting point regarding forthcoming bidirectional empirico-philosophical research. This re-investigation loop particularly reflects the non-reductive principle of concept-fact iterativity (Northoff 2014a).

In summary, non-reductive neurophilosophy takes a *brain-based stance* (in opposite to a brain-reductive stance of reductive neurophilosophy) which absolutely includes the brain but also goes beyond it by taking the world in respect to the brain's functionality as well as mental features into consideration. Accordingly, the feedback-loop system of non-reductive neurophilosophy's research as described above searches for a "common currency" as the linkage between mental features and the brain's neuronal activity (Northoff 2019a, b; Northoff et al. 2019). A reduction of both subjective first-person phenomenal experience including the ontological determination of the self to only the brain's empirical functionality is thus rejected.[6]

Consequently, (self-)consciousness as well as specific mental features are considered to hold an ontological status within the perspective of non-reductive neurophilosophy, constituted by the relational neuro-ecological structure between the brain, body (accounting for embodiment) and the environment (accounting for embeddedness), which is conceptualized and ultimately traced back to the

[5] In perspective of the self, hereof involved are especially the overlapping cortical midline structures (CMS) and the default-mode-network (DMN); in accordance to the empirical phenomena of self-relatedness, both the CMS and DMN are neuroscientifically associated with the self (Northoff 2013a; Qin and Northoff 2011; Scalabrini et al. 2018; Qin et al. 2013; Qin et al. 2016; Wolff et al. 2018).

[6] Phenomenal experience of the self, i.e., what phenomenology defines as»*ipseity*« of the»*experiential self, core self* or *minimal self* « (Gallagher 2000; Parnas and Henriksen 2019; Zahavi 2005, 2014, 2019), that is immediate and intrinsically melt of a basic sense-of-self within the stream-of-consciousness, is not reduced to the brain's neuronal activity within the approach of non-reductive neurophilosophy.

empirical-ontological»*World-brain relation*« within the Temporo-spatial theory of consciousness (TTC) (Northoff 2014b, c, 2016a, b, 2018b; Northoff and Huang 2017). Such ecological view of the brain contradicts a brain-reductive stance, as the latter claims that consciousness, the self, and mental features are reducible to and especially *caused* by the brain's neuronal activity, which is still commonly presupposed in theories about consciousness by empirical neuroscience, as seen in the Integrated Information Theory (IIT) (Tononi 2004, 2008; Tononi and Koch 2008; Tononi et al. 2016), the Global Neuronal Workspace Theory (GNWT) (Baars 2005; Baars and Franklin 2007) and reductive neurophilosophy (Churchland 1985, 1989, 2002a, b, 2013a, b).[7] On the contrary in the perspective of non-reductive neurophilosophy, (self-)consciousness including mental features are neither seen as reducible to the brain's neuronal activity nor caused by the latter. Instead, an intrinsic correspondence, that is, a neuro-mental transformation by the brain's temporo-spatial dynamics between neuronal activity and mental features, is suggested. Therefore, the distinction into two distinct entities as well as the reduction from one level to the other is rejected (Northoff et al. 2019).

3 Conclusion

Even though Patricia Churchland (1989) explicitly introduced the term neurophilosophy into the academic discourse of philosophy and its possible connection with neuroscience more than 30 years ago, neither widely accepted abstract principles of neurophilosophy nor the methodologies concerning its practical implementation about neurophilosophical research exist as at date. Therefore, the chosen threefold differentiation in the article between reductive neurophilosophy, non-reductive neurophilosophy, and parallelism as a strict separation between the disciplines of neuroscience and philosophy, presented a brief insight into today's distinguishable perspectives on the project of neurophilosophy. While parallelism between neuroscience and philosophy denies the possibility of a merging collaboration between their branches and methodologies, even in the light of fascinating results and possibilities that neuroscience developed especially within the last 25 years, non-reductive neurophilosophy reaches out for exactly this bidirectional interaction: a neuro-phenomenological linkage consisting of neuroscientific third-person-perspective data with corresponding first-person-perspective's experience of (self-)consciousness is one of its aims, leading to a broader understanding of consciousness in general as well as specific mental features in particular, and therefore ultimately of human existence. Correspondingly, the project of non-reductive neurophilosophy goes along without a reductionism of ourselves to the brain's neuronal activity. As paradigmatically presented, this approach also applies to neurophilosophical inspired research on the topic of the self. While philosophy in the past focused on many contrasting concepts of the self, e.g. as mental substance (Descartes, 1641/1993), which are most widely rejected today, or as the distinction between the subjective "I" and objective "me" (James

[7] Since this causal relationship between neuronal activity and mental features implies that both are sort of distinct entities, it may very well reflect the criticized neo-Cartesianism in current neuroscience by Bennett and Hacker (2003) as well as by German psychiatrist and philosopher Thomas Fuchs (2018).

1890a, b), it is common in contemporary reductive neurophilosophy and neuroscience to reduce or eliminate the phenomenal self in favor of the brain (Churchland 2013b; Metzinger 2004, 2009). In other terms, the self is only empirically conceptualized, e.g. as a higher-order cognitive function (Churchland 2002b; Damasio 1999, 2000; Dennett 1991), without further and sufficient philosophical consideration, including respective implications. This one-sided notion on the self consequently leads to significant neglect of phenomenological aspects of the self and its present-day frequent ontological denial, similar to the self's rejection by Hume (1739–1740/2003), so that wide parts of neuroscience, and especially reductive neurophilosophy, reverted into the other one-sided extreme in form of a neuronal reductionism. Non-reductive neurophilosophy, however, takes both neuroscience and philosophy seriously for any field of investigation, e.g. in respect to the self (Northoff 2014b, 2016c, 2018b, 2019a, b), therefrom reflecting a brain-based stance and cooperative naturalism form of the naturalization of philosophy (rather than a brain-reductive and replacement naturalistic stance). In conclusion, non-reductive neurophilosophy does not stand in competition with neuroscience and philosophy, instead, its approach should be seen as complementary to genuine particular sciences and it especially preserves philosophy alive by actively considering and taking philosophical concepts into the interdisciplinary investigation, that is, both as input and output.

Funding Open Access funding enabled and organized by Projekt DEAL.

References

Baars, B. J. (2005). Global workspace theory of consciousness: Toward a cognitive neuroscience of human experience. *Progress in Brain Research, 150,* 45–53.

Baars, B. J., & Franklin, S. (2007). An architectural model of conscious and unconscious brain functions: Global Workspace Theory and IDA. *Neural Networks, 20*(9), 955–961.

Bechtel, W., Mandik, P., Mundale, J., & Stufflebeam, R. S. (Eds.). (2001). *Philosophy and the neurosciences: A reader*. Oxford: Wiley.

Bennett, M. R., & Hacker, P. M. S. (2003). *Philosophical foundations of neuroscience*. Oxford: Blackwell.

Bennett, M. R., & Hacker, P. M. S. (2012). *History of cognitive neuroscience*. Oxford: Wiley.

Bermúdez, J. L. (2005). *Philosophy of psychology. A contemporary introduction*. New York: Routledge.

Bickle, J. (2003). *Philosophy and neuroscience: A ruthlessly reductive account*. Dordrecht: Kluwer Academic Publishers.

Bickle, J. (2006). Reducing mind to molecular pathways: Explicating the reductionism implicit in current cellular. *Synthese, 151*(3), 411–434.

Bickle, J. (Ed.). (2009). *The Oxford handbook of philosophy and neuroscience*. New York: Oxford University Press.

Bickle, J. (2019). Lessons for experimental philosophy from the rise and "fall" of neurophilosophy. *Philosophical Psychology, 32*(1), 1–22.

Brüntrup, G. (2018). *Philosophie des Geistes. Eine Einführung in das Leib-Seele-Problem*. Stuttgart: Kohlhammer.
Churchland, P. M. (1981). Eliminative materialism and the propositional attitudes. *The Journal of Philosophy, 78*(2), 67–90.
Churchland, P. M. (1985). Reduction, qualia and the direct introspection of brain states. *The Journal of Philosophy, 82*(1), 8–28.
Churchland, P. S. (1989). *Neurophilosophy: Toward a unified science of the mind–brain*. Cambridge, MA: MIT Press.
Churchland, P. S. (2002a). *Brain-wise: Studies in neurophilosophy*. Cambridge, MA: MIT Press.
Churchland, P. S. (2002b). Self-representation in nervous systems. *Science, 12*(296), 308–310.
Churchland, P. M. (2013a). *Matter and consciousness*. Cambridge, MA: MIT Press.
Churchland, P. S. (2013b). *Touching a nerve. The self as brain*. New York: W.W. Norton.
Damasio, A. R. (1999). How the brain creates the mind. *Scientific American, 281*(6), 112–117.
Damasio, A. R. (2000). *The feeling of what happens. Body and emotion in the making of consciousness*. Boston, MA: Mariner books.
Damasio, A. R. (2010). *Self comes to mind. Constructing the conscious brain*. New York: Pantheon Books.
D'Argembeau, A., Collette, F., Van der Linden, M., Laureys, S., Del Fiore, G., Delgueldre, C., et al. (2005). Self-referential reflective activity and its relationship with rest: A PET study. *NeuroImage, 25*(2), 616–624.
D'Argembeau, A. (2013). On the role of the ventromedial prefrontal cortex in self-processing: The valuation hypothesis. *Frontiers in Human Neuroscience, 7,* 372.
Dennett, D. C. (1991). *Consciousness explained*. Boston: Little, Brown and Company.
Descartes, R. (1641/1993). *Meditations on first philosophy*. Indianapolis: Hackett Publishing Company.
Esfeld, M., & Lam, V. (2008). Moderate structural realism about space-time. *Synthese, 160*(1), 27–46.
Esfeld, M., & Lam, V. (2011). Ontic structural realism as a metaphysics of objects. In A. Bokulich & P. Bokulich (Eds.), *Scientific structuralism* (pp. 143–159). Dordrecht: Springer.
Frewen, P., Schroeter, M. L., Riva, G., Cipresso, P., Fairfield, B., Padulo, C., et al. (2020). Neuroimaging the consciousness of self: Review, and conceptual-methodological framework. *Neuroscience and Biobehavioral Reviews, 112,* 164–212.
Fuchs, T. (2018). *Ecology of the brain*. New York: Oxford University Press.
Gallagher, S. (2000). Philosophical conceptions of the self: Implications for cognitive science. *Trends in Cognitive Sciences, 4*(1), 14–21.
Gazzaniga, M. S. (2000). *The mind's past*. Berkeley, CA: University of California Press.
Gazzaniga, M. S. (2005). Forty-five years of split-brain research and still going strong. *Nature Reviews Neuroscience, 6*(8), 653–659.
Göhmann, D. (2018). Neurophilosophie. In D. Schubbe & M. Koßler (Eds.), *Schopenhauer-Handbuch* (pp. 375–378). Stuttgart: J. B. Metzler.
Hacker, P. M. S. (2007). *Human nature. The categorial framework*. Oxford: Blackwell Publishing.
Hacker, P. M. S. (2010). Hacker's challenge. *The Philosophers' Magazine, 51*(51), 23–32.
Hacker, P. M. S. (2013). *The intellectual powers. A study of human nature*. New York: Wiley.
Hume, D. (1739–1740/2003). *A treatise of human nature*. Mineola, NY: Dover Publications.
James, W. (1890a). *The principles of psychology* (Vol. 1). New York, NY: Holt.
James, W. (1890b). *The principles of psychology* (Vol. 2). New York, NY: Holt.
Kant, I. (1781/1996). *Critique of pure reason*. Indianapolis: Hackett Publishing Company.
Khachouf, O. T., Poletti, S., & Pagnoni, G. (2013). The embodied transcendental: A Kantian perspective on neurophenomenology. *Frontiers in Human Neuroscience, 7,* 611.
Klein, S. B. (2012). Self, memory, and the self-reference effect: An examination of conceptual and methodological issues. *Personality and Social Psychology Review, 16*(3), 283–300.
Knyazev, G. G. (2013). EEG correlates of self-referential processing. *Frontiers in Human Neuroscience, 7,* 264.
Kutschera, F. (2006). *Die Wege des Idealismus*. Paderborn: Mentis.
Kutschera, F. (2009). *Philosophie des Geistes*. Paderborn: Mentis.
LeDoux, J. (2003). *Synaptic self. How our brains become who we are*. New York: Penguin Books.
Locke, J. (1690/1996). *An essay concerning human understanding*. Indianapolis: Hackett Publishing Company.
Liu, J., Corbera, S., & Wexler, B. E. (2014). Neural activation abnormalities during self-referential processing in schizophrenia: An fMRI study. *Psychiatry Research: Neuroimaging, 222*(3), 165–171.

Lutz, A., & Thompson, E. (2003). Neurophenomenology integrating subjective experience and brain dynamics in the neuroscience of consciousness. *Journal of Consciousness Studies, 10*(9–10), 31–52.
Merleau-Ponty, M. (1945/2013). *Phenomenology of perception*. London: Routledge.
Metzinger, T. (1999). *Subjekt und selbstmodell*. Paderborn: Mentis.
Metzinger, T. (Ed.). (2001). *Bewusstsein*. Mentis: Beiträge zur Gegenwartsphilosophie. Paderborn.
Metzinger, T. (2004). *Being no one. The self-model theory of subjectivity*. Cambridge, MA: MIT Press.
Metzinger, T. (2009). *The ego tunnel. The science of the mind and the myth of the self*. New York: Basic Books.
Newen, A. (2013). *Philosophie des Geistes. Eine Einführung*. München: Verlag C. H. Beck.
Northoff, G. (2004). What is neurophilosophy? A methodological account. *Journal for General Philosophy of Science, 35*(1), 91–127.
Northoff, G. (2012). *Das disziplinlose Gehirn—Was nun, Herr Kant? Auf den Spuren unseres Bewusstseins mit der Neurophilosophie*. München: Irsania.
Northoff, G. (2013a). What the brain's intrinsic activity can tell us about consciousness? A tri-dimensional view. *Neuroscience and Biobehavioral Reviews, 37*(4), 726–738.
Northoff, G. (2013b). Brain and self—a neurophilosophical account. *Child and Adolescent Psychiatry and Mental Health, 7,* 28.
Northoff, G. (2014a). *Minding the brain. A guide to philosophy and neuroscience*. Basingstoke: Palgrave Macmillan.
Northoff, G. (2014b). *Unlocking the brain. Volume 1: Coding*. New York: Oxford University Press.
Northoff, G. (2014c). *Unlocking the brain. Volume 2: Consciousness*. New York: Oxford University Press.
Northoff, G. (2016a). Spatiotemporal psychopathology I: No rest for the brain's resting state activity in depression? Spatiotemporal psychopathology of depressive symptoms. *Journal of Affective Disorders, 190,* 854–866.
Northoff, G. (2016b). Spatiotemporal psychopathology II: How does a psychopathology of the brain's resting state look like? Spatiotemporal approach and the history of psychopathology. *Journal of Affective Disorders, 190,* 867–879.
Northoff, G. (2016c). *Neuro-philosophy and the healthy mind. Learning from the unwell brain*. New York: W.W. Norton.
Northoff, G. (2018a). Neurophilosophy and Neuroethics: Template for Neuropsychoanalysis? In H. Boeker, P. Hartwich, & G. Northoff (Eds.), *Neuropsychodynamic psychiatry* (pp. 599–615). Berlin: Springer.
Northoff, G. (2018b). *The spontaneous brain. From the mind–body to the world-brain problem*. Cambridge, MA: MIT Press.
Northoff, G. (2019a). Lessons from astronomy and biology for the mind-Copernican revolution in neuroscience. *Frontiers in Human Neuroscience, 13,* 319.
Northoff, G. (2019b). Phenomenological psychopathology and neuroscience. In G. Stanghellini, M. R. Broome, A. V. Fernandez, P. Fusar-Poli, A. Raballo, & R. Rosfort (Eds.), *The Oxford handbook of phenomenological psychopathology* (pp. 909–924). New York: Oxford University Press.
Northoff, G., & Huang, Z. (2017). How do the brain's time and space mediate consciousness and its different dimensions? Temporo-spatial theory of consciousness (TTC). *Neuroscience and Biobehavioral Reviews, 80,* 630–645.
Northoff, G., & Tumati, S. (2019). Average is good, extremes are bad"—Non-linear inverted U-shaped relationship between neural mechanisms and functionality of mental features. *Neuroscience and Biobehavioral Reviews, 104,* 11–25.
Northoff, G., Wainio-Theberge, S., & Evers, K. (2019). Is temporo-spatial dynamics the "common currency" of brain and mind? In Quest of "Spatiotemporal Neuroscience". *Physics of Life Reviews* (**in press**).
Panksepp, J. (1998). The pre-conscious substrates of consciousness: Affective states and the evolutionary origin of the SELF. *Journal of Consciousness Studies, 5*(5–6), 566–582.
Panksepp, J. (2003). The neural nature of the core SELF: Implications for understanding schizophrenia. In T. Kircher & A. David (Eds.), *The self in neuroscience and psychiatry* (pp. 197–213). Oxford: Oxford University Press.
Parnas, J., & Henriksen, M. G. (2019). Selfhood and its disorders. In G. Stanghellini, M. R. Broome, A. V. Fernandez, P. Fusar-Poli, A. Raballo, & R. Rosfort (Eds.), *The Oxford handbook of phenomenological psychopathology* (pp. 465–474). New York: Oxford University Press.
Penfield, W. (1975). *The mystery of the mind: A critical study of consciousness and the human brain*. Princeton: Princeton University Press.
Popper, K. R., & Eccles, J. C. (1985). *The self and its brain*. Berlin: Springer.

Qin, P., & Northoff, G. (2011). How is our self related to midline regions and the default-mode network? *NeuroImage, 57*(3), 1221–1233.
Qin, P., Duncan, N. W., & Northoff, G. (2013). Why and how is the self-related to the brain midline regions? *Frontiers in Human Neuroscience, 7,* 909.
Qin, P., Grimm, S., Duncan, N. W., Fan, Y., Huang, Z., Lane, T., et al. (2016). Spontaneous activity in default-mode network predicts ascription of self-relatedness to stimuli. *Social Cognitive and Affective Neuroscience, 11*(4), 693–702.
Quine, W. V. O. (1951). Two dogmas of empiricism. *The Philosophical Review, 60,* 20–43.
Quine, W. V. O. (1969). *Ontological relativity and other essays*. New York: Columbia University Press.
Scalabrini, A., Mucci, C., & Northoff, G. (2018). Is our self related to personality? A neuropsychodynamic model. *Frontiers in Human Neuroscience, 12,* 346.
Schopenhauer, A. (1819/2011). *Die welt als wille und vorstellung*. München: Deutscher Taschenbuch Verlag.
Searle, J. R. (1999). The future of philosophy. *Philosophical Transactions of the Royal Society of London, 354*(1392), 2069–2080.
Searle, J. R. (2004). *Mind. A brief introduction*. Oxford/New York: Oxford University Press.
Tononi, G. (2004). An information integration theory of consciousness. *BMC Neuroscience, 5,* 42.
Tononi, G. (2008). Consciousness as integrated information: A provisional manifesto. *Biological Bulletin, 215*(3), 216–242.
Tononi, G., & Koch, C. (2008). The neural correlates of consciousness: An update. *Annals of the New York Academy of Sciences, 1124*(1), 239–261.
Tononi, G., Boly, M., Massimini, M., & Koch, C. (2016). Integrated information theory: From consciousness to its physical substrate. *Nature Reviews: Neuroscience, 17*(7), 450–461.
Turk, D. J., Heatherton, T. F., Macrae, C. N., Kelley, W. M., & Gazzaniga, M. S. (2003). Out of contact, out of mind: The distributed nature of the self. *Annals of the New York Academy of Sciences, 1001,* 65–78.
Vacariu, G. (2016). *Illusions of human thinking. On concepts of mind, reality, and the universe in psychology, neuroscience, and physics*. Wiesbaden: Springer Fachmedien.
Varela, F. J. (1996). Neurophenomenology: A methodological remedy for the hard problem. *Journal of Consciousness Studies, 3*(4), 330–349.
Walter, H. (1998/2009). *Neurophilosophy of free will. From libertarian illusions to a concept of natural autonomy*. Cambridge, MA: MIT Press.
Wolff, A., Di Giovanni, D. A., Gómez-Pilar, J., Nakao, T., Huang, Z., Longtin, A., & Northoff, G. (2018). The temporal signature of self: Temporal measures of resting-state EEG predict self-consciousness. *Human Brain Mapping, 40*(3), 789–803.
Zahavi, D. (2005). *Subjectivity and selfhood. Investigating the first-person perspective*. Cambridge, MA: MIT Press.
Zahavi, D. (2014). *Self and other: Exploring subjectivity, empathy, and shame*. Oxford: Oxford University Press.
Zahavi, D. (2019). Self. In G. Stanghellini, M. R. Broome, A. V. Fernandez, P. Fusar-Poli, A. Raballo, & R. Rosfort (Eds.), *The Oxford handbook of phenomenological psychopathology* (pp. 299–305). New York: Oxford University Press.