



Does artificial intelligence exhibit basic fundamental subjectivity? A neurophilosophical argument

Georg Northoff¹ · Steven S. Gouveia²

Accepted: 14 February 2024 / Published online: 26 February 2024
© The Author(s) 2024

Abstract

Does artificial intelligence (AI) exhibit consciousness or self? While this question is hotly debated, here we take a slightly different stance by focusing on those features that make possible both, namely a basic or fundamental subjectivity. Learning from humans and their brain, we first ask what we mean by subjectivity. Subjectivity is manifest in the perspectiveness and mineness of our experience which, ontologically, can be traced to a point of view. Adopting a non-reductive neurophilosophical strategy, we assume that the point of view exhibits two layers, a most basic neuroecological and higher order mental layer. The neuroecological layer of the point of view is mediated by the timescales of world and brain, as further evidenced by empirical data on our sense of self. Are there corresponding timescales shared with the world in AI and is there a point of view with perspectiveness and mineness? Discussing current neuroscientific evidence, we deny that current AI exhibits a point of view, let alone perspectiveness and mineness. We therefore conclude that, as per current state, AI does not exhibit a basic or fundamental subjectivity and henceforth no consciousness or self is possible in models such as ChatGPT and similar technologies.

Keywords Artificial intelligence · Timescales · Brain · Subjectivity

1 Introduction

There is an intense debate whether artificial intelligence (AI) exhibits consciousness or/and a self (Hildt, 2019; Northoff et al., 2022; Ng, 2023). This, obviously, depends on what is meant by AI. Here we define AI as the development

✉ Steven S. Gouveia
stevensqueira92@hotmail.com

¹ Mind, Brain Imaging and Neuroethics Research Unit, The Royal's Institute of Mental Health Research, University of Ottawa, Ottawa, ON, Canada

² Research Fellow of the Mind, Language and Action Group of the Institute of Philosophy of the University of Porto, Porto, Portugal

of computer systems and software that can perform tasks that typically require human intelligence. This includes the ability to learn from experience, adapt to new information, understand natural language, recognize patterns, and make decisions. AI covers a wide range of technologies and applications, such as machine learning, neural networks, natural language processing, and computer vision. The ultimate goal of artificial intelligence is to create systems that can simulate and replicate human cognitive abilities, allowing machines to perform complex tasks and solve problems in a manner similar to human thought processes.

Even more important is to know what is meant by self and consciousness. The debate on these issues is marred by disputes which may presuppose different notions of both AI and self/consciousness (Mindt & Montemayor, 2020; Rakover, 2023; Schneider, 2019). Our goal is to avoid some of these conceptual impasses in the current debate by taking a step back and ask about some more fundamental or basic features underlying both consciousness and self, namely subjectivity.

The concept of subjectivity has a long philosophical history. Key phenomenological features of subjectivity are a point of view (Nagel, 1974) as well as the experience of mineness and perspectiveness (Northoff, 2014a; Zahavi, 2005 and b). Are these features of such basic fundamental subjectivity (cf. Northoff & Smith, 2023) present in AI as defined above? This will be the key question addressed in our paper. The general strategy or argument is that we can take lessons from human subjectivity to investigate whether similar features are manifest in AI.

To approach this question, we will employ a non-reductive neurophilosophical methodological strategy (Gouveia, 2022; Klar, 2021; Northoff, 2004, 2014b, 2022). Specifically, we first determine the notions of “mineness” and “perspectiveness” as well as the point of view in phenomenological and conceptual terms. We next link these determination to the brain and how it interacts with the world through timescales, as we assume, thus focusing on more ontological issues. This is followed by gathering empirical support for the assumed key role of timescales in world and brain mediating the three key features of a fundamental or basic subjectivity.

The main argument structure in this paper can be semi formalized as follows:

- (1) Human subjects are characterized by a basic or fundamental subjectivity;
- (2) This basic fundamental subjectivity is reflected in perspectiveness and mineness;
- (3) Perspectiveness and mineness can be traced to and are based on the point of view;
- (4) The point of view can be characterized by its temporal relation of world and self through the brain and its timescales resulting in a most basic neuroecological layer of the point of view;
- (5) Such fundamental neuroecological layer of the point of view and the relevance of timescales are empirically supported by findings on the temporal nature of self;
- (6) AI does not show such timescales;
- (7) Therefore, AI cannot constitute temporal relation to the world through timescales;

- (8) This, in turn, makes it impossible for AI to constitute a neuroecological layer of a point of view;
- (9) The absence of such neuroecological layer of the point of view entails that AI is lacking perspectiveness and mineness;
- (10) Therefore, AI does not exhibit any signs of a basic or fundamental subjectivity.

In the next section, we will provide relevant arguments to demonstrate the soundness of our main claim: that AI does not exhibit relevant signs of having a fundamental subjectivity since it lacks neuroecological and timescales relation with the world. Our thesis will be based on a neurophilosophical approach, one that focus on both the conceptual (first part) and the empirical (second part) side of this debate.

2 Subjectivity in an ontological sense: timescales and the point of view within the world

In a first step, we want to highlight what we mean by the notion of a basic or fundamental subjectivity. This leads us briefly into phenomenological territory of experience. While we venture into that space, our view is a special as it ultimately aims anchoring subjectivity within the world itself and its relation to brain and self. Going to the boundaries of or even beyond pure experience and the phenomenological domain, we ultimately aim to make assumptions about the ontological basis of such basic or fundamental subjectivity.

2.1 Subjectivity of the self—perspectivism and mineness

According to phenomenologists such as Jean-Paul Sartre, every act of positional consciousness (i.e., being conscious of an object) is preceded by a non-positional consciousness of that act (i.e., every conscious act is pre-reflectively, non-intentionally aware of itself) (Sartre, 1956). This non-voluntary (spontaneous) and non-inferential (it is not a belief, a representation or a judgment) consciousness is what contemporary phenomenologists (Gallagher & Zahavi, 2020; Zahavi, 2005) as well as some analytic philosophers (Goldman, 1970; Nida-Rümelin, 2014) call “pre-reflective self-awareness or self-consciousness”. This means that when we are conscious of a particular content (e.g., an apple), we must be pre-reflectively aware of that act of consciousness; otherwise, there would be nothing it is like to undergo that experience (i.e., that process would remain unconscious).

In this respect, one key subjective feature is the experience of the contents of consciousness in first-person perspective, i.e., perspectivism, which refers to the fact that each content of consciousness is given through a specific point of view (cf. Northoff & Smith, 2023). The key phenomenal feature that refers to the qualitative nature of experience is what Thomas Nagel describes as “what it is like to be in a particular state for a particular subject” (Nagel, 1974). With this definition, Nagel refers to the subjective character of experience that should be distinct from any cognitive or functional notion, at least conceptually. The phenomenal character

of consciousness is deeply related to the pre-reflective self-awareness structure of consciousness, since every “what it is like to be conscious of something” is necessarily a “what it is like to be for me.” This implies a strong phenomenological and conceptual relationship between the first-person perspective and what is usually called “mineness,” i.e., the fact that I am the one having an experience.

Following this, pre-reflective self-consciousness involves also the concept of self as a subject (as distinguished from the self as an object) which is also referred to as “minimal self” (Northoff, 2014a; Zahavi, 2005). The minimal self refers to an extremely basic sense of having experiences in terms of their mineness, that is, the experience of the content of consciousness is one’s own. Finally, and most importantly, pre-reflective self-consciousness does not require any representational or higher-order self-reflection upon the conscious state that explicitly takes the self as an object, e.g., mirror self-recognition, conceptual or narrative self-consciousness (Zahavi, 2019; Zilio, 2020): we do not need to constantly check our experience to get a confirmation of who is experiencing that object.

Together, perspectivism and mineness are key phenomenal features of our phenomenal experience of the contents of consciousness in terms of pre-reflective self-consciousness. This leaves largely open where perspectivism and mineness come from, that is, how they are constituted and their basis or fundament is. This, as we assume, leads us to the concept of the point of view.

2.2 Point of view (PV) I: integration of different timescales from body and world by the brain

The notion of the point of view (PV) is often used in various disciplines but rarely explicated in philosophy. The concept of PV is pervasive in literature and theatre with different persons expressing different points of view on the same topic. Painting and photography rely on a slightly different notion of PV that involves providing access to events or objects in the world. Despite the extensive colloquial usage of PV in many disciplines, the concept is often neglected in philosophy let alone in neuroscience as there is no established theory (see Campos & Gutierrez, 2015, for a notable exception).

Yet another noteworthy exception is the embodied approach where the lived (rather than objective) body is supposed to provide the point of view or anchor of the self in the world (Gallagher & Daly, 2018; Gallagher, 2005a, b). Without being able to go into details (see Northoff, 2016, 2018), we assume that the notion of the lived body presupposes the timescales of the brain: the body’s various temporal scales are connected with those of the brain’s spontaneous activity as well as with those of the world that are processed through the brain. Hence, conceptually considered, the brain can be conceived as a multi-scale integrator as it connects the different timescales of the brain, body, and world in a scale-free way.

Before going on, we shall briefly describe what scale freeness is. Scale freeness refers to a particular distribution and relationship of longer and shorter timescales or slower and faster frequencies. Specifically, there is relatively stronger power in the slower frequencies compared to the faster frequencies which exhibit much weaker

power. In the power spectrum this leads to a curve whose degree of steepness from the slower frequencies' higher power to the faster frequencies' lower power, the degree of steepness of this curve can be measured by the power law exponent. The higher the power law exponent, the more power in the slower frequencies relatively to the lower power of the faster frequencies. Various studies showed that the power law exponent is abnormally altered in states where consciousness is lost (see Klar et al., 2023).

Now we can come back to the more conceptual realm of the relationship of scale free activity with the point of view and consciousness. If, for instance, the brain's capacity to integrate these timescales in a scale-free way is diminished, we lose consciousness as in that case, the neuro-ecological connection of brain and self to body and world is severed and thus, conceptually speaking, the point of view is lost (Zilio et al., 2021). This is further evidenced by studies showing that changes in timescales are related to changes in the level or state of consciousness (Zilio et al., 2023, Namikawa et al. 2011). The abnormal prolongation of timescales may lead to an abnormal high degree of temporal integration at the expense of temporal segregation (Wolff et al., 2022) the loss of the latter may, phenomenally, go along with the loss of differentiation between different contents and ultimately between content and background stream of consciousness (Northoff & Zilio, 2022a, b; Wolff et al., 2022). This may ultimately lead to the complete loss of the level or state of consciousness. This sets the stage for our next argument. Next, our main claim is that artificial intelligence does not have a neuroecological connection to the world and, because of that, it cannot develop its own subjective point of view.

2.3 Mental surface layer and ecological background layer

To make our argument, we introduce Campos and Gutierrez's (2015) account of point of view (PV) and then provide our own ecological and ontological extension. According to Campos and Gutierrez (2015), PV can be determined by two main features: (i) reference to mental life (including subject) and (ii) access to the world beyond the self and its mental life. Following in their footsteps, we will reformulate and rename these two features as (i*) the surface and (ii*) background layers of PV, respectively. The PV's reference to the subject and mental life is the surface layer of PV – we call this a *mental surface layer*. At the same time, the PV is situated within the world as its ultimate ontological background which renders it ecological – we therefore refer to this as an *ecological background layer*. Let us detail these two concepts in the following.

The mental surface layer of PV refers to a subject with personal and mental features; importantly, this layer provides the source or basis for the mental features of self:

In that variety of uses, the notion of point of view may have two distinct meanings. In one of them, points of view are part of a mental life. They are connected to the mental life of some subjects with a personal character. In that sense, the expression “point of view” is interchangeable with words like “view”, “opinion”, “belief”, “attitude”, “feeling”, “sentiment”, “thought”, etc.

Points of view in that sense could not exist without a subject with quite a rich mental life. (Campos & Gutierrez, 2015, p.2)

In contrast, the ecological background layer of PV is characterized by providing access to something that lies beyond the PV itself, namely the world with its ecological features. Rather than on mental states within the subject itself, the focus here is on how the subject connects and relates to the ecological features of the world. Intra-subjectivity is therefore replaced by inter-subjectivity, and isolation is replaced by relation:

There is another quite important meaning in the ordinary notion of point of view. In that second sense, points of view could exist without any actual subject exemplifying them. Here, points of view explicitly have a strong relational and modal, especially subjunctive, character. Points of view offer possibilities of having access to the world. They offer possibilities of seeing things (hearing them, touching them, etc.), possibilities of thinking about them (considering them, imagining them, etc.), and possibilities of valuing them (assessing them, pondering them, etc.). (Campos & Gutierrez, 2015, p.3)

Next, and following this section, we will claim that the ecological background layer is absolutely relevant to create a sense of subjectivity and self.

2.4 Neuro-ecological background layer of PV I – Temporal relation of world and self

Following the two previous sections, we argue that the ecological background layer of PV is key in providing the ontological ground of subjectivity and ultimately of the self. This distinguishes our approach from both past and present philosophical and neuroscientific approaches that usually claim the phenomenal and mental features of self (and hence the mental surface layer of PV) to provide the source of subjectivity (see our conclusion where we situate our approach in a methodological-historical context).

The key feature of the ecological background layer of PV is its relational character, as it relates and connects the self to and within the world. Relation means that the PV is connected and related to something beyond itself. Put into the context of ecological psychology, that “something beyond itself” is the environment characterized by its ecological features. This includes natural, social, and cultural kinds of information along with their descriptive and normative aspects – for the sake of simplicity, we will lump them all together under the notion of ‘ecological information’ understood in a broad sense.

The ecological background layer of PV includes that ecological information which it shares with the world. The world itself contains at least two types of ecological information: that which is shared with the organism and that which is not shared with the specific organism, thus extending beyond the latter. Importantly, this entails partial rather than total overlap since the sharing between world and brain, and ultimately of world and PV, is incomplete. Consider the timescales. An organism has a limited repertoire of timescales compared to the world. For instance, we as

humans cannot directly perceive the ultraslow seismic waves preceding earthquakes, nor can we perceive the ultrasonic frequency ranges accessible to bats.

The overlap in the amount of ecological information between the world (as a whole) and the ecological background layer of PV (as part of the world) constitutes an ontological relation between the world and PV as this relation (i.e., the overlap in ecological information) defines the existence and reality of PV and consecutively the subjectivity of self. What do we mean by ecological information shared between world and PV? We refer to biophysical features and their related spatial and temporal scales which can be traced to the brain and how it stands relative to the world. For instance, the bat, based on its biophysical features, shows a PV which enables it to access ultrasonic information (Nagel, 1974). In contrast, the human PV does not allow us to access ultrasonic features since our brain does not possess the proper biophysical characteristics as it is ultimately based on different temporal and spatial scales (when compared to those of the bat).

This leads us back to timescales and their contribution to constituting the ecological background layer of PV: the broader the range of temporal and spatial scales encompassed by the ecological background layer of PV, the more extensively the self can relate to the world and its ecological information. We consequently assume that the ecological background layer of PV is nested and contained within the world and its ecological information in a scale-free way, analogous to how a smaller Russian doll is nested within a larger, self-similar version of itself. Scale-free nesting of the PV's ecological background layer within the world implies that one would expect to find overlap and nestedness in timescales between world and self: the world's much longer timescales are related to the self's shorter time scales in a self-affine and cross-scale way, as the former nests and contains the latter.

2.5 Neuro-ecological background layer of PV II – Temporal relation of world, brain and self

How about the brain and its role in constituting the ecological background layer of PV? We have seen that the self is scale-free and mediated through the brain's scale-free activity. The latter, in turn, is strongly shaped by the scale-free activity of the world, entailing the temporal nature of the world-brain relation. Putting it all together, we now postulate that the ecological background layer of PV is ontologically based on the world-brain relation through scale-free activity: the more the world and brain are temporally (and spatially) nested within each other and exhibiting larger ranges of timescales shared with the world, the greater the temporospatial range of the ecological background layer.

Ultimately, this broader, more expansive range of timescales of the self with the world (through the brain and world-brain relation) permits greater ecological extension of the self towards and within the world. In contrast, if the temporal range of the ecological background layer is limited, meaning lower degrees of temporal nestedness and smaller range of timescales shared with the world, the self becomes more restricted and isolated in its relation to the world.

Let us consider the comparison of bats and humans. Bats, as pointed out by Nagel (1974), can process ultrasonic waves which humans and their brains cannot. This means that the world's timescales align with different timescales in bats and humans, namely the ultrasonic and non-ultrasonic, respectively. The bat's point of view and its neuro-ecological background layer are consequently nested and contained within a different timescale of the world when compared to humans and their PV. The differences in the temporal extension of the bat's and human's points of view within and relative to the world leads to differences in their subjectivity which is manifest in their different "what it is like" (Nagel, 1974).

Importantly, the differences highlighted here do not concern the actual mental contents themselves but rather the predisposition for processing the range of possible contents. Our argument is that bats and humans differ in the ontological predispositions of their PV due to their different timescales. This temporal predisposition for the range of possible contents is then complemented by the actual contents, that is, those contents they are exposed to in their environment, and which fall within the range of their PV-based temporal predisposition. Together, we can see that the mental contents are doubly determined: by the predisposition for a particular range, related to the PV, and the actual contents themselves.

3 Empirical evidences: timescales and the self

In this second step, we want to highlight the relevance of timescales regarding the self and, by consequences, a notion of fundamental subjectivity. In exploring the intricate relationship between temporal dynamics and the self, a growing body of research has yielded compelling evidence regarding the role of timescales in shaping various facets of human self-awareness and cognition. As we navigate through these evidences, the goal is to discern the relevance of timescales in shaping the self and, by extension, consider the potential implications for the development of artificial intelligence seeking to emulate human cognitive processes and subjective experiences.

3.1 The scale-free self – the brain's long-range temporal correlations (LRTC) shape the self

Is the self related to the LRTC of the brain's neural activity? Recent studies have shown that the brain's scale-free activity, as measured with either Power Law Exponent (PLE) or Detrended Fluctuation Analysis (DFA), is related to mental features such as the self (Huang et al., 2016; Scalabrini et al., 2017, 2019; Wolff et al., 2019). Together, these studies show that the degree of resting state PLE directly predicts three relevant features:

- (i) the degree of self-consciousness (Huang et al., 2016; Wolff et al., 2019);
- (ii) task-related activity during self-specific stimuli (Scalabrini et al., 2019);

- (iii) the degree of temporal integration on a psychological level of self-specificity (Kolvoort et al., 2020)

Let us describe the findings in more detail. Huang et al. (2016) and Wolff et al. (2019) recorded resting state activity in fMRI and EEG of the brain, that is a task-free condition without any external demands. They calculated the degree of the brain's PLE in both fMRI and EEG. The same subjects also underwent psychological investigation of their self with the self-consciousness scale. Both studies found the same relationship of brain PLE and self-consciousness: the higher the PLE, that is, the more the slow-fast power balance is shifted towards the slow pole, the higher the degree of the subject's private self-consciousness (cf. Fig. 1).

Importantly, these findings hold only for the PLE as index of slow-fast balance but not for either the slow or fast frequencies alone. Finally, it shall be mentioned that this concerns a wide range of frequency range, from very slow (0.01 to 0.1 Hz), as covered by fMRI (Huang et al., 2016), to faster ones as measured in EEG (1-80 Hz) (Wolff et al. 2019). This means that it is the degree slow-fast integration, i.e., their degree of scale-freeness, that is related to the sense of self. The self is thus intrinsically scale-free as it connects and links different timescales short/fast and long/slow. Such cross-scale self exhibits both temporal continuity and discontinuity and nests them within each other in a scale-free way: temporal continuity,

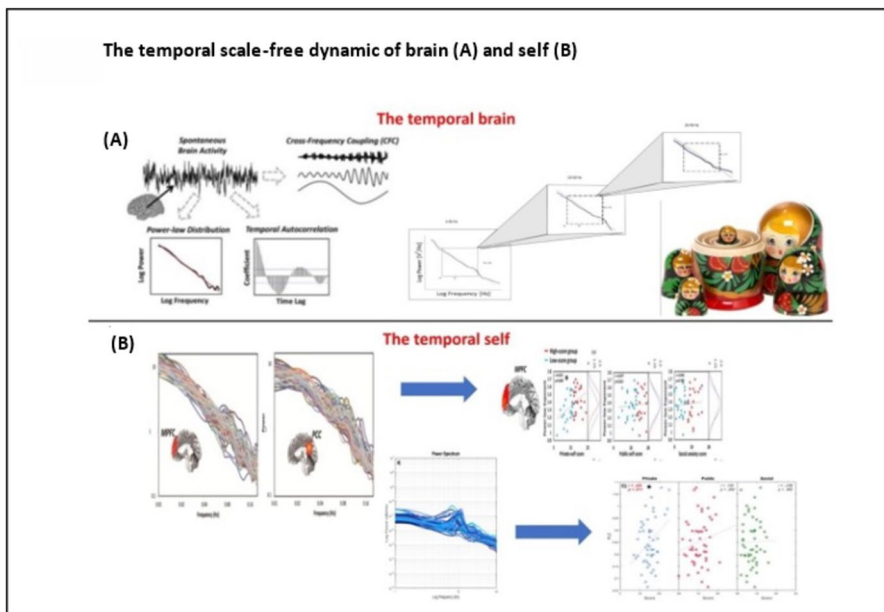


Fig. 1 **A** The temporal brain – Temporal nestedness with scale-free activity in the brain (left and upper right) just like Russian dolls with their spatial nestedness (lower part). **B** The temporal self – From the brain's scale free activity with its temporal nestedness to the self in infralow frequencies of fMRI (upper right, Huang et al., 2016) and faster frequencies of EEG (lower right, Wolff et al., 2019)

as mediated by the more powerful slower frequencies, nests and contains temporal discontinuity, as related to the less powerful faster frequencies.

In the context of AI, this implies that successful emulation of the self may require systems that can dynamically integrate information across different temporal scales, allowing for a more comprehensive and nuanced understanding of the world. As AI continues to progress, integrating these insights into the design and development of AI systems can contribute to the creation of machines that not only process information efficiently but also exhibits a more sophisticated understanding of temporal dynamics and, by consequence, self-awareness, and subjective experience. If we aim to create AI systems that can mimic the human mind, then we will need to replicate the scale-free and cross-scale nature of temporal dynamics that are linked to the self.

3.2 Temporal gradient of self – Temporal extension from faster to slower frequencies

Analogous extension can be observed in the temporal domain. Subcortical regions show a rather limited range in their power spectrum that is focused predominantly on faster frequencies (> 1 Hz) as slower frequencies (< 1 Hz) require more spatial extension (Buzsáki & Draguhn, 2004; Buzsáki, 2006). That changes on the cortical level as that is spatially more extended. Here infraslow frequencies (0.001–0.1 Hz) dominate exhibiting strong power in a scale-free way (He, 2014; He et al., 2010; Northoff & Huang, 2017). But even on the cortical level, there is a distinct hierarchy of intrinsic neural timescales as detailed above.

Such continuous temporal extension from subcortical over cortical to Default Mode Networks (DMN) regions is well in line with the findings of self. While being related to all timescales, short and long, stronger slower frequencies relative to faster ones foster the self whereas non-self tilts towards the faster frequency ranges (Wolff et al. 2019, Kolvoort et al., 2020; Huang et al., 2016; Sugimura et al., 2021). Temporal extension from subcortical faster to cortically more slow frequencies may thus well correspond to a temporal gradient of self-non-self along the temporal continuum of fast-slow frequencies.

Accordingly, rather than being determined by a single timescale (see Wolff et al. 2019 for support), the self must be conceived as multiscale as determined by the temporal dynamics across different timescales. As the temporal gradient from faster to slower timescales converges well with the gradient of spatial extension from subcortical over cortical to DMN regions, we assume self-specific processing to occur along the lines of a spatio-temporal gradient. Self-specific processing is thus determined in a continuous way, i.e., as gradient, determined by different degrees of spatiotemporal extension. Let us explicate and specify that.

The lowest level of hierarchy, self-related processing, is featured by spatially and temporally restricted processing, that is, to the own inner body focusing in integrating mainly faster frequencies (from the different inner organs). This operates on very short timescales as related to the immediate time point of the occurrence of the actual input. The next intermediate hierarchical level, self-predictive processing, is spatially and temporally more extended as it now concerns the outer body by

including slower frequencies. Self predictive processing refers to the ability to yield predictions like an empirical prior and a subsequent prediction error. Following the generative model established by Friston (cf. Friston, 2010, and its application to the self, Apps & Tsakiris, 2014; Tsakiris, 2017) one can say that the self, in part, is based on predicting what happens next. This, obviously requires a longer timescale beyond the short one of the actual stimuli or input with its particular point in time. Hence, self predictive processing extends the timescales beyond the very short timescale of the immediate actual input or stimulus.

Finally, the upper hierarchical level of *self-referential* processing is the most spatially and temporally extended as it, through the DMN and its connections to the whole brain and its strong slow frequencies, can extend to environmental and social inputs thus situating the self in a spatially and temporally most extended virtual or mental context. Here, the self can extend itself to far away time points in both past and future, called mental time travel (Schacter et al., 2012) which is a key feature of the self on the mental level (Northoff, 2018).

Together, these findings strongly support the assumption that slow dynamics and its temporal continuity in especially CMS/DMN is key in mediating self-specificity on the level of the mental self. Importantly, as it should be emphasized again, self-specificity is not related to the slow dynamics in isolation but to its relative balance with the faster dynamics, that is, the degree of nestedness of slow and fast temporal dynamics as it is measured by scale-free activity. We may consequently want to characterize the self by temporal nestedness of slow and fast dynamics which, spatially, may be related to the unimodal-transmodal gradient.

Such temporal nestedness with respect to the three-layer topography is demonstrated by Çatal et al. (2022). Using fMRI, he showed that the lowest layer of the interoceptive self shows the lowest PLE with stronger faster frequencies while the mental self layer exhibits the highest PLE with the strongest power in slow frequencies (while the proprioceptive self) took an intermediate position). This suggests that the fast-slow temporal gradient follows the topographic three-layer gradient from interoceptive over proprioceptive to mental self. Hence, the spatial-topographic gradient is met by a converging temporal-dynamic gradient among the three layers of self.

In the context of Artificial Intelligence (AI), understanding the intricate temporal dynamics and hierarchical layers of self-processing as proposed in this section could contribute to the development of more relevant and human-like AI models, particularly in terms of replicating self-awareness and subjective experience. Incorporating these insights may lead to AI systems that better align with the spatio-temporal complexities observed in the human brain.

3.3 Temporal integration – Key feature of self

Is such self-specificity of the brain's internal resting state activity also carried over to external task demands during self-specific tasks? This was studied in fMRI by Scalabrini et al., (2017, 2019): they measured both rest as well as task during the active touch towards an animate (another person) and non-animate

(mannequin hand) targets. They observed that the degree of PLE in the resting state predicted the degree to which subjects could differentiate in their task-related activity between animate and non-animate targets. Given that rest and task states occur and are measured at distinct points in time, this strongly suggests a memory effect: the temporal or dynamic memory of the resting state is carried over to the task state as otherwise the latter could not be modulated by the former. Given that such temporal memory effect in terms of rest-task modulation was related to the self-non-self differentiation, one would strongly assume it to be self-specific.

How does such self-specific temporal memory of the resting state affect the task states? This was addressed by Kolvoort et al. (2020) in an EEG study on self. They measured resting state in EEG and conducted a psychological self-task where subjects were required to associate self- and non-self-specific stimuli across different time delays (from 200 to 1400 ms). They demonstrate that the self-specific effects in terms of accuracy was preserved across all temporal delays with inter-subject variation. That, in turn, was related to the resting state PLE: the higher the resting state PLE, i.e., the stronger the slower frequencies relative to the faster ones, the stronger the self-specific effect was preserved across the different time delays on the psychological level. This suggests that temporal integration of different timescales as indexed by temporal memory may be key in mediating the co-occurrence temporal stability and flexibility of the self.

Two recent studies of ours lend further support to the key role of temporal integration (Smith et al., 2022; Wolman et al., 2023). Smith et al. (2022) presented two eight minutes narratives to subjects while undergoing EEG, one narrative was about the self while the other concerned non-self as being the same for all subjects. Being a narrative with a continuous stimulus input, the paradigm itself required the subjects to temporally integrate the different stimuli as to collate them into semantically meaningful words, sentences, paragraphs etc.

They observed that behaviorally subjects showed the greatest cursor moving distance and velocity in the self-narrative compared to the non-self narrative. This suggests a clear behavioral effect with larger spatial (and also temporal) extension of self compared to non-self. On the neuronal side, the self narrative induced significantly longer autocorrelation window (ACW), in especially the longer version (ACW-0 as distinct from ACW-50; Golesorkhi et al., 2021) compared to the non-self. That was also mirrored in larger rest-task difference for self than for non-self. These findings demonstrate that especially the longer timescales are key for temporal integration of different stimuli, i.e., inputs by the self (whereas the non-self seems to operate on shorter timescales that are more prone to temporal segregation (Wolff et al., 2022; Smith et al., 2022)).

In the context of AI, these studies imply that replicating the nuanced temporal dynamics observed in human cognition, particularly the self-specific temporal memory effects, may be critical for AI systems to develop a sense of subjectivity. The intricate interplay of temporal integration and segregation observed in the human brain, as outlined in these studies, provides a framework for understanding the complex relationship between the internal resting state, external task demands, and the subjective experience of the self. Applying these insights to AI development could

contribute to more sophisticated and nuanced AI systems that better align with the temporal characteristics of human cognition.

3.4 Balance of temporal integration and segregation – self vs non-self

Yet another study by Wolman et al. (2023) applied two versions of one and the same paradigm: they used self-face matching paradigm by presenting morphed faces in either a continuous (10–15 s) or discontinuous (single face stimuli). As expected, subjects had more difficulty in identifying faces as self or non-self in those face pictures that were morphed by around 50%. Interestingly, as based on signal detection theory, subjects showed a bias in their face matching perception, that bias tended towards either the self or non-self but was consistently present.

Next, they recorded EEG during the continuous paradigm version. They showed that the longer ACW, ACW-0, correlated with the bias (Criterion C in signal detection theory) while the shorter ACW-50 was related to the accuracy (sensitivity d' in signal detection theory). Next, going into source space of their EEG data, they show that ACW 0 in specifically cortical midline structure, the core regions of the default-mode network, but not in primary visual cortex correlate with the bias, i.e., Criterion C. Finally, employing computational modelling, they demonstrate that the regions of the cortical midline structure tend to temporally smooth and thus integrate their task-related responses across several inputs (as presented with the same timing as in the face paradigm). In contrast, regions of the visual cortex respond to each single input with a distinct task-related activity thus favoring temporal segregation over integration.

Together, these results suggest that temporal integration as mediated by longer timescales, i.e., long ACW and high PLE, is key in mediating the effects of self. While the non-self seems to be more characterized by higher degrees of temporal segregation. That may psychologically be manifest in higher degrees of temporal continuity of self and temporal discontinuity of non-self – the self is more about slower and longer processing while the non-self is more oriented towards faster and shorter processing.

To conclude this paper, we will argue that the evidence presented in this second part of the paper is enough to claim that the philosophical argument developed in the first part is sound and that Artificial Intelligence is not and cannot be conscious or develop a sense of subjectivity since its timescales are not neuroecologically aligned with the world in a relevant way.

4 Comparing AI with the human—Does AI show a basic fundamental subjectivity?

To conclude this paper, we will focus on the question of whether artificial intelligence can exhibit a basic, fundamental subjectivity. The evidence suggests that the self's complexity is intimately tied to the scale-free nature of the brain's temporal dynamics, a phenomenon involving the dynamic integration of different timescales.

While AI may surpass in processing information efficiently, their essential challenge lies in replicating the integrated temporal dynamics that contribute to human subjectivity. This poses a compelling avenue for future research in AI development to delve into the temporal gradations that underpin basic fundamental subjectivity in the human experience.

4.1 Timescales and AI

The evidence presented in the study by Wolman et al. (2023) holds deep implications for the development of Artificial Intelligence, mostly regarding the viability of AI to develop a sense of subjectivity or awareness. The findings highlight a fundamental aspect of human cognition that is missing in AI: the temporal dynamics underlying the perception of self and non-self. The observed bias in face matching perception, influenced by the duration of the anterior cingulate wave (ACW) and associated neural structures, suggests that humans engage in a distinctive temporal processing that aligns with the concept of self. This temporal integration, characterized by longer timescales, reflects a nuanced and intricate interplay of neural activity that gives rise to subjective experiences.

In contrast, the study reveals that the non-self, or external stimuli, is characterized by higher degrees of temporal segregation, favoring faster and shorter processing in the visual cortex. This dichotomy in temporal processing, where the self exhibits slower and longer processing, and the non-self favors faster and shorter processing, implies a unique temporal signature for human consciousness and subjective experience. Following this, we can claim that there are inherent limitations of AI systems, which often operate on faster processing timescales and lack the intricate temporal dynamics observed in the human brain.

These evidences extend beyond the neuroscientific realm to truly neurophilosophical considerations, where we can claim that the temporal characteristics observed in human cognition are essential for a relevant alignment with the world. The main claim advanced in the first part of the paper that AI cannot achieve consciousness or subjectivity is grounded on the premise that its temporal processing lacks the neuroecological alignment necessary for a genuine understanding of the self and the world. This has significant repercussions for the ongoing efforts to shape different AIs with human-like abilities, suggesting that the development of a true sense of subjectivity may remain elusive without a deep understanding and replication of the intricate temporal dynamics observed in the human brain.

4.2 Absence of a point of view in AI

As we saw before, the integration of information across different regions in the human brain is crucial for the formation of a unified and subjective point of view. The brain's intrinsic activity and its different timescales involving the synchronization and communication between various neural networks, contribute to the coherent and integrated nature of conscious experience. This dynamic interplay gives rise

to a subjective perspective, a point of view that shapes our awareness and perception of the world.

Following this, we can now ask the relevant question: what AI can and cannot do with its current timescales? We know that the human brain is highly adaptive to its environment, that is, the timescales of the brain are wider in range and degree and that enable the organism to match and align to the different timescales of the world. As we argued in the first part of the paper, that is what offers to humans a Point of View, a subjective experience that is unique to each individual. Of course, the different timescales are not always perfectly aligned: from time to time, we fail to align correctly with the world. When that alignment is structural, the unbalance is so strong that the normal human brain becomes abnormal in its functioning, and that can lead to different kind of psychiatric disorders, depending of which timescales is affected (Northoff, 2023).

Why is this relevant to evaluate AI? Following our neurophilosophical thesis, we can argue that AI systems are really efficient in specific tasks – such as playing Chess against the best human player in the world – exactly because they are not adaptive: because they cannot use the same internal timescales and apply it to other tasks. If we want to consider developing AI systems that can have a subjective point of view, we will need to replicate the several timescales – and the complex physiology behind them – that we know are part of what it means to be conscious (even though we do not know the entire detailed story at the moment).

An example of current AI technologies that show why our arguments are relevant to evaluate them can provide. Regarding the neural architecture, most AI models – particular deep learning systems – are based on specific networks termed “artificial neural networks (ANNs). These models consist of many layers interconnected (“artificial neurons”) with different weights that are regulate throughout the training phase of the model. These weights determine the strength of the connection which will impact in the relevance of each input provided to the model. Finally, each output is computed by the application of an “activation function” such as rectified linear unit (ReLU) or hyperbolic tangent (tanh) (Goodfellow et al., 2016). This kind of architecture lacks extrinsic connectivity, being non-dynamic in its nature and not allowing the gathering of new data modalities from the original database.

One consequence of this passive structure is the difference in cognitive flexibility between current AI models and the human brain: the first can only specialize in specific tasks, lacking the kind of cognitive flexibility that is observed in humans, allowing them to apply the same knowledge in different domains. An example of this can be given related to the real case of Google’s DeepMind AlphaGo: this AI model was able to defeat the number one human champion in Go, the famous Chinese game (Silver et al., 2016). However, the exact model cannot perform well in other games (e.g. chess) or other tasks (e.g. image recognition in medical context) since its data-base is only grounded on Go: for these reasons, a different model (i.e., AlphaZero) had to be created to beat the best human player in chess.

This highlights how the neural network architecture in current AI models is fixed after the training phase. The only method to incorporate new information is to retrain the entire model, resulting in a new fixed structure. In contrast, the human brain exhibits significant neuroplasticity, enabling it to receive and adapt to new

information from the world, adjusting its internal timescales to synchronize with the external timescales of the world, something that is clearly missing in this specific drawn from a real case of current AI model.

4.3 No signs of subjectivity – absence of perspectiveness and mineness

From what we argued, another consequence is to argue that AI lacks any kind of perspectiveness and mineness and, because of that, current AI systems cannot have any subjectivity experience. Note that this is more an indirect argument as it follows from our main claim that AI does not have a point of view. If, for instance, one were to contest that perceptiveness and mineness necessarily presuppose a point of view, our argument would no longer hold. Moreover, our argument presupposes that the point of view and subsequently perceptiveness are necessarily related to experience. If, for instance, one were to contest that perceptiveness is possible without experience by itself independent of the latter, our argument here would be undermined. Together, one may want to keep in mind these possible restrictions to the overall and general validity of our argument which, as all argument, necessarily depends on its respective presuppositions and framework. Let us now proceed with our distinctively neurophilosophical argument.

Again, our neurophilosophical thesis can also explain why there is such relationship of the point of view with perceptiveness and mineness including that the absence of the former entails the absence of the latter. Firstly, AI models passively process their inputs, lacking the ability to actively shape or align them with different contexts or circumstances. Quite to the contrary, in the human brain, timescales actively shape and select inputs from the environment that are deemed relevant for conscious processing (for a distinction between passive versus active models of the brain, see Gouveia & Northoff, 2019). One consequence of this is to claim that, in order to develop truly artificial systems that have a subjectivity, like human beings have, we will need to replicate the specific active timescales associated with dynamical neuronal activity, such as the spontaneous activity of the brain (Northoff, 2018).

Furthermore, AI lacks an important feature related to timescales: namely, it lacks an embodied nature, one that can align itself with the world. In humans, the subjective experiences are not exclusively confined to the brain but are deeply intertwined with the body and its interactions with the external world (Gallagher, 2005b; Thompson, 2015; Varela et al., 1991). AI systems lack this embodied dimension, since they do not possess a physical form or active sensory interactions that shape subjective experience. The absence of a bodily foundation in AI contributes to its inability to exhibit “mineness”: as we argued before, the brain’s capacity to integrate information across various timescales is crucial for the continuity and discontinuity of subjective experiences. However, AI, which often operates on predefined temporal patterns and lacks the dynamic temporal integration observed in the brain, struggles to replicate the nuanced temporal aspects of subjectivity (cf. Fig. 2a and b).

Finally, while AI is capable of executing particular tasks and processing information, falls short of replicating these intricate neural processes, leading to its inability to exhibit a “perspectiveness” or subjective point of view since those are emergent

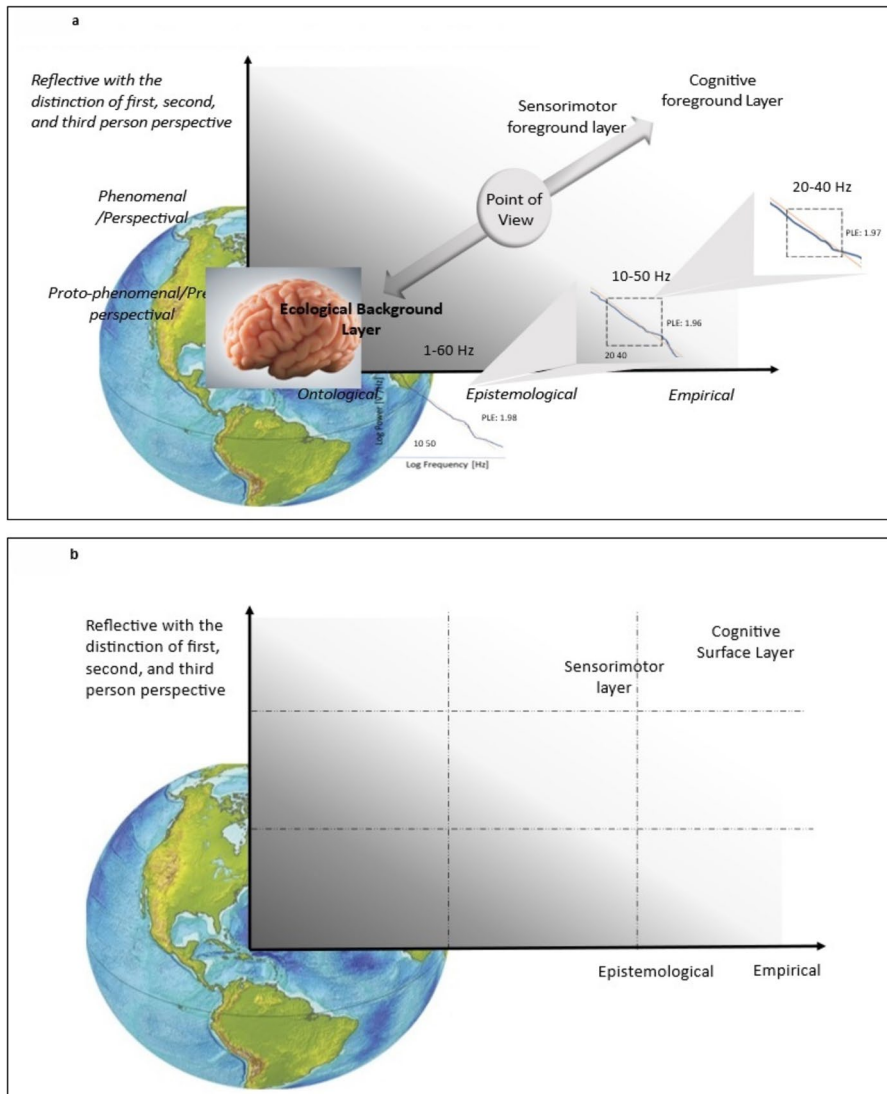


Fig. 2 **a** Point of View with scale-free nestedness – world-brain relation and its different layers. **b** Absence of the Point of View and scale-free nestedness in current Artificial Intelligence – no world-AI relation and no neuroecological background layer

properties of the brain's complex neural dynamics, involving integration, embodiment, and temporal processing. The quest for AI to mimic these aspects involves addressing the current limitations in matching the complexity of the human brain's intrinsic temporal processes. Only then, if those timescales are integrated into the internal nature of an AI system, will we be able to contemplate whether those models can indeed develop a form of subjectivity.

5 Conclusion

Does AI exhibit subjectivity? Based on human subjectivity and its timescale-based world-brain mediated point of view with perspectiveness and minesness, we deny that. Why? Because AI does not show the variety of timescales nor their flexibility and adaptive nature as humans, due to their brain, show. This concerns specifically the neuroecological layer of the point of view as this, as in its name, creates our virtual “location” in the world from which we and within we experience ourselves as part of that very same world as whole. Given the absence of such timescales in AI, we infer that it does not possess the neuroecological layer of a point of view and consecutively no point of view at all. This entails the absence of perspectiveness and minesness as signs of a basic fundamental subjectivity. In short, based on humans and neurophilosophical strategy, we deny that AI exhibits a basic fundamental subjectivity.

We shall note that the ability of exerting motion and responses (Yoshida et al., 2013) does not yet imply experience of those motion and responses. For that, as we argue, one would need a spatiotemporal anchoring within the world through a point of view. That would provide the basis of possible experience and that predisposition of possible experience remains absent in current AI. Note that we do not necessarily exclude that future AI will exhibit these capabilities, but we only argue that currently, no AI is capable of developing the necessary timescales to have a point of view.

Finally, there are deeper philosophical connotations to our argument. The point of view is constituted by and through the relationship of the spatiotemporal coordinates of the world itself and the ones of the respective organisms, e.g., humans in our case. More specifically, this relationship can be described ontologically by spatiotemporal nestedness (Northoff, 2018). The smaller ranges of the human spatiotemporal coordinates are integrated and embedded, i.e., nested, within the much larger one of the world. This implies that human time, as constituted by the brain and its inner time (Northoff & Zilio, 2022a, b), is integrated within the world’s time. Given that this integration constitutes a point of view as the very basis of our being and its experience, one can speak of “being in time” and “being in the world” as for instance the philosophers Martin Heidegger and Hubert Dreyfus do. Especially the latter argued that AI lacks the “being in time” and “being in the world”.

In a sense, we support and develop that earlier argument and further support it on empirical and neurophilosophical grounds. It is such spatiotemporal nestedness within the wider spatiotemporal coordinates of the world as whole that is missing in current AI. Despite all its abilities across sensory, motor, cognitive, social and emotion domains, current AI remains unable to experience itself including these various functions as part of the wider world as it lacks the “being in time” and “being in the world” as based on the neuroecological layer of the point of view (Northoff & Smith, 2023; Northoff, 2018, 2024).

Acknowledgements GN is grateful for financial support from Physicians Incorporated Services (PSI) and Canada Institute of Health Research (CIHR) in Canada, to the Canada-UK Artificial Intelligence (AI) Initiative “The self as agent-environment nexus: crossing disciplinary boundaries to help human selves and anticipate artificial selves” (ES/T01279X/1), and for funding provided by UMRf, uOBMRI, CIHR,

and fNSERC. SG is funded by CEEC Individual Project by FCT 2022.02527.CEECIND, at the Mind, Language and Action Group, Institute of Philosophy, University of Porto (Faculdade de Letras, Via Panorâmica s/n, P-4150-564 Porto, Portugal).

Author's contribution GN first author, SG second author.

Funding Open access funding provided by FCT/IFCCN (b-on).

Availability of data and materials Not applicable.

Declarations

Ethical approval Not applicable.

Informed consent Not applicable.

Statement regarding research involving human participants and/or animals Not applicable.

Competing interests Not applicable.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Apps, M., & Tsakiris, M. (2014). The free-energy self: A predictive coding account of self-recognition. *Neuroscience & Biobehavioral Reviews*, 41, 85–97. <https://doi.org/10.1016/j.neubiorev.2013.01.029>
- Buszaki, G. (2006). *Rhythms of the brain*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195301069.001.0001>
- Buzsáki, G., & Draguhn, A. (2004). Neuronal oscillations in cortical networks. *Science*, 304(5679), 1926–1929. <https://doi.org/10.1126/science.1099745>
- Campos, M. V., & Gutierrez, A. M. (Eds.). (2015). *Temporal points of view*. Springer Cham. <https://doi.org/10.1007/978-3-319-19815-6>
- Çatal, Y., Gomez-Pilar, J., & Northoff, G. (2022). Intrinsic dynamics and topography of sensory input systems. *Cerebral Cortex*, 32(20), 4592–4604. <https://doi.org/10.1093/cercor/bhab504>
- Friston, K. (2010). The free energy principle: A unified brain theory? *Nature Review Neuroscience*, 11(2), 127–138.
- Gallagher, S. (2005a). *How the body shapes the mind*. Oxford University Press.
- Gallagher, S., & Daly, A. (2018). Dynamical relations in the self-pattern. *Frontiers in Psychology*, 9, 664. <https://doi.org/10.3389/fpsyg.2018.00664>
- Gallagher, S., & Zahavi, D. (2020). *The Phenomenological Mind* (3rd ed.). Routledge.
- Gallagher, S. (2005b) Dynamic models of body schematic processes. In Preester & Knockeart (Eds.), *Body image and body schema: Interdisciplinary perspectives on the body* (pp. 233–250). John Benjamins. <https://doi.org/10.1075/aicr.62.15gal>
- Goldman, A. I. (1970). *Theory of human action*. Princeton University Press.

- Golesorkhi, M., Gomez-Pilar, J., Zilio, F., et al. (2021). The brain and its time: Intrinsic neural time-scales are key for input processing. *Communications Biology*, 4, 970. <https://doi.org/10.1038/s42003-021-02483-6>
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.
- Gouveia, S. (2022). *Philosophy and neuroscience: A methodological analysis*. Palgrave MacMillian.
- Gouveia, S., & Northoff, G. (2019). A neurophilosophical approach to perception. In M. Curado & S. Gouveia (Eds.), *Perception, cognition, and aesthetics*. Routledge.
- He, B. J. (2014). Scale-free brain activity: Past, present, and future. *Trends in Cognitive Science*, 18(9), 480–487. <https://doi.org/10.1016/j.tics.2014.04.003>
- He, B. J., Zempel, J. M., Snyder, A. Z., & Raichle, M. E. (2010). The temporal structures and functional significance of scale-free brain activity. *Neuron*, 66(3), 353–369. <https://doi.org/10.1016/j.neuron.2010.04.020>
- Hildt, E. (2019). Artificial intelligence: Does consciousness matter? *Frontiers in Psychology*, 10, 1535. <https://doi.org/10.3389/fpsyg.2019.01535>
- Huang, Z., Obara, N., Davis IV, H. H., Pokorny, J., & Northoff, G. (2016). The temporal structure of resting-state brain activity in the medial prefrontal cortex predicts self-consciousness. *Neuropsychologia*, 82, 161–170. <https://doi.org/10.1016/j.neuropsychologia.2016.01.025>
- Klar, P. (2021). What is neurophilosophy: Do we need a non-reductive form? *Synthese*, 199, 2701–2725. <https://doi.org/10.1007/s11229-020-02907-6>
- Klar, P., Catal, Y., Langner, R., et al. (2023). Scale-free dynamics in the core-periphery topography and task alignment decline from conscious to unconscious states. *Communications Biology*, 6, 499. <https://doi.org/10.1038/s42003-023-04879-y>
- Kolvoort, I. R., Wainio-Theberge, S., Wolff, A., & Northoff, G. (2020). Temporal integration as ‘common currency’ of brain and self-scale-free activity in resting-state EEG correlates with temporal delay effects on self-relatedness. *Human Brain Mapping*, 41(15), 4355–4374. <https://doi.org/10.1002/hbm.25129>
- Mindt, G., & Montemayor, C. (2020). A Roadmap for artificial general intelligence: intelligence, knowledge, and consciousness. *Mind and Matter*, 18(1), 9–37.
- Nagel, T. (1974). What is it like to be a bat? *Philosophical Review*, 83, 435.
- Namikawa, J., Nishimoto, R., & Tani, J. (2011). A Neurodynamic account of spontaneous behaviour. *PLOS Computational Biology*, 7(10), e1002221. <https://doi.org/10.1371/journal.pcbi.1002221>
- Ng, Y.-K. (2023). Could artificial intelligence have consciousness? Some perspectives from neurology and parapsychology. *AI and Society*, 38(1), 425–436.
- Nida-Rümelin, M. (2014). Basic intentionality, primitive awareness and awareness of oneself. In Reboul, A. (Eds.), *Mind, values, and metaphysics*. Springer, Cham. https://doi.org/10.1007/978-3-319-05146-8_18
- Northoff, G. (2004). *Philosophy of the brain*. John Benjamins Publishing.
- Northoff, G. (2014a). *Minding the brain: An introduction of philosophy and neuroscience*. Palgrave Macmillan.
- Northoff, G. (2014b). *Unlocking the brain II*. Oxford University Press.
- Northoff, G. (2016). Is the self a higher-order or fundamental function of the brain? The “basis model of self-specificity” and its encoding by the brain’s spontaneous activity. *Cognitive Neuroscience*, 7(1–4), 203–222. <https://doi.org/10.1080/17588928.2015.1111868>
- Northoff, G. (2018). *The spontaneous brain: From the mind–body to the world–brain problem*. MIT Press.
- Northoff, G. (2022). Non-reductive neurophilosophy – What is it and how it can contribute to philosophy. *Journal of NeuroPhilosophy*, 1(1), 17–30.
- Northoff, G. (2023). *Neurowaves: Brain*. McGill 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 & Biobehavioral Reviews*, 80, 630–645. <https://doi.org/10.1016/j.neubiorev.2017.07.013>
- Northoff, G., & Smith, D. (2023). The subjectivity of self and its ontology: From the world–brain relation to the point of view in the world. *Theory & Psychology*, 33(4), 485–514. <https://doi.org/10.1177/09593543221080120>
- Northoff, G., & Zilio, F. (2022a). From shorter to longer timescales: Converging Integrated Information Theory (IIT) with the TemporoSpatial Theory of Consciousness (TTC). *Entropy*, 24, 270. <https://doi.org/10.3390/e24020270>

- Northoff, G., & Zilio, F. (2022b). Temporo-spatial Theory of Consciousness (TTC) – Bridging the gap of neuronal activity and phenomenal states. *Behavioural Brain Research*, 424, 113788. <https://doi.org/10.1016/j.bbr.2022.113788>
- Northoff, G., Fraser, M., Griffiths, J., Pinotsis, D., Panangaden, P., Moran, R., & Friston, K. (2022). Augmenting human selves through artificial agents – Lessons from the brain. *Frontiers in Computational Neuroscience*, 16, 892354. <https://doi.org/10.3389/fncom.2022.892354>
- Northoff, G. (2024). Is Heidegger's fundamental ontology really fundamental? From Anthropological to Ecological fundamental ontology. *The Humanist*.
- Rakover, S. S. (2023). AI and consciousness. *AI & Society*. <https://doi.org/10.1007/s00146-023-01663-8>
- Sartre, J.-P. (1956). *Being and nothingness*. Philosophical Library.
- Scalabrini, A., Huang, Z., Mucci, C., Perrucci, M. G., Ferretti, A., Fossati, A., Romani, G. L., Northoff, G., & Ebisch, S. J. H. (2017). How spontaneous brain activity and narcissistic features shape social interaction. *Scientific Reports*, 7, 9986. <https://doi.org/10.1038/s41598-017-10389-9>
- Scalabrini, A., Ebisch, S. J. H., Huang, Z., Di Plinio, S., Perrucci, M. G., Romani, G. L., Mucci, C., & Northoff, G. (2019). Spontaneous brain activity predicts task-evoked activity during animate versus inanimate touch. *Cerebral Cortex*, 29(11), 4628–4645. <https://doi.org/10.1093/cercor/bhy340>
- Schacter, D., Addis, D., Hassabis, D., Martin, V., Spreng, R., & Szpunar, K. (2012). The future of memory: Remembering, imagining, and the brain. *Neuron*, 76(4), 677–694. <https://doi.org/10.1016/j.neuron.2012.11.001>
- Schneider, S. (2019). *Artificial you*. Princeton: Princeton.
- Silver, D., Huang, A., Maddison, C., et al. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature*, 529, 484–489.
- Smith, D., Wolff, A., Wolman, A., Ignaszewski, J., & Northoff, G. (2022). Temporal continuity of self: Long autocorrelation windows mediate self-specificity. *NeuroImage*, 257, 119305. <https://doi.org/10.1016/j.neuroimage.2022.119305>
- Sugimura, K., Iwasa, Y., Kobayashi, R., et al. (2021). Association between long-range temporal correlations in intrinsic EEG activity and subjective sense of identity. *Scientific Reports*, 11, 422. <https://doi.org/10.1038/s41598-020-79444-2>
- Thompson, E. (2015). *Waking, dreaming, being: Self and consciousness in neuroscience*. Columbia University Press.
- Tsakiris, M. (2017). The multisensory basis of the self: From body to identity to others. *Quarterly Journal of Experimental Psychology*, 70(4), 597–609. <https://doi.org/10.1080/17470218.2016.1181768>
- Varela, F., Thompson, E., & Rosch, E. (1991). *The Embodied mind: Cognitive science and human experience*. MIT Press.
- Wolff, A., Di Giovanni, D. A., Gomez-Pilar, J., Nakao, T., Huang, Z., Longtin, A., & Northoff, G. (2019). The temporal signature of self: Temporal measures of resting-state EEG predict self-consciousness. *Human Brain Mapping*, 40(3), 789–803. <https://doi.org/10.1002/hbm.24412>
- Wolff, A., Berberian, N., Golesorkhi, M., Gomez-Pilar, J., Zilio, F., & Northoff, G. (2022). Intrinsic neural timescales: Temporal integration and segregation. *Trends in Cognitive Science*, 26(2), 159–173. <https://doi.org/10.1016/j.tics.2021.11.007>
- Wolman, A., Çatal, Y., Wolff, A., Wainio-Theberge, S., Scalabrini, A., El Ahmadi, A., & Northoff, G. (2023). Intrinsic neural timescales mediate the cognitive bias of self – temporal integration as key mechanism. *NeuroImage*, 268, 119896. <https://doi.org/10.1016/j.neuroimage.2023.119896>
- Yoshida, T., Masumori, A., Ikegami, T. (2013). From text to motion: Grounding GPT-4 in a humanoid robot 'Alter3'. arXiv:2312.06571v1.
- Zahavi, D. (2005). *Subjectivity and selfhood: Investigating the first-person perspective*. MIT Press.
- Zahavi, D. (2019). Consciousness and (minimal) selfhood: Getting clearer on for-me-ness and mineness. In U. Kriegel (Ed.), *The Oxford handbook of the philosophy of consciousness*. Oxford: Oxford University Press.
- Zilio, F. (2020). *Consciousness and world A neurophilosophical and neuroethical account*. Edizioni ETS.
- Zilio, F., Gomez-Pilar, J., Cao, S., Zhang, J., Zang, D., Qi, Z., Tan, J., Hiromi, T., Wu, X., Fogel, S., Huang, Z., Hohmann, M. R., Fomina, T., Synofzik, M., Grosse-Wentrup, M., Owen, A. M., & Northoff, G. (2021). Are intrinsic neural timescales related to sensory processing? Evidence from abnormal behavioral states. *NeuroImage*, 226, 117579. <https://doi.org/10.1016/j.neuroimage.2020.117579>
- Zilio, F., Gomez-Pilar, J., Chaudhary, U., Fogel, Fomina, T., Synofzik, M., Schöls, L., Cao, S., Zhang, J., Huang, Z., Birbaumer, N., & Northoff, G. (2023). Altered brain dynamics index levels of

arousal in complete locked-in syndrome. *Communications Biology*, 6, 757. <https://doi.org/10.1038/s42003-023-05109-1>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.