Phenomenology and the Cognitive Sciences
https://doi.org/10.1007/s11097-020-09667-1

Brain imaging technologies as source for Extrospection: self-formation through critical self-identification

Ciano Aydin[1] · Bas de Boer[1]

Published online: 26 March 2020
© The Author(s) 2020

Abstract

Brain imaging technologies are increasingly used to find networks and brain regions that are specific to the functional realization of particular aspects of the self. In this paper, we aim to show how neuroscientific research and techniques could be used in the context of self-formation without treating them as representations of an inner realm. To do so, we show first how a Cartesian framework underlies the interpretation and usage of brain imaging technologies as functional evidence. To illustrate how material-technological inventions and developments can have a significant and lasting impact on views of the self, we show how this framework was influenced by another technology: the camera obscura. Subsequently, we show that brain imaging technologies challenge the idea that privileged access to the self can be obtained merely through introspection, indicating a strong discontinuity between the Cartesian and the current neuroscientific framework. Building on these insights, we reframe the self in terms of self-formation. This view neither regards the brain as an independent realizer of aspects of the self, nor assumes that self-knowledge can be obtained through introspection. From this perspective, self-formation is realized through *critical self-identification*: instead of offering representational knowledge of an 'inner self,' the potential use of brain imaging technologies within this framework lies in their capacity to offer what we call 'extrospective knowledge' that pragmatically can contribute to self-formation. Brain imaging technologies contribute to this process because they foreground our neurophysiology, which helps to critically integrate biological aspects into self-formation.

Keywords Critical self-formation · Brain imaging technologies · Extrospection · Technological mediation

✉ Ciano Aydin
c.aydin@utwente.nl

Bas de Boer
s.o.m.deboer@utwente.nl

[1] Department of Philosophy, University of Twente, Drienerlolaan 5, 7500AE Enschede, The Netherlands

1 Introduction

Brain imaging technologies are increasingly used to find neuronal networks and brain regions that are specific to the functional realization of particular aspects of the self (e.g., Christoff et al. 2011; Legrand & Ruby 2009). These studies inform debates in philosophy of mind, as well as philosophical, psychological and legal conceptions of phenomena like freedom and responsibility (e.g., Klemm 2010; Vincent 2013). On the one hand, this employment of brain imaging technologies raises important new philosophical questions, offering new perspectives on age-old philosophical debates. On the other hand, these outcomes and their philosophical interpretations seem to rely strongly on the idea that the brain acts as an autonomous agent with regard to realizing (particular aspects of) the self, thereby presupposing a divide between the internal realization of the self and the external world.

Inspired by technological mediation theory (e.g., Ihde 1990; Verbeek 2005), we will illustrate that this idea can be traced back to a Cartesian dualism that was influenced—mediated—by another technology: the camera obscura. Although this suggests a continuity between the camera obscura and brain imaging technologies, we will also point out a strong and far-reaching discontinuity: in contrast to the Cartesian view, the usage of brain imaging technologies in this regard presupposes that aspects of the self and their functioning can be externally accessed by (neuro-)scientific means (Aydin 2018a).

Our main aim in this paper is to show how neuroscientific research and techniques could be used in the context of self-formation without being treated as representations of an inner realm. To do so, we propose, by appropriating a stronger *pragmatist account* of technological mediation theory (Aydin 2018b), a view of the self that neither regards the brain as the locus of the self, nor understands brain imaging technologies as offering representations or photographs of the brain. Instead, we reframe the self in terms of self-formation. From this perspective the self continuously forms itself in the world by pursuing certain goals and ideals and developing matching habits. The potentially interesting use of brain imaging technologies within this framework lies in their capacity to offer what we call *extrospective knowledge*. In doing so, they foreground our (neuro-)biological functioning: this is a type of technological mediation that opens up the space to ask to what extent our neurophysiological interactions are on par with or pose limits and constraints to our personal or social goals and ambitions.

This view shows similarities to recent enactivist accounts of self, mind and brain. Characteristic of these approaches is that they do not conceive of the self as a localizable, unified entity that brings about experiences such as self-awareness, self-recognition, or sense of agency, which can be reduced to particular patterns of brain activity. Rather, the subject's experiences in the world are understood in relation to its environment as constituting a pattern correlating with what is called a self (e.g., Gallagher 2013; Vogeley & Gallagher 2011). The proposed view has a different focus than enactivist perspectives but also has the potential to contribute to them: it takes as a starting point environments becoming technological environments and focuses on how brain imaging technologies, increasingly used in those technological environments, mediate and allow for new ways of self-formation.

2 How brain imaging technologies could compete with introspective knowledge

In recent decades, there have been several attempts in (cognitive) neuroscience to relate several aspects that we associate with self or selfhood to the brain. For example, our capacity for self-recognition has been related to a network involving frontal, parietal and occipital areas (Devue & Brédart 2011), and our sense of agency to prefrontal and parietal areas (Haggard 2017).[1] However, as several reviews have already clearly pointed out, there is evidence neither that a specific area of the brain nor that a specific network in it can be said to be self-specific (e.g., Gillihan & Farah 2005; Legrand & Ruby 2009; Northoff et al. 2006).

Even though no demarcated self-specific area in the brain seems detectable, current brain research continues to seek for networks in the brain through which self-related actions and experiences are realized (e.g., Christoff et al. 2011, Davey et al. 2016). Various studies have indicated that neuroscientific research should overcome the idea that the self is to be found somewhere in the brain, rather than in the world (Gallagher & Daly 2018; Vogeley & Gallagher 2011). Nevertheless, the brain continues to be seen as constitutive of aspects of a 'real self' or as the 'control room' that instigates cognitive processes, moves the body and manipulates the external world (Vidal & Ortega 2017).

We propose that this conception of the brain as being the control room in which the self is constituted is due to a stubborn Cartesian legacy in current neuroscience that can be, at least partly, traced back to a view that Descartes drew from the camera obscura. Lee W. Bailey and, more recently, a number of technological mediation theorists have argued that the *camera obscura* is one of the root metaphors that helped generate the image of the mind as an inner realm, ontologically separated from the external world, an image that is characteristic of our modern suppositions of subjectivity and autonomy. The *camera obscura*, they argue, strongly affected the Cartesian view because its central perspective makes the idea of an aloof spectator position, situated outside a geometrical *res extensa*, a plausible starting point for reasoning (Bailey 1989; Ihde 2007; Kockelkoren 2003). Leaving open the question how decisive the camera obscura was for Descartes, this argument aims to illustrate how, besides ideas, *things,* and particularly technologies, impact fundamental philosophical notions.

In the sixteenth century the camera obscura was widely used for perspective drawing. Descartes was very familiar with this piece of technology. In his *Optics*, he describes it and uses it as a metaphor to explain how the world enters our mind through our senses.[2] This indicates that the *camera obscura* inspired Descartes' view that in methodical self-contemplation the self discovers that the mind is completely detached

[1] Giving a comprehensive review of the many neuroscientific studies aiming to tie aspects often associated with the self to the brain is beyond the scope of this paper, also because several accessible summaries of the development of neuroscientific research on the self and its philosophical implications are available (e.g., Gallagher 2013, Vogeley and Gallagher 2011).

[2] Descartes writes: "The objects we look at do imprint very perfect images on the back of our eyes. Some people have very ingeniously explained this already, by comparison with the images that appear in a chamber, when having it completely closed except for a single hole, and having put in front of this hole a glass in the form of a lens, we stretch behind, at a specific distance, a white cloth on which the light that comes from the objects outside forms these images. For they say that this chamber represents the eye; this hole, the pupil; this lens, the crystalline humour, or rather, all those parts of the eye which cause some refraction; and this cloth, the interior membrane, which is composed of the extremities of the optic nerve" (Descartes 1965, p. 91).

from the outside world and that our access to the contents of our thoughts (introspection) radically differs from how we encounter the external world of material things: we can have an authoritative (others cannot challenge our belief of being in a particular mental state), privileged (we always know the contents of our own minds better than we know the contents of the minds of others) and immediate (knowledge of our mental contents is non-inferential and non-evidence based) access to the contents of our thoughts, which we lack towards the external world of material things (cf. Macdonald 2007).

This capacity makes the fundamental distinction that Descartes draws between our thinking (*res cogitans*) and the material world (*res extensa*) *appear* reasonable: since introspective beliefs about my mind are epistemologically radically different from my beliefs about the outside world, my mind must be—ontologically speaking—a completely different substance than the outside world, including my body. Therefore, the sense of being an autonomous agent is derived from the *cogito*, which is ontologically separated from the external world but in some way capable of exerting influence on it.

The imperative prejudice that Descartes did not put in brackets in his methodical doubt experiment is that it is altogether *possible* to fundamentally detach oneself from the world and evaluate different ideas on the basis of clarity and distinctness. Although this assumption has been extensively challenged throughout the history of philosophy, especially from phenomenological perspectives (intentionality, being-in-the-world, embodied embedded extended, enactive cognition, etc.), not much attention has been given to the material/technological conditions that made this view appear cogent. The *camera obscura* is, thus, not an innocent and neutral metaphor that Descartes used to explain his epistemology but mediated the idea of the mind as an internal representation of a world that can be observed (and acted on) by a homunculus situated somewhere outside the world.

Moreover, the Cartesian framework, influenced by the *camera obscura*, does not mark a historical event that is left behind, but one that has been transported and rehabilitated by other technologies, such as contemporary brain imaging technologies. In the context of neuroscience, *functional Magnetic Resonance Imaging* (fMRI), *Electroencephalography* (EEG), and *Magnetoencephalography* (MEG) are the dominant mediating technologies on the basis of which the realization of (aspects of) the self and its psychology are investigated. Neuroscientists often treat imaging studies as offering depictions of the brain in an isolated realm, such that it can be approached as an independent inner agent (Vidal & Ortega 2017). This way a categorical distinction between 'inside' and 'outside' is assumed and preserved. The brain is attributed a privileged status because it is an interface presumably detached from the outside world, yet able to process information and to determine, as a kind of *causa sui*, our decisions and actions.

There is a continuity between the camera obscura and brain imaging technologies insofar as both mediate a particular view of self, but there is also an important and relevant difference between them: in Descartes' epistemology, the *cogito* is a 'black box' that is only accessible through subjective introspection, whereas brain imaging technologies promise objective knowledge of how (aspects of) the self are realized. fMRI is treated as having—given the status attributed to it in neuroscientific research on the self—the potential to eventually display the interior of the 'dark room' and to

disclose not only *what* really determines the realization of our self and our behavior, but also *how* it is able to do it.

The credibility of approaching the self as something that can be objectified in a scientific setting points to an important *discontinuity* with the Cartesian framework of the mind mediated by the camera obscura. In the Cartesian framework, the self is located 'within', and privileged: authoritative knowledge of it can only be obtained through introspection. By definition then, attaining reliable knowledge of the self cannot be outsourced to processes in the external world. This picture changes drastically when the human brain is thought to be constitutive of the human self. The self that allegedly is located 'within' is no longer a private realm.

Visualizing (aspects of) the self on brain scans supposes the possibility of making inferences about the subject's beliefs. In fact, introspective knowledge and knowledge that is based on brain scans may contradict each other, and studies suggest that there is no clear factor to determine which of the two knowledge claims is valid (e.g., Christoff et al. 2011; Legrand & Ruby 2009). For example, the truth-value of an introspective claim such as "I believe to have executed action X", might be effectively challenged by inferences made on the basis of brain imaging technologies (Farrer & Frith 2002). Another example that illustrates how imaging technologies can challenge subjective beliefs can be found in neuroscientific research on the phenomenon of cognitive dissonance, a psychological phenomenon that displays that a subject can believe that her behavior is consistent with the values she holds while it is actually not the case (e.g., Festinger 1957). Neuroscientific studies may contradict such beliefs by showing that observed patterns of brain activity reveal that the consistency of one's beliefs is *in fact* an instance of cognitive dissonance (e.g., Kitayama et al. 2013; de Vries et al. 2015). Accordingly, the belief "I had good reasons to do X" might be challenged through neuroscientific research.

We remain agnostic about whether such challenges on the basis of recent neuroimaging studies are valid, but still want to point out that nowadays brain imaging technologies *can* be used to influence beliefs 'from the outside'. Because of this, the view of the self as mediated by brain imaging technologies is not one in which self-knowledge can exclusively be obtained 'from within.' If brain scans can indeed be used as a means to settle debates about (in)voluntary action and the consistency of one's beliefs, the self and its beliefs become potentially accessible for external informed observers. As a consequence, brain imaging technologies offer the prospect of allowing scientists to make predictions about beliefs of people, which makes it possible that they are treated as epistemologically competitive to introspection; it must be clear that within this current framework 'introspective self-reports' and knowledge derived 'from without' could also complement one another, although we will propose going beyond this inside-outside distinction.

This discontinuity instigated by brain imaging technologies affects the idea that self (as inner) and world (as outer) are fundamentally detached from one another. While brain imaging technologies provide an image of the brain as detached from the world—which is continuous with a Cartesian framework, they undermine the Cartesian idea that introspection is a source of privileged, authoritative knowledge of the self. Through brain imaging technologies, knowledge of the self can potentially be obtained *in* the external world (*res extensa*), thereby challenging the assumption that there is an ontological divide between brain and world, and, hence, mind and world.

3 The mediated self as self-formation

Our alternative approach to the self as something that is continuously forming itself in and through the world breaks with the idea that brain imaging technologies can be understood as representations of an inner self. From our perspective, the 'inner self' or 'real self' is not to be found 'in the head' (see also Clark and Chalmers 1998). In fact, our approach aims to surpass the discussion of whether the self should be understood as residing inside or outside the human body by making obsolete the question where to locate the self. Rather, the self is understood as an ongoing formative process that takes place in a world mediated by institutions, scientific developments, technologies, and the biological functioning of the human body.

If the self can acquire knowledge of itself through brain imaging technologies, this knowledge could help the self to relate to its own development. It could inform the self about its inclinations, habits and conduct, and raise the question of whether its present conduct coincides with its preferred conduct. This information would allow the self, in other words, to evaluate its conduct and form itself in a preferred direction.

Our view builds on older approaches of the self as a self-referential being: Sartre, for example, illustrates with his many examples that the self never completely coincides with itself; my experiences do not only take place 'in me' but they are also 'for me' (Sartre 1943; Zahavi 2015). From the proposed point of view, the formation of the self is conceived more specifically as a process of developing habits that allow orienting towards long-term goals or ideals, which are set against the opportunities and constraints encountered during the process of formation. When understanding the self in this way, not only is the question concerning self-localization surpassed (and thereby the distinction between 'in me' and 'for me'), but the interest also shifts to how the self forms itself in relation with the world in which it is unfolding.

Peirce offers a general characterization of the conditions for self-formation (see Aydin 2009). First of all, according to Peirce, self-formation requires self-control. It is only possible to give ourselves a certain form if we have some control over our interactions. Self-control presupposes in its turn self-criticism. Self-criticism means in this context that an actor consciously reviews each of her (important) actions and compares them with certain longer-term goals and ideals that she wants to achieve (Peirce EP II, p. 377). The decision to commit to an ideal and to try to realize it will eventually provoke the agent to adjust her (natural or acquired) tendencies or to form new habits (Peirce CP 8.320).

Peirce proposes a kind of reciprocal reinforcement between ideals and habit formation: the more we devote ourselves to a certain ideal, the more we will be able to form and cultivate matching habits; and the more we form and cultivate certain habits, the more we will be sensitive to the attractive power of the ideal that we wish to incorporate. However, certain habits could also be the result of inclinations and contingent interactions. Ideals (or higher ideals) enable us to assess and adjust those (unquestioned) habits (Peirce CP 5.421).

There are some similarities (but also differences) between Peirce's view of the self and Frankfurt's concept of a person (Frankfurt 1971). Frankfurt elaborates on how people can have different first order (wanting something) and second order desires (desires *about* desires to do this or that). If someone not only wants to have a certain desire but also wants that desire to be effective (i.e., to be the product of her will), she

has, according to Frankfurt, a 'second order volition'. A person is somebody who deliberately and decisively identifies herself with a particular desire, who, in Peirce's (and our) terminology, commits herself to an ideal. Along similar lines, we believe that a person is able to discover in herself a discord between instant inclinations and higher-order volitions that are guided by longer-term wishes and goals. However, we depart, as will become clear later on, in our views regarding the 'source' for pursuing those longer-term wishes and goal: we will focus on the knowledge and self-criticism acquired through brain imaging technologies.

Charles Taylor's idea that being a self is an achievement rather than a given also resembles our Peircian proposal. Taylor holds that to be a self is to stand in a particular interpretative and evaluative relation to oneself; the self is constituted by organizing and unifying our experiences and actions according to narrative structures (Taylor 1989). However, the self can never escape the fact that all its goals, including its most personal preferences, are greatly determined by society, history and tradition. Taylor offers an 'instrument' to evaluate whether my actions are on a par with my 'real' self, instead of being externally determined (1992). If there is no tension between my actions and the long-term goals and ideals that I pursue, which form to a great extent (depending on the degree of critical identification) my 'real' self, I then can also often manage to fit my actions into a comprehensible story that is 'my own'. I can still persuade others and myself that such actions can be given a meaningful place in a coherent narrative about myself.

While Taylor emphasizes the importance of narratives in maintaining a relative degree of independence in self-formation, our account centralizes the ability of the self to test its actions pragmatically in relation to its long-term goals and ideals. First, we take it as phenomenological evidence based on everyday life experience that we can set distal goals and ideals for ourselves. It is made possible by our ability to reflect on ourselves, which is often instigated by a certain discord between how we act and how we would like to act. However, what is less evident and of vital importance for self-formation, is whether conscious long-term intentions can really influence our behaviour. Some opponents of free will argue that proximal conscious intentions and decisions have no causal efficacy and power because they are produced by preceding unconscious processes in the brain (Wegner & Wheatley 1999; Soon et al. 2008; Fried et al., 2011). Other opponents of free will claim that they have no causal efficacy and power because they are necessitated by situational influences (Williams & Bargh 2008; Doris, 2002). Although these views are controversial, we can leave here in the middle whether and in what sense proximal intentions and decisions necessitate behaviour since they do not affect the alternative perspective that we are attempting to develop. What is a relevant question for our thesis, is whether the arguments used by opponents of free will also apply to distal or longer-term intentions.

Slors, for one, attributes causal efficacy and control to distal conscious intentions (2015). On the one hand, he points at the wealth of empirical evidence, summarized by Baumeister, Masicampo & Vohs, that demonstrates how consciousness can influence our actions in the long run (2011; see also Monroe & Malle 2010; Stillman et al., 2011). On the other hand, he also appeals to simple everyday-life experience that strongly indicates that distal intentions are efficacious: "When I form the intention, while checking the airline website and making the relevant reservation, to take the 2:15 flight to London next Wednesday, I usually find myself in that airplane on that day" (Slors

2015, p.98). Slors further refines his argument by borrowing from Dretske a distinction between two kinds of causation: 'triggering causes' and 'structuring causes' (Dretske 1988, p.43). He argues that effective conscious intentions are indirect and should be seen as structuring causes of our actions, while our actions are possibly always triggered unconsciously by internal or external stimuli. Conscious distal intention formation is, according to him, a structuring cause, which could be conceived as a form of 'self-programming': "it causes us to be 'programmed' to be responsive in specific ways to specific stimuli in specific circumstances" (Slors 2015, p.106).

This fits well with the Peircean framework: deliberate habit-formation could be understood as a form of programming. A steady orientation towards certain preferred goals and ideals enables us to regulate our behaviour and ultimately form certain preferred habits, (i.e., to program ourselves).

So we have good reasons to believe that we not only can set distal goals and ideals for ourselves but also that these goals and ideals could affect our future behaviour and enable us to form a self that we would like to identify with. Longer-term goals and ideals provide orientation and prevent us from fully disintegrating in contingent interactions. They enable us to transcend our current state: we can identify ourselves with some urges and inclinations that we find in ourselves and distance ourselves from others. I can, for example, set for myself the longer-term goal to live healthier. This goal can help me to regulate my behaviour: it enables me to monitor and control my eating habits, for example. I will be responsive to urges (e.g., "mmm, I would kill for a hamburger") and short-term conscious decisions (e.g., "let's drive to Burger King for the last time this week") that do not fit my health programme, find ways not to succumb to these urges and intentions, and might gradually adjust my habits in correspondence with my long-term goal.

Now it also becomes clearer how biological inclinations, as well as conscious short-term intentions and decisions, can be reinterpreted from the proposed perspective. Biological inclinations and short-term conscious intentions, which might be instigated by our brain and/or our surroundings, can influence or even greatly determine our actions, but we do not necessarily have to identify ourselves with them. By virtue of longer-term goals and ideals, we can reflect on and effectively adjust our inclinations and short-term intentions and form ourselves in a more preferred direction.

Our account emphasizes that the self is not an isolated, a priori existing entity that 'immediately' determines its decisions and actions from something 'within', but that it rather discovers and forms itself by virtue of an orientation towards goals and ideals provided by its surroundings. From this perspective, the self is never a completely autonomous author of its decisions and actions but is greatly heteronomous and receptive. The notion of the 'mediated self' neatly expresses that the self is no entity from which original self-conception and self-experience are derived, but that the self rather 'indirectly' conceives, experiences and forms itself by virtue of certain identifications.

4 Brain imaging technologies and self-formation

Our proposal to understand the self in terms of self-formation attempts to offer an alternative account to the view that the self is primarily a manifestation of brain activity. At the same time, the proposed perspective aims to recognize that brain imaging

technologies could play a significant role in processes of self-formation. Brain imaging technologies can potentially explicate how our neurophysiology both sets constraints on and offers opportunities for the development of specific habits that help or fail to help us realizing our long-term ideals.

The use of brain imaging technologies breaks, as we have argued, with the Cartesian idea that privileged, authoritative knowledge of the self can only be acquired through introspection. In section 2, we touched upon potential far-reaching implications of this break: we have shown how neuroscientific research might, for example, be able to expose cases of cognitive dissonance, thereby challenging an individual's belief that she has good reasons to engage in certain behavior. The potential of brain imaging technologies to externally acquire knowledge that could even challenge (though might also complement) subjective beliefs of the self seems to make them unique. However, earlier psychological techniques such as Rorschach-tests or Intelligence quotient (IQ) tests have also been used as external sources to attain knowledge of the self. What then makes brain imaging more exceptional than these older techniques?

First of all, we acknowledge that there is a continuum between brain imaging technologies and earlier psychological techniques: both attempt to define or shed light on people's personalities, as well as on how people relate to themselves, thereby providing scientific discourses that offer an implicit normative foundation of the self by setting standards that constitute what is 'normal' or 'good' (e.g., Danziger 1994; Hacking 1986). Analogously, we can say that both embody principles that dictate what are good processes of self-formation, and offer normative guidelines that we could set for ourselves. In the words of Ian Hacking: brain imaging technologies can also be understood as a way to "make up people" (1986). That is, they could, along with other factors, constitute, facilitate and constrain the possibilities through which we can interpret and shape ourselves.

Moreover, methodological challenges to the idea of brain scans as direct representations both of the brain, and of the self, seem to further relativize the difference between imaging technologies and for example, Rorschach-tests. For example, Roskies argues that brain scans should not be understood as pictures of the brain because the information needed to interpret them is for the most part external to the scans themselves (2007). Furthermore, she illustrates that the reading of brain scans cannot be cut loose from the experimental circumstances in which several interpretational choices are already made. Additionally, Klein indicates that brain regions and networks are involved in multiple processes and cannot be tied to particular functions. This suggests a variety of possible explanations for the activity that is visually represented on a brain scan (Klein 2010).

Both earlier psychological techniques and brain imaging technologies seem to offer a particular view of the self, which, when accepted, competes with other possible descriptions of human behavior and self-formation. A Rorschach-test conceptualizes the self in terms of different personality traits, whereas brain scans offer a view of the self that is grounded in neurophysiological variables. This continuity between brain imaging and older technologies suggests taking a critical approach to the omnipresence of the neuroscientific discourse (e.g., Dumit 2004; Slaby 2010).

However, the fact that brain scans are not representations of the brain/self does not imply that there is no significant difference between them and older technologies. In contrast to techniques like Rorschach-tests, it is characteristic of brain imaging

technologies that they *foreground* physiological responses. In doing so, brain imaging technologies allow us to make sense of—and possibly influence—how our neurology co-constitutes the habits that we develop, and in turn enables us to orient towards our long-term ideals. This foregrounding is a new form of technological mediation that opens up new ways of self-formation to which brain imaging technologies can pragmatically contribute. In the case of earlier psychological techniques, neurophysiological interactions remained silent processes implicitly running in the background, yet of significant influence on our actions and habits. If, however, our biological and neurological possibilities and constraints significantly influence our self-formation, foregrounding those provides the opportunity to more explicitly and deliberately relate to them.

Neurofeedback, developed in the last decade, is an illustrative example that embodies the potential for individuals to respond to physiological processes, and see the physiological results of their interventions in real-time. This is not to say that technologies such as neuroimaging and neurofeedback must be considered an objective correction of our beliefs and experiences, since the relevance of physiological interactions remains something to be judged and acted upon by the self in light of its long-term ideals. Rather, we suggest that these technologies open up the opportunity to critically reflect on the relation between urges and habits that we are aware of, and the physiological processes that correlate with them. Making this correlation explicit helps us in establishing a relation with something the self could not previously relate to, which allows it to develop an explicit stance towards it in light of its long-term goals. Accordingly, the opportunities and constraints that individuals are confronted with can be pragmatically formed, not only on a psychological, but also on a physiological level.

5 Extrospection as critical self-identification

Neuroscientific developments seemingly allow individuals to interact with patterns of brain activity, which could contribute to forming a preferred or more preferable self. Our proposal to understand such interactions is similar to the promising 'neurophenomenology' program launched by Varela, which aims to align first-person and third-person perspectives on conscious experience (Varela 1996). Neurophenomenology attempts to integrate phenomenological qualitative experiences with existing quantitative research in the neurosciences in order (1) to allow neuroscientists to better interpret neurophysiological data that is relevant to conscious experience (Ibid.), and (2) to open up new possibilities for individuals to relate to physiological processes that correlate with their conscious experiences (Lutz & Thompson 2003). Thus far, primarily the former has been examined, focusing on how phenomenology can be used to epistemologically warrant neuroscientists to scientifically investigate conscious experience (e.g., Gallagher & Varela 2003; Garrisson et al. 2013; Lutz & Thompson 2003). In the present context, especially the latter is of interest because it takes seriously the potential to use measurements of brain activity for self-formation purposes.

We share the aim of neurophenomenology to move beyond a computational theory of cognition that interprets brain scans as representing how human cognition is functionally realized (Varela et al., 1991). However, our emphasis on *critical* self-

formation distinguishes our proposal from the project of neurophenomenology. Our aim is not to develop a methodology grounded in phenomenology that allows for the scientific investigation of conscious experience. Instead, we ask how brain imaging technologies can provide knowledge that allows relating our neurophysiology to the long-term goals and ideals we have set. The knowledge offered by brain imaging technologies we interpret as pragmatically allowing individuals to critically interact with processes that influence their self-formation. Labeling this process *extrospection* indicates that brain imaging technologies instantiate the existence of new perceptual experiences that can be used to deliberately intervene in our self-formation (Aydin 2015).[3]

By making explicit whether our habits resonate with our ideals in terms of their neurophysiological origin, brain imaging technologies offer insight into how our neurological set-up affects and is affected by the way we form ourselves. This allows for establishing an explicitly critical relation with the way our conscious and deliberate self-formation interferes with—and *is* interfered in by—our neurophysiology. Brain imaging technologies allow establishing such a critical relation by foregrounding our neurophysiology, which is an aspect of our world that was previously out of sight. Consequently, brain imaging technologies can be incorporated in the Peircean framework outlined in section 3.

However, it is pivotal to show to what extent it is really possible to *critically* relate to our self-formation because understanding the self as a formative process in the world runs the risk of falling into a tenacious determinism, as is expressed by Gallagher and Crisafi (2009). They propose that scientific, social, legal, and educational institutions can be understood in terms of an extended mind at work; in their words: "we take one of these cognitively produced things and we use it for further cognitive production. Our subsequent acts of cognition are facilitated or enhanced or made possible by particular tools or institutional mechanisms" (Ibid., p.51). While they believe that it is productive to think about cognition as distributed over humans, nonhumans, and institutions, there also seems to be a downside: "the extended mind can come back to bite us; it can place limitations on our thinking, as easily as it can enable great and wonderfully extended cognitive performances" (Ibid., p.51). Social and institutional forces not only offer opportunities for self-formation, but also put constraints on how self-formation can take place. How does one deal with the possibility that our very thinking as "a human enterprise, and as an individual practice" has been formed by the particular institutions "that we have invented to facilitate or enhance cognition" (Gallagher & Crisafi 2009, p.51)? In sum, are we still free to relate to and shape our mind, if it is embedded in socio-institutional mechanisms? And, if so, how must the potential to critically relate to self-formation be understood in light of the forces that influence its development?

[3] In the last decade, neuroscientists have made important steps towards the possibility of decoding mental states using *brain-computer interfacing* (BCI), fMRI, and other techniques (e.g., Haynes & Rees 2006; Poldrack 2011). Most famously, such developments bear the promise of decoding mental states of patients with locked-in syndrome such that these can be communicated to caregivers and family members (e.g., Chaudhary et al. 2017; Vansteensel et al. 2016). In our vocabulary, it can be said that these new techniques allow to obtain extrospective knowledge about the mental states of *others* (e.g., locked-in syndrome patients) from a third-person perspective, rather than that they allow the self to relate to its *own* physiology from a first-person perspective. Because of this, an extensive discussion of these important developments is beyond the scope of this paper.

Catherine Malabou has strongly highlighted this downside: she has argued that there is a close connection between understanding ourselves in terms of the plasticity of the brain, and the societal demand to be flexible citizens and employees (2008). Plasticity refers to the idea that neuronal interactions in our brain are not fixed once and for all, but change in relation to both societal demands and our own preferences. According to Malabou, when uncritically accepted the plasticity of our brain functions as a scientific legitimization of the societal status quo, as a neurological legitimation of societal demands. She holds that the idea of the plasticity of our brain conforms to the demand to be capable of easily switching between different tasks, thereby denying the importance of individual variability: "In effect, anyone who is not flexible deserves to disappear" (Ibid., p.46).

Even if we do not fully agree with Malabou's diagnosis, it points out that perhaps the institutional frameworks through which we are shaped render critical self-identification and self-formation impossible: the self might be shaped by forces and constraints with which it cannot critically identify. Malabou's analysis brings to the fore that the idea itself that our actions, goals and ideals are predetermined by our neurophysiology can be understood as one of the institutional forces that put constraints on our own process of self-formation.

However, the plasticity of the brain not only allows society to impose its demands on it, but it also makes it possible for us to actively engage and manipulate it. It is not either-or. Malabou recognizes this 'affordance' too: plasticity does not only allow the environment to shape the brain but it also makes it possible for us to form our neuronal interactions relative to our own preferred ideals (Ibid., p.11). For example, someone who desires to become a piano player could counterbalance an environment that does not encourage developing musical skills by practicing regularly on a piano available on, for example, trains stations (which has become fashionable in the Netherlands), thereby shaping the neuronal and synaptic interactions in the brain through which piano skills are co-shaped. Forming these interactions is the consequence of the desire to develop oneself in this way. Accordingly, Malabou argues that the structure of society is not a mere reflection of what is going on in the brain anyway, but that it is possible to actively form and adjust it, also even if in dissonance with societal preferences (Ibid., p.11).

By foregrounding our neurophysiology and offering extrospective knowledge, brain imaging technologies can disclose possibilities to form the brain more effectively. Instead of offering theoretical-conceptual knowledge of the plasticity of our brain—what seems to be Malabou's proposal—they allow for the practical-perceptual visualization of one's neurophysiology. Doing so, brain imaging technologies can offer insight in how we more effectively can pursue our long-term ideals by revealing how our habits relate to our neurophysiological make-up. Moreover, they could display that a change in our habits is also accompanied with a neurophysiological change. Brain imaging technologies could help us to gradually make the transition from who we are and who society want us to be to who we eventually want to be; we can, for example, do something about our urges to give in to instigations from the market to consume fast food.

In short, brain imaging technologies need not be understood as tools to represent our inner realm, but can be alternatively viewed as tools to help us orient to our own ideals by offering extrospective knowledge of biological factors that conflict with them. This new form of knowledge is a form of *critical* self-identification because it makes possible a relation with the neurological constraints and possibilities that affect our

behaviour. The development and use of technologies such as brain stimulation and neurofeedback takes, as we will show in the next section, Malabou's narrative one step further: these technologies illustrate how we could actively change brain processes that both enable and constrain the habits that we develop.

6 Self-formation and Neurofeedback

The proposed framework can be further clarified when discussed in the context of neurofeedback. Recently, it has been suggested that the paradigm of 'neurofeedback' can have a unique function in reducing *food craving*, a phenomenon closely associated with developing obesity (e.g., Bowell & Kober 2016; Volkow et al., 2011; Weingarten & Elston 1990). The neural pathways that give rise to *food craving* could display the neurological constraints and opportunities in attempts to form ourselves in a particular direction. Food craving can be defined as an intense desire to consume a particular food or food type that is difficult to resist (Weingarten & Elston 1990). From the perspective of self-formation, such short-term desires function as constraints on the long-term ideal to form oneself as a healthy individual. Food craving correlates with patterns of brain activity that potentially undermine the realization of a particular long-term ideal. If these patterns are the consequence of (earlier) 'bad' habits, they potentially can also be altered by developing new habits.

Previous studies suggest that non-invasively stimulating specific regions of the brain significantly reduces food craving by modulating neuronal interactions (e.g., Goldman et al. 2011; Ljubisavljevic et al. 2016). Apparently, our neurology not only functions as a constraint on our long-term ideals, but also offers the opportunity to develop new habits by altering neural pathways. Brain stimulation technologies could potentially help us to develop new habits that would counterbalance those patterns of brain activity which instantiate food craving.

However, while non-invasive brain stimulation still relies on external intervention, neurofeedback is closely associated with neurophenomenology and explicitly directed at giving individuals the possibility to relate to their neurophysiology by making it perceptually present (e.g., Bagdasaryan & Le Van Quyen 2013). Through neurofeedback—and this is what differentiates it from brain stimulation techniques—individuals have the opportunity to perceive patterns of localized brain activity related to food craving, and voluntarily regulate them (e.g., Bagdasaryan & Le Van Quyen 2013; Bartholdy et al. 2013). Respondents receive feedback via real-time displays of EEG or fMRI on how they respond to certain stimuli. Relating to these displays enables participants to actively influence their brain activity by attending to it in a more focused manner. This fits well with our view of self-formation: while brain stimulation techniques influence habit formation by *force*, neurofeedback makes it possible for the self to establish a relation with something previously invisible by turning brain activity into something that the self can act upon.

Although the clinical merit of neurofeedback is still in a trial phase, positive results were found in areas beyond food craving. For example, the use of neurofeedback reportedly had a positive effect on the ability of children diagnosed with ADHD to concentrate (Gevensleben et al. 2009), the capability of individuals to control addiction-related impulses by managing dopamine levels (Sulzer et al. 2013), and helped increasing the mood of people suffering from *Major Depressive Disorder* (MDD) by self-regulating amygdala responses (Young et al. 2014).

These studies support our idea that brain imaging technologies can potentially inform processes of self-formation, because they offer extrospective knowledge, reportedly enabling experimental participants to be in better control of their own brain activity (e.g., Micoulaud-Franchi et al. 2014). This illustrates how brain imaging technologies could contribute to the formation of a self that better satisfies our ambitions and life projects. Since technologies mediate our self-identification, they can be integrated in our practices in such a way that they can help to bridge possible tensions between how we act and how we would like to act. Neurofeedback makes the discrepancy between short-term urges and long-term goals accessible in terms of one's neurophysiology, bringing to the fore a dialectic between the process of self-formation and self-identification. When being confronted with the neurological constraints that constitute our habits, at the same time we are confronted with the opportunity to actively manipulate them, and turn them into ways to more effectively realize our long-term goals. Let us stress again that it is still unclear to what extent neurofeedback is effective (cf. Thibault et al. 2018). Although we are agnostic about its scientific status, this example illustrates how brain imaging technologies *could* be used in self-formation processes.

7 Conclusion

In this paper, we have proposed a pragmatic understanding of brain imaging technologies as a source for extrospection that can be used in self-formation. This pragmatic approach to brain imaging technologies shares with enactivist and neurophenomenological perspectives the idea that the self is not a distinct identifiable entity, and that subjective experiences should be integrated into scientific investigations on how aspects of the self are realized (cf. Gallagher 2007; Gallagher & Varela 2003). However, because of our focus on self-formation, we have brought to the fore the relevance of the knowledge that can be obtained through brain imaging technologies for the experiencing subject.

We suggested that brain imaging technologies can be understood as constituting the possibility to develop new ways of self-formation. By making previously unnoticed neurophysiological processes perceptually present, brain imaging technologies confront individuals with how their habits are co-constituted by neurophysiological processes. Our pragmatic approach reveals that brain imaging technologies do not offer representations of the self, but they allow us to attain extrospective knowledge of self-formation. We have used the example of neurofeedback to show how current neuroscientific developments could be integrated in our proposal: using neurofeedback, individuals can critically reflect on whether their habits are consistent with the long-term goals or ideals they are oriented towards.

Instead of understanding brain imaging technologies in terms of their representational qualities, we propose to conceive them in terms of how they mediate new opportunities for self-formation by *making present* our neurophysiology and the way it co-constitutes our habits. By attaining extrospective knowledge of our physiology, brain imaging technologies potentially allow for intervening in the habits we form when pursuing our existing long-term ideals. Understood in this way, brain imaging technologies can contribute to self-formation by making explicit the neurological constraints and possibilities that coincide with our (unwanted) habits, without relying on a representationalist framework.

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

Aydin, C. (2009). On the significance of ideals. Peirce and the good life. *Transactions of the C.S. Peirce Society, 45*, 422–443.

Aydin, C. (2015). The artifactual mind: Overcoming the 'inside-outside' dualism in the extended mind thesis and recognizing the technological dimension of cognition. *Phenomenology and the Cognitive Sciences, 14*(1), 73–94.

Aydin, C. (2018a). From camera obscura to fMRI: How brain imaging technologies mediate free will. In J. Aagaard, J. K. Berg Friis, J. Sorenson, O. Tafdrup, & C. Hasse (Eds.), *Postphenomenological methodologies: New ways in mediating techno-human relationships* (pp. 103–121). Lanham: Rowman Littlefield Press/Lexington Books.

Aydin, C. (2018b). World oriented self-formation as sublimation: Or why postphenomenology needs Peircean pragmatism. *Cognitio, 19*(2), 204–219.

Bagdasaryan, J., & Le Van Quyen, M. (2013). Experiencing your brain: Neurofeedback as a new bridge between neuroscience and phenomenology. *Frontiers in Human Neuroscience, 7*, 680.

Bartholdy, S., Musiat, P., Campbell, I. C., & Schmidt, U. (2013). The potential of neurofeedback in the treatment of eating disorders: A review of the literature. *European Eating Disorders Review, 21*(6), 456–463.

Bailey, L. W. (1989). Skull's darkroom: The camera Obscura and subjectivity. In P. T. Durbin (Ed.), *Philosophy of technology: Practical, historical and other dimensions* (pp. 63–79). Dordrecht: Kluwer Academic Publishers.

Baumeister, R. F., Masicampo, E. J., & Vohs, K. D. (2011). Do conscious thoughts cause behavior? *Annual Review of Psychology, 62*, 331–361.

Bowell, R. G., & Kober, H. (2016). Food cue reactivity and craving predict eating and weight gain: A meta-analytic review. *Obesity Reviews, 17*(2), 159–177.

Chaudhary, U., Xia, B., Cohen, L. G., & Birbaumer, M. (2017). Brain-computer interface-based communication in the completely locked-in state. *PLoS Biology, 15*(1), e1002593.

Christoff, K., Cosmelli, D., Legrand, D., & Thomspon, E. (2011). Specifying the self for cognitive neuroscience. *Trends in Cognitive Sciences, 15*(3), 104–112.

Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis, 58*(1), 7–19.

Danziger, K. (1994). *Constructing the subject: Historical origins of psychological research*. Cambridge: Cambridge University Press.

Davey, C. G., Pujol, J., & Harrison, B. J. (2016). Mapping the self in the brain's default mode network. *Neuroimage, 132*, 390–397.

Descartes R. (1965). *Discourse on method, optics, geometry, and meteorology*. In Paul J. Olscamp (transl.). New York/Kansas City: The Bobbs-Merrill company, Inc.

Devue, C., & Brédart, S. (2011). The neural correlates of visual self-recognition. *Consciousness and Cognition, 20*(1), 40–51.

Doris, J. (2002). *Lack of character: Personality and moral behavior*. Cambridge: Cambridge University Press.

Dretske, F. (1988). *Explaining behavior: Reasons in a world of causes*. Cambridge: MIT Press.

Dumit, J. (2004). *Picturing personhood: Brain scans and biomedical identity*. Princeton: Princeton University Press.

Farrer, C., & Frith, C. D. (2002). Experiencing oneself vs another person as being the cause of an action: The neural correlates of the experience of agency. *Neuroimage, 15*(3), 596–603.

Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford: Stanford University Press.

Frankfurt, H. (1971). Freedom of the will and the concept of a person. *The Journal of Philosophy, 68*, 5–20.

Fried, I., Mukamel, R., & Kreiman, G. (2011). Internally generated preactivation of single neurons in human medial frontal cortex predicts volition. *Neuron, 69*, 548–562.

Gallagher, S. (2007). Neurophilosophy and Neurophenomenology. In L. Embree & T. Enon (Eds.), *Phenomenology 2005* (Vol. 5, pp. 293–316). Bucharest: Zeta Press.

Gallagher, S. (2013). A pattern theory of self. *Frontiers in Human Neuroscience, 7*, 443.

Gallagher, S., & Crisafi, A. (2009). Mental institutions. *Topoi, 28*(1), 45–51.

Gallagher, S., & Daly, A. (2018). Dynamical relations in the self-pattern. *Frontiers in Psychology, 9*, 664.

Gallagher, S., & Varela, F. (2003). Redrawing the map and resetting the time: Phenomenology and the cognitive sciences. *Canadian Journal of Philosophy. Supplementary, 29*, 93–132.

Garrisson, K. A., Scheinost, D., Worhunsky, P. D., Elwafi, H. M., Thornhill IV, T. A., et al. (2013). Real-time fMRI links subjective experience with brain activity during focused attention. *Neuroimage, 81*(November 1), 110–118.

Gevensleben, H., Holl, B., Albrecht, B., Schlamp, D., Kratz, O., et al. (2009). Distinct EEG effects related to neurofeedback training in children with ADHD: A randomized controlled trial. *International Journal of Psychophysiology, 74*, 149–157.

Gillihan, S. J., & Farah, M. J. (2005). Is self special? A critical review of evidence from experimental psychology and cognitive neuroscience. *Psychological Bulletin, 131*(1), 76–97.

Goldman, R. L., Borckardt, J. J., Frohman, H. A., O'Neil, P. M., Madan, A., et al. (2011). Prefrontal cortex transcranial direct current stimulation (tDCS) temporarily reduces food cravings and increases the self-reported ability to resist food in adults with frequent food craving. *Appetite, 56*(3), 741–746.

Hacking, I. (1986). Making up people. In T. L. Heller, M. Sosna, & D. E. Wellbery (Eds.), *Reconstructing individualism: Autonomy, individuality, and the self in Western thought* (pp. 222–236). Stanford: Stanford University Press.

Haggard, P. (2017). Sense of agency in the human brain. *Nature Reviews Neuroscience, 18*, 196–207.

Haynes, J.-D., & Rees, G. (2006). Decoding mental states from brain activity in humans. *Nature Reviews Neuroscience, 7*, 523–534.

Ihde, D. (1990). *Technology and the lifeworld: From garden to earth*. Indianapolis: Indiana University Press.

Ihde, D. (2007). Art precedes science. In P. Kockelkoren (Ed.), *Mediated vision* (pp. 25–37). Arnhem: ArtEZ/ Veenman Publishing.

Kitayama, S., Chua, H. F., Tompson, S., & Han, S. (2013). Neural mechanisms of dissonance: An fMRI investigation of choice justification. *Neuroimage, 69*, 206–212.

Klein, C. (2010). Images are not the evidence in neuroimaging. *The British Journal for the Philosophy of Science, 61*(2), 265–278.

Klemm, R. (2010). Free will debates: Simple experiments are not so simple. *Advances in Cognitive Psychology, 6*, 47–65.

Kockelkoren, P. (2003). *Technology: Art, fairground and theatre*. Rotterdam: NAi Publishers.

Legrand, D., & Ruby, P. (2009). What is self-specific? Theoretical investigation and critical review of neuroimaging results. *Psychological Review, 116*(1), 252–282.

Ljubisavljevic, M., Maxood, K., Bjekic, J., Oommen, J., & Nagelkerke, N. (2016). Long-term effects of repeated prefrontal cortex transcranial direct current stimulation (tDCS) on food craving in normal and overweight young adults. *Brain Stimulation, 9*(6), 826–833.

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.

Macdonald, C. (2007). Introspection and authoritative self-knowledge. *Erkenntnis, 67*(2), 355–372.

Malabou, C. (2008). *What should we do with our brain? (S. Rand, trsl.)*. New York: Fordham University Press.

Micoulaud-Franchi, J. A., Quiles, C., Fond, G., Cermolacce, M., & Vion-Dury, J. (2014). The covariation of independent and dependant variables in neurofeedback: A proposal framework to identify cognitive processes and brain activity variables. *Consciousness and Cognition, 26*, 162–168.

Monroe, A. E., & Malle, B. F. (2010). From uncaused will to conscious choice: The need to study, not speculate about people's folk concept of free will. *Review of Philosophy and Psychology, 1*, 211–224.

Northoff, G., Heinzel, A., de Greck, M., Bermpohl, F., Dobrowolnu, H., & Panksepp, J. (2006). Self-referential processing in our brain – A meta-analysis of imaging studies on the self. *Neuroimage, 31*, 440–457.

Peirce, C.S. (1931-1935, 1958). *Collected papers of C.S. Peirce*. In Ch. Hartshorne & P. Weiss; A. Burks (Eds). Cambridge: Harvard University press, CP 8.320 (abbreviation: CP followed by '[volume].[page]').

Peirce, C.S. The essential Peirce. Hauser, N. & Kloesel, N.; The Peirce Edition Project (Eds.). Bloomington and Indianapolis: Indiana University Press. 1992 and 1998 (EP followed by '[volume], [page]').

Poldrack, R. A. (2011). Inferring mental states from neuroimaging data: From reverse inference to large-scale decoding. *Neuron, 72*(5), 692–697.
Roskies, A. L. (2007). Are neuroimages like photographs of the brain? *Philosophy of Science, 74*(5), 860–872.
Sartre, J. P. (1943). *Being and nothingness: An essay on phenomenological ontology (H.E. Barnes, trsl.)*. London: Routledge.
Slaby, J. (2010). Steps towards a *Critical Neuroscience*. *Phenomenology and the Cognitive Sciences, 9*(3), 397–416.
Slors, M. (2015). Conscious intending as self-programming. *Philosophical Psychology, 28*(1), 94–113. https://doi.org/10.1080/09515089.2013.803922.
Soon, C .S., Brass, M., Heinze, H.-J., & Hayned, J.-D. (2008). Unconscious determinants of free decisions in the human brain. *Nature Neuroscience, 11*, 543–545.
Stillman, T. F., Baumeister, R. F., & Mele, A. R. (2011). Free will in everyday life: Autobiographical accounts of free and unfree actions. *Philosophical Psychology, 24*, 381–394.
Sulzer, J., Sitaram, R., Blefari, M. L., Kollias, S., Birbaumer, N., Stephan, K. E., Luft, A., & Gassert, R. (2013). Neurofeedback-mediated self-regulation of the dopaminergic midbrain. *Neuroimage, 83*, 176–184.
Taylor, C. (1989). *Sources of the self: The making of modern identity*. Cambridge: Harvard University Press.
Taylor, C. (1992). *The ethics of authenticity*. Cambridge: Harvard University Press.
Thibault, R. T., MacPherson, A., Lifshitz, M., Roth, R. R., & Raz, A. (2018). Neurofeedback with fMRI: A critical systematic review. *Neuroimage, 172*, 786–807.
Vansteensel, M. J., Pels, E. G. M., Bleichner, M. G., Branco, M. P., Denison, T., Freudenburg, Z. V., Gosselaar, P., Leinders, S., Ottens, T. H., Van Den Boom, M. A., Van Rijen, P. C., Aarnoutse, E. J., & Famsey, N. F. (2016). Fully implanted brain-computer interface in a locked-in patient with ALS. *The New England Journal of Medicine, 375*(21), 2060–2066.
Varela, F. (1996). Neurophenomenology: A methodological remedy for the hard problem. *Journal for Consciousness Studies, 3*(4), 330–349.
Varela, F., Thompson, E., & Rosch, E. (1991). *The embodied mind: Cognitive science and human experience*. Cambridge: The MIT Press.
Verbeek, P. P. (2005). *What things do: Philosophical reflections on technology, agency, and design*. University Park: Pennsylvania State University Press.
Vidal, F., & Ortega, F. (2017). *Being brains: Making the cerebral subject*. New York: Fordham University Press.
Vincent, N. A. (Ed.). (2013). *Neuroscience and legal responsibility*. Oxford: Oxford University Press.
Vogeley, K., & Gallagher, S. (2011). Self in the brain. In S. Gallagher (Ed.), *The Oxford handbook of the self* (pp. 111–136). Oxford: Oxford University Press.
Volkow, N. D., Wang, G., & Baler, R. D. (2011). Reward, dopamine and the control of food intake: Implications for obesity. *Trends in Cognitive Science, 15*(1), 37–46.
de Vries, J., Byrne, M., & Kehoe, E. (2015). Cognitive dissonance induction in everyday life: An fMRI study. *Social Neuroscience, 10*(3), 268–281.
Wegner, D. M., & Wheatley, T. (1999). Apparent mental causation: Sources of the experience of will. *American Psychologist, 54*, 480–492.
Weingarten, H. P., & Elston, D. (1990). The phenomenology of food cravings. *Appetite, 15*(3), 231–246.
Williams, L. E., & Bargh, J. A. (2008). Experiencing physical warmth promotes interpersonal warmth. *Science, 322*, 606–607.
Young, K. D., Zotev, V., Phillips, R., Misaki, M., Yuan, H., Drevets, W. C., & Bodurka, J. (2014). Real-time fMRI neurofeedback training of amygdala activity in patients with major depressive disorder. *PLoS One, 9*(2), e88785.
Zahavi, D. (2015). Self and other: From pure ego to co-constituted we. *Continental Philosophy Review, 48*(2), 143–160.

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