PhiMiSci
Philosophy and the Mind Sciences

Breaking the self

Radical disruptions of self-consciousness and impossible conscious experiences

Wanja Wiese[a] (wawiese@uni-mainz.de)

Abstract
Are there logically possible types of conscious experience that are *nomologically impossible*, given independently justified assumptions about the neural underpinnings of consciousness in human beings? In one sense, this is trivial: just consider the fact that the types of perceptual experiences we can have are limited by our sensory organs. But there may be non-trivial types of conscious experience that are impossible. For instance, if there is a basic type of self-consciousness, corresponding to a phenomenal property that is nomologically necessary for consciousness, then experiences lacking this phenomenal property will be (nomologically) impossible. More generally, it may be that there are causal dependencies between the neural mechanisms that are required to instantiate distinct phenomenal properties (in human beings). If this is the case, instantiating one of these phenomenal properties without certain others may be impossible, which means there are non-trivial cases of nomologically impossible types of conscious experience. This paper clarifies this hypothesis, outlines a general methodology for its investigation, and relates it to research on radical disruptions of self-consciousness.

Keywords
Consciousness · Impossible conscious experiences · Self-consciousness · Sense of agency · Sense of boundaries · Sense of time

This article is part of a special issue on "Radical disruptions of self-consciousness", edited by Thomas Metzinger and Raphaël Millière.

1 Introduction

This paper explores the hypothesis that some logically possible configurations of self-consciousness are *nomologically impossible*, given certain assumptions about consciousness and its neural underpinnings in human beings. Put differently, an impossible conscious experience, as defined here, is an experience that cannot be

[a]Johannes Gutenberg University Mainz

Wiese, W. (2020). Breaking the self: Radical disruptions of self-consciousness and impossible conscious experiences. *Philosophy and the Mind Sciences*, *1*(I), 10.
https://doi.org/10.33735/phimisci.2020.I.32
 https://philosophymindscience.org ISSN: 2699-0369

generated by human brains.[1] Radical disruptions of self-consciousness involving a total loss of self-consciousness are logically possible, but some authors doubt that they are nomologically possible (given the human neural architecture). If correct, this would provide an example of an impossible conscious experience. However, as some of the other contributions to this special issue show, that claim is at least controversial (and probably false, see Letheby, 2020; Metzinger, 2020; Millière, 2020, all in this special issue; for the opposite view, see Sebastián, 2020, this special issue) – hence, I will not take a stance on the question whether conscious experiences without any kind of self-consciousness whatsoever are possible.

Instead, I will pursue slightly more general goals. First, I will argue that research on impossible conscious experiences is relevant to understanding consciousness. If we had a complete theory of consciousness (including detailed knowledge of the neural mechanisms of consciousness), we would be able to say which types of conscious experience are possible and which are not. I shall argue that the reverse statement is also true: impossible experiences can provide insights into the neural mechanisms of consciousness. Secondly, I will propose a general methodology for finding evidence for impossible conscious experiences. In addition, I will argue that radical disruptions of self-consciousness (whether they involve a total loss of self-consciousness or not) provide particularly good opportunities for finding candidates for impossible experiences.

In what follows, I will first define what an impossible conscious experience is (section 2), and why research on impossible conscious experiences is relevant (section 3). Then I will propose general strategies to investigate impossible conscious experiences (section 4). Finally, I will apply one of these strategies to an example, viz. the hypothesis that having a sense of agency without a sense of time is impossible (section 5).

2 What exactly is an impossible conscious experience?

The brain is an extremely plastic organ that is capable of generating a seemingly unbounded number of different conscious experiences.[2] At the same time, the brain is a finite organ, which has not been selected for its ability to generate the largest number of logically possible experiences, but for its ability to enable flexible control of behavior in changing environments. Therefore, it is likely that some types of conscious experience are logically possible, but in fact impossible *for creatures like us.* This is the main hypothesis of this paper: There are logically possible

[1] Unless indicated otherwise, I will use "possible" and "impossible" to refer to nomological possibility throughout this paper.

[2] I will use the terms "conscious experience", "experience", and "conscious state" interchangeably in this paper.

Wiese, W. (2020). Breaking the self: Radical disruptions of self-consciousness and impossible conscious experiences. *Philosophy and the Mind Sciences, 1*(I), 10.
https://doi.org/10.33735/phimisci.2020.I.32

types of conscious experience that are *nomologically impossible*,[3] given independently justified assumptions about the neural underpinnings of consciousness in human beings.[4] In the following, I shall use the term "impossible (conscious) experience" as a shorthand for this more complicated formulation (unless indicated otherwise).

This formulation raises a few questions: 1) How do we individuate types of conscious experience? 2) What are independently justified assumptions about consciousness and its neural underpinnings? 3) What does it mean to say that a conscious experience is nomologically impossible, given certain assumptions? The first question asks for individuation criteria for types of conscious experience. The second question asks about the relationship between types of experience and types of neural activity and structure. The third asks what bearing this relationship has on whether a given type of experience is nomologically possible or not. Let us address these questions in turn.

2.1 Individuation criteria

Here, I will stipulate that types of conscious experience can be individuated in terms of their phenomenal character, i.e., by the phenomenal properties instantiated by these experiences. Examples of phenomenal properties are experienced perceptual features, such as experienced redness, the experienced smell of a rose, a feeling of warmth in your body, the sense of mental ownership accompanying

[3] This is trivially true, in the sense that we cannot have experiences that would, say, require radically different sensory organs. I distinguish this from a different, substantial sense of impossibility in section 3.1 below.

[4] One could try to subsume this under a notion of "biological impossibility", derived from Tim Bayne's concept of *biological necessity*:

> Many features of human beings are not necessitated by our design specifications but vary from one biologically normal human being to another. [...] But other features of our nature – such as the possession of a heart – are mandated by our design specifications; they are included within our functional blueprint. [...] [A] feature can be 'biologically necessary' without being an essential feature of all possible human beings. (Bayne, 2014, pp. 526–527)

Biologically necessary features are features that all "biologically normal" human beings possess. As Bayne emphasizes, such features need not be essential features of all *possible* human beings, because one could change a person's "normal" features by intervening with their body. After the intervention, the person may continue to be a human being, without being "biologically normal". Instead of using the notion of biological normality to define biological necessity, one could use it to refer to features that *no* "biologically normal" human being possesses, and call these features "biologically impossible". However, this raises the question: What counts as normal and what does not? Answers to this question will, to a large extent, be a matter of debate, and there are likely to be many borderline cases. Furthermore, restricting the notion *biologically normal* to "normal human brains" raises similar questions. For this reason, I am using the more complicated formulation "nomologically impossible, given independently justified assumptions about consciousness and its neural underpinnings in human beings". I unpack this formulation in the main text.

Wiese, W. (2020). Breaking the self: Radical disruptions of self-consciousness and impossible conscious experiences. *Philosophy and the Mind Sciences*, *1*(I), 10.
https://doi.org/10.33735/phimisci.2020.I.32

many conscious thoughts, or a feeling of effort while thinking about a puzzling philosophical problem.

I will, furthermore, assume that phenomenal character is grounded in (phenomenally represented) intentional content. In particular, I assume that conscious experiences have accuracy conditions, and that they determine what it is like to have those experiences. The intentional content of a conscious experience is then a *phenomenal* content. Some philosophers distinguish between phenomenal content and the phenomenal mode of presentation:[5] the idea would be that the difference between, say, consciously seeing a fluffy puppy and consciously touching its fur is not exhausted by a difference in phenomenal content; as a result, such authors would distinguish between two types of phenomenal properties (see Chudnoff, 2013, pp. 565–566), or would argue that phenomenal properties are not determined by *pure* representational properties (i.e., one would also have to specify the particular phenomenal manner in which something is represented, such as a visual or auditory phenomenal manner – see Chalmers, 2004, pp. 159–161, for discussion). Here, I shall assume that there are no differences in phenomenal character that cannot be accounted for in terms of differences between representational content, and that one does not have to distinguish between two types of phenomenal properties: there are only phenomenal properties corresponding to phenomenal contents, and differences in the phenomenal mode or manner of representation can be analyzed as differences in contents – but note that the claims argued for later in this paper do not depend on these background assumptions.

Must two (token) experiences of the same type have the *same* phenomenal character? Or does *similarity* of phenomenal character suffice for membership of a given type of conscious experience (Koksvik, 2014, p. 129)? Relatedly, how determinate do the phenomenal properties picked out by a description of a type of experience have to be (Kriegel, 2015, p. 10)? For the purposes of this paper, nothing hinges on how we answer these questions. After all, we want to find descriptions of conscious experiences that cannot be created by the human brain. The (token) conscious experiences falling under a given description may have the same or only similar phenomenal characters (for instance, because they instantiate different determinates of the same determinable phenomenal property) – and there may be reasons to regard them as tokens of the same or of different types of experience. But what counts is what they have in common: they are nomologically impossible (independently of whether they share membership of the same type of experience or not).

[5] Just as one can distinguish between intentional content and intentional mode (see Crane, 2009, p. 477; Searle, 1983, p. 6), or between representative content and psychological mode (see Searle, 1979, p. 82).

Wiese, W. (2020). Breaking the self: Radical disruptions of self-consciousness and impossible conscious experiences. *Philosophy and the Mind Sciences, 1*(I), 10.
https://doi.org/10.33735/phimisci.2020.I.32

2.2 Consciousness and the brain

We cannot *find* impossible conscious experience (in the way in which black swans can be found), so we can at most hypothesize that a given description of a conscious experience does not refer to any nomologically possible conscious experiences. How can this hypothesis be tested? I shall suggest that we find evidence for a possible explanation as to *why* the type of conscious experience in question is impossible. Such an explanation must make some assumptions about the neural underpinnings of consciousness, because it purports to explain why the brain *cannot* generate this type of conscious experience. Furthermore, these assumptions should ideally either be statements for which strong empirical support already exists, or they should enable us to derive empirical hypotheses that can be tested in practice (not just in principle). In addition to this, there may also be useful assumptions that are to some extent theoretically justified (for instance, assumptions about general computational principles implemented by neural activity). These should at least be consistent with currently available empirical evidence. I shall say more about such assumptions below, in section 4.2.

2.3 Nomological possibility and the brain

Why do we have to make further assumptions about the brain to justify the claim that a given type of experience is nomologically impossible? After all, don't the laws of nature already put constraints on possible conscious experiences? This is certainly true, and there may be many interesting types of conscious experience that are nomologically impossible, but logically possible, regardless of how consciousness is generated in the human brain. However, as intimated in the previous paragraph, "finding" impossible experiences is methodologically difficult. Adding constraints that are more specific than constraints derived from physics alone can therefore facilitate finding candidates for impossible experiences. To the extent that these additional constraints suggest empirically testable hypotheses, we do not have to rely on theoretical considerations alone to justify the claim that a given type of experience is impossible.

A description of an impossible conscious experience, in the sense used in this paper, must therefore specify a phenomenal character (i.e., a phenomenal property) and suggest why the human brain is incapable of generating an experience with this phenomenal character. But why should some types of experience be impossible in the first place?

Here is a *prima facie* reason why we should expect that there are systematic limits on the types of conscious experience that the brain is capable of creating. The ability to generate consciousness is likely to serve an adaptive function, and this includes the ability to regulate consciousness or to shut it down. Extreme pain can cause subjects to faint, and this may have an adaptive value. Crucially, it also suggests that there is a threshold above which no pain experience is possible (because consciousness will be shut down if this threshold is crossed). Assuming

Wiese, W. (2020). Breaking the self: Radical disruptions of self-consciousness and impossible conscious experiences. *Philosophy and the Mind Sciences, 1*(I), 10.
https://doi.org/10.33735/phimisci.2020.I.32

that the intensity of logically possible pain experiences is unbounded (i.e., for every logically possible pain experience, there is another that is even more painful), this suggests that certain phenomenal properties ("more than extreme" painfulness) cannot be instantiated by the human brain. More generally, there may be other "extreme" phenomenal properties, or combinations of phenomenal properties, that cannot be instantiated by the human brain, because of the way in which the brain generates and regulates conscious experience.

To clarify and test these ideas, we need a more systematic way of arriving at descriptions of impossible conscious experiences; this will also allow us to specify how such experiences are still *logically* possible. In the following section, I will motivate and describe three different strategies.

3 Why are impossible conscious experiences relevant for consciousness research?

Any scientific theory needs a clear description of its target domain. Phenomenal consciousness is particulary vexing in this respect, because defining the explanandum is notoriously difficult. Although it seems we can, conceptually, distinguish phenomenal consciousness from mere tonic alertness (Metzinger, 2020), attention, the capacity for integrated control of behavior, and other cognitive capacities (see Chalmers, 1995; Revonsuo, 1998), we cannot completely dissociate phenomenal consciousness from these phenomena, and their relationship remains debated. This points to a central feature of theory-building in consciousness research, viz. that more rigorous specifications of the target phenomenon (i.e., phenomenal consciousness) have to rely on empirical results. As research progresses, we gain deeper insights into the relationship between phenomenal consciousness and cognitive capacities. For instance, we learn more about which perceptual and cognitive tasks can be performed unconsciously (Lamme, 2015) and which require consciousness.[6]

Discovering phenomena that do not (necessarily) involve conscious processing enables one to circumscribe the concept of consciousness. In addition to this, it is also possible to constrain the concept of consciousness from within, by considering tasks that cannot even be performed consciously. For instance, although some savants are able to perform highly remarkable forms of mental arithmetic, there are limits to what is humanly possible. Some people are able to consciously ex-

[6] One could object that such empirical discoveries do not tell us anything about the concept of consciousness, but merely contingent facts about how consciousness and cognitive capacities are related in actual human beings. In other words, it is conceivable that there are beings with the same cognitive capacities, but without consciousness, or beings with conscious experiences that are impossible for us. First, this objection seems to presuppose a strong separation between empirical investigations and conceptual analyses. Secondly, even if our current, folk-psychological concept of consciousness is extremely general, it can be useful to develop a more specific, scientific concept of consciousness, which will mainly apply to actual human beings.

Wiese, W. (2020). Breaking the self: Radical disruptions of self-consciousness and impossible conscious experiences. *Philosophy and the Mind Sciences, 1*(I), 10.
https://doi.org/10.33735/phimisci.2020.I.32

perience the process of multiplying astronomically large numbers, but no human being is able to mentally multiply two arbitrary numbers with one million digits each. Speaking more generally, there are limits on cognitive capacities (e.g., on the capacity of working memory[7]), and therefore also limits on consciousness, because one cannot[8] experience the exercise of a cognitive capacity if one does not possess this capacity in the first place.[9] Call this the *ability approach* to impossible conscious experiences.

A further approach is to consider phenomenal properties that vary on a continuum. A possible example is the experience of toothache, which varies from a dull feeling of pressure to a sharp pain of tremendous intensity. Other possible examples include the experienced brightness of a visual stimulus, the experienced loudness of a sound, or the experienced sweetness of a piece of food.[10] The core assumption underlying this approach is that there are levels of intensity that cannot be consciously experienced by human beings. Call this the *intensity approach*.[11] As a simple example (which is not related to consciousness), consider a flexible stick. The stick can be bent to some extent, but (as we shall assume) it will eventually break, if the bend becomes too extreme. Bending the stick beyond the point at which, as a matter of fact, it breaks is logically possible, but nomologically impossible (given the physical nature of the stick). Similarly, instantiating extreme phenomenal properties may be logically possible, but nomologically impossible (given the nature of our brains). This might also apply to some forms of self-consciousness: finding out at which point "the self breaks" can provide hints about the underlying neural mechanisms.

In this paper, I will pursue a third approach. The main assumption is that there are phenomenal properties that are dependent on each other, such that it is

[7] As is well-known, human working memory is limited to around 7 "items" (Miller, 1956). While the precise capacity of working memory may vary developmentally or be context-dependent (for a review, see Cowan, 2010), it is out of the question *that* the capacity of working memory is limited. For instance, states of consciousness that involve imagining a number of 30 specific chunks in working memory are nomologically impossible.

[8] Even if one imagined exercising a capacity one does not have, the experience could not be one of exercising that capacity, but would be a different experience.

[9] I am grateful to Martin Dresler for suggesting this.

[10] The claim that such properties vary on a continuum should be taken with a grain of salt. In particular, I do not want to claim that there are infinite different levels of intensity, nor that they can be measured in the way temperature can be measured. Following a suggestion by the editors of this special issue, it may even be more appropriate to say that what differently intense feelings of pain have in common is that they pertain to the same (natural) kind of phenomenal property, even though they are distinct phenomenal properties. The idea would then be that certain members of this kind of phenomenal property cannot be instantiated by human brains.

[11] Upon closer inspection, this approach may partly overlap with the ability approach. While the latter focuses on cognitive capacities, it may be possible to characterize the current approach in terms of perceptual abilities. For instance, the fact that there are limits on experienced loudness could also be formulated as the fact that the loudness of sounds cannot be discriminated beyond a certain threshold. This would presuppose that phenomenal differences go along with discriminatory abilities, which is not uncontested.

Wiese, W. (2020). Breaking the self: Radical disruptions of self-consciousness and impossible conscious experiences. *Philosophy and the Mind Sciences, 1*(I), 10.
https://doi.org/10.33735/phimisci.2020.I.32

nomologically impossible to experience one without the other. The (hypothesized) reason for this is that there are causal dependencies between the neural underpinnings of these phenomenal properties. Call this the *dependency approach*. Below, I will give the sense of time (i.e., experiencing the passage of time) and the sense of agency as a potential example: the hypothesis would be that the sense of time is causally necessary for the sense of agency, because of dependencies between the neural realizers of those phenomenal properties.

More generally, self-consciousness provides a useful testbed for the dependency approach. Self-consciousness is not a unitary, simple phenomenon. Instead, it is characterized by a variety of features, such as *de se* thought, body ownership, spatial self-location (Millière, 2020, this issue, p. 5), a pre-reflective self/non-self distinction, a sense of agency, patterns of affective states (Gallagher & Daly, 2018, p. 4), a sense of time, or a sense of center (Ataria, Dor-Ziderman, & Berkovich-Ohana, 2015, p. 142). None (or few) of these features are necessary for self-consciousness (in fact, it is even controversial whether all of them exist, and whether they are really distinct). Put differently, self-consciousness should be treated as a mongrel concept (just as the more general concept "consciousness" – see Block, 1995). Similar versions of this view have been expressed as *a pattern theory of self* (Gallagher, 2013; Gallagher & Daly, 2018; Newen, 2018), or as the claim that "self-consciousness may be best construed as a multidimensional construct" (Millière, Carhart-Harris, Roseman, Trautwein, & Berkovich-Ohana, 2018, p. 1). What I take to be central to this position is that particular types of self-consciousness are best characterized in terms of a variety of *features*.

Crucially, some of these features may be sufficient without being necessary for self-consciousness. This means that different types of self-consciousness can be instantiated (probably independently of each other). In addition, this enables the hypothesis that some combinations of features are nomologically impossible (although they are logically possible). Again, evidence regarding how "the self breaks" in radical disruptions of self-consciousness can suggest hypotheses about which combinations of phenomenal properties related to self-consciousness are possible and which are not. I will elaborate on this in section 5.

In sum, investigating impossible conscious experience is valuable, because it increases our understanding of consciousness (by narrowing down the concept of consciousness). But in addition to that, explanations of why certain conscious experiences are impossible may yield insights into the neural mechanisms underpinning consciousness.

3.1 Two types of impossible conscious experiences

Before moving on, it will be important to make a distinction between two types of impossible conscious experiences. One of them is rather trivial.[12] The other

[12] I am grateful to an anonymous reviewer for pressing me on this issue.

Wiese, W. (2020). Breaking the self: Radical disruptions of self-consciousness and impossible conscious experiences. *Philosophy and the Mind Sciences*, *1*(I), 10.
https://doi.org/10.33735/phimisci.2020.I.32

turns the claim that there are impossible conscious experiences into an interesting hypothesis. I will introduce this distinction by using an analogy.

Imagine you are painting with a brush and a small set of watercolors. Provided you have the three primary colors, cyan, yellow, and magenta, you can mix all other colors. If you only have cyan and magenta, you will only be able to mix a more restricted set of colors. For instance, painting in orange will be impossible for you, given your watercolor equipment. Still, painting in orange is logically possible, and even nomologically possible, if we relax the constraints on watercolors that are at your disposal. Human consciousness is similarly constrained. We have a certain set of sensory organs, which shape the way we experience the world. There are types of experience we will never have, for the simple reason that we lack the required sensory organs and do not have the neural machinery that would control and make sense of signals received through them. With a nod to Thomas Nagel (1974), one could point out that we will never have a bat's conscious experience. But this type of limitation is trivial, because it does not reflect the way our brains create consciousness. In order to overcome this type of limitation, we would have to augment our senses (or the way we use them, cf. Teng & Whitney, 2011); this would enable us to have conscious experiences that were impossible beforehand (just as adding a primary color extends the set of possible colors you can paint with). We might even be able to have experiences that are similar to experiences of bats (even if this might not show us what being a bat is like *for a bat*).

Now imagine you are painting on a dark, absorbent piece of wood. You have all three primary colors, but the consistencies of the colors differ: your yellow color is nice and pasty, but your magenta and cyan are too thin to leave a visible mark on the wooden board. The yellow color, by contrast, can readily be applied to the ground. Fortunately, you find out that you can still mix cyan and magenta with yellow, and the mixed color can be applied as nicely as the pure yellow color. This means you can paint in pure yellow, in green, orange, and any colors that are obtained by mixing all three primary colors. However, you will not be able to paint in pure cyan or pure magenta – because you will always have to mix those colors with a bit of yellow, in order to apply the color on your painting surface. In other words, some colors will be impossible for you, not because you lack the required primary colors, but because of how your painting equipment constrains the way in which you can paint.

If consciousness is constrained in this way (and this is the hypothesis explored in this paper), then there will be dependencies between phenomenal properties (just as blue and red "depend on yellow" in the analogy). These dependencies are contingent on the neural mechanisms underlying consciousness in human brains. Although there is a sense in which the first type of impossible conscious experiences (e.g., perceptual experiences based on sonar signals) is also contingent on the neural mechanisms underlying consciousness, there is a crucial difference, because the first type is not created by *dependencies*.

Wiese, W. (2020). Breaking the self: Radical disruptions of self-consciousness and impossible conscious experiences. *Philosophy and the Mind Sciences, 1*(I), 10.
https://doi.org/10.33735/phimisci.2020.I.32

Let me further illustrate the second type of impossible experiences with a toy example: assume every human conscious experience is characterized by at most two phenomenal properties, and that these properties are binary (either the phenomenal property is instantiated or not; there are no differences in the degree to which the property is expressed). Every conscious experience can then be described as an ordered pair, in which each element is either zero or one (zero means the respective phenomenal property is present, and one means it is absent). The set of *nomologically* possible experiences, given our neural machinery (which can, by assumption, only instantiate two different phenomenal properties), is then a set of four conscious experiences (or three, if at least one phenomenal property must be instantiated). The set of *logically* possible experiences is much greater, because there are further phenomenal properties. Experiences that require instantiating further phenomenal properties are impossible experiences of the first type.

Now assume that there is a dependency between the two phenomenal properties (which is contingent on the way our brains create consciousness): the first phenomenal property cannot be instantiated without the second. If this is the case, then the space of nomologically possible experiences, given our neural machinery, has a hole: an experience in which the first phenomenal property is present, while the second is absent, is impossible. This is an impossible experience of the second type. This type of impossible experience is non-trivial. For, knowing that a system can have experiences with certain phenomenal properties, it is not obvious that some combinations of these phenomenal properties are impossible.

3.2 Constraining models of consciousness

What can impossible experiences tell us about the neural mechanisms of consciousness? Note first that it is straightforward to see that the opposite is true: if we had a complete theory of consciousness (including detailed knowledge of the neural mechanisms of consciousness), we would be able to say which types of conscious experience are possible and which are not. But can impossible experiences, conversely, yield insights into the mechanisms of consciousness? I will argue that they can.

We can see this most clearly by considering a strong type of dependency between phenomenal properties. If there is a phenomenal property that is nomologically *necessary* for consciousness, then conscious experiences without this phenomenal property will be impossible in the sense that is relevant here. Furthermore, this will suggest that all other phenomenal properties depend on this property (at least causally).

Certain types of self-consciousness may provide controversial examples: if there is a basic sense of self that is constituted by a particular phenomenal property, viz. a *self quale* (or "ur quale", as it is called in Perlis, 1997, p. 509) that is necessary for consciousness, then conscious experience without this self quale would be impossible. If correct, this would suggest that there is a causal dependency between

Wiese, W. (2020). Breaking the self: Radical disruptions of self-consciousness and impossible conscious experiences. *Philosophy and the Mind Sciences, 1*(I), 10.
https://doi.org/10.33735/phimisci.2020.I.32

the neural underpinnings of this basic type of self-consciousness and the underpinnings of all other phenomenal properties. Hence, such a dependency between a single phenomenal property and all types of conscious experience is likely to reveal insights into the neural mechanisms of consciousness. A crucial step would be to find the neural correlate of the self quale, and then to relate it to candidates for total neural correlates of consciousness (that are involved whenever a person is conscious, but not when she is unconscious). However, as already noted, the example is controversial.

What is more, proponents of the claim that consciousness necessarily involves a basic sense of self usually explicitly reject the claim that this corresponds to a self quale (*contra* Perlis, 1997). For instance, Zahavi & Kriegel (2016) point out that "for-me-ness is not a detachable self quale that one could introspect in isolation from any other content of consciousness" (Zahavi & Kriegel, 2016, p. 39). This suggests that your visual experience of a red tomato and my visual experience of a red tomato never have any (maximally determinate) phenomenal properties in common. They may share the determinable property of seeing a red tomato, but always instantiate more determinate phenomenal properties that do not overlap.

Another way of interpreting this view is that there is a sense in which a basic sense of self is a (necessary) phenomenal property, but one that cannot be instantiated without instantiating at least one further phenomenal property. This would still be in line with Zahavi's and Kriegel's claim that one cannot introspect the self quale "in isolation from any other content of consciousness". Furthermore, it illustrates another interesting potential result of investigating candidates for impossible conscious experiences. It could turn out that what we described as two distinct phenomenal properties F and G are in fact constitutively dependent properties: G may actually be a composite phenomenal property constituted by F and F', i.e., $G \equiv F\&F'$. The fact that G depends on F would then be trivial (just as F trivially depends on itself).

In a recent article, Zahavi (2018) points out that the basic sense of self is not an "additional experiential object" (Zahavi, 2018, p. 5). However, in line with the interpretation that all phenomenal properties constitutively depend on a self quale, one could turn this statement on its head: all *other* contents of consciousness, such as a visually experienced red tomato, are not additional experiential objects, but always involve self-content.

The general idea of constraining models of consciousness by research on impossible experiences is as follows. Instead of starting with an assumption of how consciousness is generated, we start with an assumption about the impossibility of certain types of experience. The evidence for impossibility can be preliminary (and non-conclusive). But from the assumption *that* certain types of experience are impossible, one can derive constraints on their neural underpinnings. In the ideal case, this will allow one to derive testable predictions. If these predictions bear out, we gain further support for the hypothesis that certain types of conscious experience are impossible, and we learn (much) about the neural underpinnings of

Wiese, W. (2020). Breaking the self: Radical disruptions of self-consciousness and impossible conscious experiences. *Philosophy and the Mind Sciences, 1*(I), 10.
https://doi.org/10.33735/phimisci.2020.I.32

those types of conscious experience. If the predictions do not bear out, we should lower our confidence in the impossibility hypothesis (and we may not learn as much about the neural underpinnings of conscious experience).

3.3 Matching descriptions at different levels of analysis

There is another, more theoretical benefit that research on impossible conscious experiences may yield: it may increase our understanding of what phenomenal properties are in the first place, by providing support for structural theories of consciousness.

If we want to understand consciousness, we have to understand the characteristic features of consciousness. Features of consciousness are usually described on a personal level of analysis (as opposed to subpersonal levels – for this distinction, see Dennett, 1969; Drayson, 2012, 2014). The personal level of analysis has a variety of properties that make it difficult to establish links to subpersonal levels, on which functional, computational, or neurobiological features of conscious systems are described. For instance, phenomenological reports are subjective (as opposed to objective) and often make reference to phenomenal and intentional properties (as opposed to functional and physical properties). Still, one can learn a lot about consciousness without worrying about these most difficult aspects of consciousness: research on consciousness usually operates under the assumption of what David Chalmers (1995, 1996) calls "the principle of structural coherence". He defines it as follows:

> This is a principle of coherence between the *structure of consciousness* and the *structure of awareness*. [...] Briefly put, we can think of awareness as *direct availability for global control*. [...] In general, any information that is consciously experienced will also be cognitively represented. [...] This principle reflects the central fact that even though cognitive processes do not conceptually entail facts about conscious experience, consciousness and cognition do not float free of one another but cohere in an intimate way. (Chalmers, 1995, pp. 212–213)

The principle of structural coherence, if true, alleviates the problem of matching phenomenological descriptions of consciousness with descriptions at subpersonal levels of analysis. For instance, research on the neural correlates of consciousness can be conducted by finding systematic relationships between availability for verbal or non-verbal report and neural activity in different areas, and differences in reports (e.g., reporting a house vs. a face) can be matched with differences in neural activity (e.g., in the parahippocampal place area vs. the fusiform face area, see Tong, Nakayama, Vaughan, & Kanwisher, 1998).

But such research does not reveal *why* neural activity of a particular type is correlated with the presence of certain types of experience, and it does not by itself clarify "what it is like" (Farrell, 1950, p. 181; Nagel, 1974) to be in certain conscious states. Part of the answer may involve explaining why the experience

Wiese, W. (2020). Breaking the self: Radical disruptions of self-consciousness and impossible conscious experiences. *Philosophy and the Mind Sciences, 1*(I), 10.
https://doi.org/10.33735/phimisci.2020.I.32

has the phenomenal character it has, and not a different phenomenal character. In particular, this leads to the question whether "inverting"[13] phenomenal properties (an "inverted spectrum") is possible: not knowing *why* a given neural activity goes along with the instantiation of a specific phenomenal property such as experienced redness, the connection between neural activity and phenomenal properties can seem more or less arbitrary, and it may seem as if one could permute a subject's phenomenal properties without evoking any differences in behavioral or cognitive capacities.

One way to block such inversion scenarios is to point to asymmetries in the space of conscious experiences (see Clark, 1996). For instance, if subjects are able to discriminate a larger variety of hues within a certain part of the color spectrum than in other parts, this could be taken as evidence that the space of consciousness (or "quality space" – see Rosenthal, 2015) is asymmetrical, and that inversion scenarios are impossible.

Showing that certain types of conscious experience are impossible is another way of discovering asymmetries. If certain combinations of phenomenal properties are impossible, the space of nomologically possible conscious experiences will have holes, which may involve asymmetries. For instance, an asymmetry exists if a phenomenal property F can be instantiated without instantiating phenomenal property G, whereas G cannot be instantiated without instantiating F. Investigating impossible experiences may therefore contribute to developing structural theories of phenomenal properties (see also Wiese, 2018).

4 How can impossible conscious experiences be discovered?

4.1 The general idea

Here, I will pursue what I called the *dependency approach* in section 3 above. This approach gives us at least two ways in which impossible conscious experiences can be discovered. A first starts with what we know about the neural mechanisms underpinning consciousness. To the extent that this enables us to see how contents of consciousness are determined by these mechanisms, it will also tell us about the limits to the contents that can be consciously experienced. A second approach

[13] In "inversion scenarios", there are subjects with deviant phenomenal properties: for instance, whenever we would have an experience as of red, a subject with an inverted spectrum would have an experience as of green. If this inversion is systematic, it may not go along with any functional or behavioral differences, and it may be impossible to detect the presence of the inverted spectrum. Block & Fodor (1972) use such scenarios to argue that behavioral and functional criteria are insufficient to provide type-identity criteria for mental states – for if an inverted spectrum is possible, two subjects can have different types of conscious experience (because their phenomenal properties differ), without any functional or behavioral differences. For an overview of the debate, see Byrne (2018).

Wiese, W. (2020). Breaking the self: Radical disruptions of self-consciousness and impossible conscious experiences. *Philosophy and the Mind Sciences, 1*(I), 10.
https://doi.org/10.33735/phimisci.2020.I.32

starts from the other side: if a given type of conscious experience is impossible, what does this tell us about the neural mechanisms underpinnings consciousness? The dialectic is a bit more complex in this case: supposing that an experience e is impossible, one tries to derive hypotheses about its neural underpinnings n. If these hypotheses about n can be tested and are corroborated, this also supports the hypothesis that e is impossible. The goal of the second approach is thus not just finding out about the limits of (contents of) consciousness, but finding phenomenological and, say, neurobiological constraints on consciousness at the same time.

But why should we suppose that a particular conscious experience e is impossible in the first place? Here, I am suggesting that we start with the observation that there is a lack of counter-examples: no report of an experience of a given type is known to exist. In other words, there is no evidence to the effect that the type of experience in question is possible. The crucial step in the argument is, of course, to move from a lack of counter-examples to positive evidence for the impossibility of a type of conscious experience.

One can start with a model of how it could possibly be that a given type of experience is impossible. This can then enable one to determine whether it is actually true (i.e., moving from a how-possibly to a how-actually model – see Machamer, Darden, & Craver, 2000, p. 21). In terms of explanation, this would involve starting with a possible epistemic explanation and using it to develop an ontic explanation (Craver, 2014, p. 37), e.g., by making hypotheses about the causal mechanisms involved, and then testing these hypotheses.

How can this reveal that certain types of experience are impossible? On the one hand, it could turn out that the information processing involved in generating a given type of experience is too complex to be realized by a human brain. This could be due to cognitive limitations or to physical constraints on the speed of information processing.[14] On the other hand, and more fundamentally, there can be causal dependencies between the neural mechanisms that instantiate different properties of experience. In this case, I shall also say that the properties are causally dependent (or that there is a causal dependency between the properties). As already noted, I will focus on this second approach in what follows.

The challenge here is to find causal dependencies between properties. One could start by searching for correlations between properties. But a correlation between two properties F and G is compatible with different causal relations: F may be causally necessary for G (or vice versa), or there may be a third property H, the instantiation of which functions as a common cause, or there might be just a correlation, without any direct causal connection between F and G. Hence, it will be useful (if not even necessary) to make further assumptions about features of consciousness and their neural underpinnings.

Examples of such assumptions can be computational, functional, or neurobi-

[14] I am grateful to Philip Gerrans for suggesting thinking about constraints on the speed of processing.

Wiese, W. (2020). Breaking the self: Radical disruptions of self-consciousness and impossible conscious experiences. *Philosophy and the Mind Sciences, 1*(I), 10.
https://doi.org/10.33735/phimisci.2020.I.32

ological. Since the exact neural underpinnings of consciousness are still unclear, it would be of little use to make specific assumptions about neural mechanisms underlying consciousness (unless they allow us to derive empirical hypotheses that can actually be tested using current neuroscientific methods). In general, the challenge here is to strike a balance between highly informative, but too specific, assumptions and non-informative, unspecific assumptions. The problem with too specific assumptions is that they are likely to be false (which, unfortunately, does not mean they can easily be falsified). However, if assumptions are too unspecific (but also more likely), they do not put any strong constraints on mechanisms underlying consciousness. Background assumptions must be specific enough to enable an explanation as to *why* a given type of conscious experience is impossible, without making controversial claims that cannot be tested in practice.

Here, I will suggest assuming, first, that computations implemented by the brain are relevant to understanding consciousness (note that this does not mean that implementing certain computations is sufficient for consciousness – see Schweizer, 2019), and, secondly, that these computations are (among others) characterized by the following two features: the ideomotor principle and the principle of spatio-temporal hierarchies.[15]

4.2 Two general computational principles as background assumptions

According to the *ideomotor principle*, representations underpinning perception and action, respectively, overlap (at least partially). Some of the neural areas that are activated when I perceive someone move are therefore also activated when I conduct this movement myself. This is the core idea of the ideomotor principle, which was first formulated by Johann Friedrich Herbart (Herbart, 1825, 1834), and was later picked up by William James (for a historical overview of ideomotor ideas, see Stock & Stock, 2004).[16] Active inference provides a formal elaboration of this idea (and offers important extensions – see Wiese, 2017a; Friston et al., 2017a, 2017b).

Note that merely adopting the ideomotor principle does not require rejecting other approaches to motor learning and control. For instance, the principle is compatible with applications of optimal control theory to motor learning and control (see Herbort & Butz, 2012) – although researchers using optimal control theory do not typically incorporate the ideomotor principle in their models (cf. Wolpert, Ghahramani, & Flanagan, 2001; for critical discussion, see Friston, 2011).

[15] Incidentally, these are also key ideas of *active inference*, a framework developed by Karl Friston and colleagues (see Friston, Samothrakis, & Montague, 2012). However, the two computational principles can be accepted without presupposing that active inference provides an adequate way (let alone the only way) of modeling brain function. But I submit that the framework can still be useful to further develop ideas sketched in this paper.

[16] A further aspect of the principle is a particular view on how actions are learned – this will not be relevant to the discussion in this article.

Wiese, W. (2020). Breaking the self: Radical disruptions of self-consciousness and impossible conscious experiences. *Philosophy and the Mind Sciences, 1*(I), 10.
https://doi.org/10.33735/phimisci.2020.I.32

Another feature that will be relevant here is the *spatio-temporal hierarchy* of information processing in the brain. The assumption behind this is that environmental features reflected in sensory signals vary with respect to their temporal and spatial grains. Some features are invariant over large spatial regions in objective space (e.g., the color of the sky), and some are invariant over long temporal intervals in objective time (for instance, the size of my hands). An efficient way of representing mixtures of features of various temporal and spatial grains is by using hierarchical representations. Lower processing levels represent features that change quickly and are more concrete. Higher processing levels represent features that are invariant over longer periods of time or larger regions of space (they may also be more abstract). This also resonates with the idea that percepts consists of hierarchical representations, in which the same objects are represented more than once (i.e., in different ways; cf. Lycan, 1996; Bayne & McClelland, 2018).

So these are two computational features which, I will suggest, will be useful to find epistemic explanations as to why certain conscious experiences are impossible: the ideomotor principle and the principle of spatio-temporal hierarchies. Without getting into any formal details, we can already derive some constraints on how the brain represents the world. In particular, this also puts constraints on possible conscious experience, as we shall see now.

Ideomotor principle. Action is partially underpinned by representations that are also activated when perceiving or imagining a similar action. This means, for instance, that there are limits on what can easily be imagined and observed at the same time. There must also be interferences between action observation and execution (for a review of empirical evidence, see Zwickel & Prinz, 2012; for a recent supporting study see Lohmann, Belardinelli, & Butz, 2019).

Spatio-temporal hierarchy. If the contents of perception are hierarchical (which is suggested by hierarchical computational models), then temporal features are also represented at different timescales (Kiebel, Daunizeau, Friston, & Sporns, 2008). If such hierarchical models not only underpin unconscious perception, but also determine the contents of *conscious* perception, then the experienced present is hierarchical as well (Wiese, 2017b). This has subtle implications for how to understand the conscious perception of temporally extended events, such as a melody: at short time-scales, we do not first experience one note, and then another note, together with an experience of "diachronic togetherness". Rather, we experience a single, temporally extended event (the melody or succession of notes) as a whole, and at the same time experience changing properties that belong to the same temporal whole. That is, we do not experience these different properties (e.g., notes) as belonging to different temporal parts of the whole event. Simon Prosser puts it thus:

> When we experience change we do not [...] have experience as of an F temporal part of an object succeeded by a *non-F* temporal part of that object, with it somehow understood that both parts belong to the same composite whole. (Prosser, 2016, p. 173)

Wiese, W. (2020). Breaking the self: Radical disruptions of self-consciousness and impossible conscious experiences. *Philosophy and the Mind Sciences, 1*(I), 10.
https://doi.org/10.33735/phimisci.2020.I.32

This may sound puzzling, because it seems to require that we somehow experience an object as having temporal parts and as not having temporal parts. But what Prosser emphasizes here is that we do not experience persisting objects as *perduring*, but as *enduring*. If they were experienced as perduring, they would be experienced as having temporal parts. Instead, we experience them as enduring, as being wholly present at any instant; and changing objects are, in addition, experienced as having properties that are only transiently instantiated.

Hierarchical models of temporal perception suggest a way to make this idea more precise: when two successive notes of a single melody are perceived, both are experienced as part of a single extended event, which is partly underpinned by a representation at a higher level (which represents properties at a more coarse-grained timescale).[17] At that level, properties are represented as being instantiated *now*, but also as just having been instantiated – they are represented as *continuing* to be instantiated. This means that those properties cannot be represented as being instantiated *now* without thereby also being represented as being extended into the recent past and the near future (in other words, the objects having those properties are represented as enduring). Properties that are transiently instantiated (such as the pitch of a single note) are represented at lower levels, and can specify in what way enduring objects change. See figure 1 for an illustration. However, the point that is crucial for the purposes of this article is that higher levels encode objects that are temporally extended.

I submit that these two computational principles, i.e., the ideomotor principle and the principle of spatio-temporal hierarchies, are general enough so as to function as relatively uncontroversial background assumptions about neural mechanisms underlying consciousness. In other words, I only assume that neural activity underpinning consciousness is consistent with the way of representing the world entailed by these two features.

At the same time, we have seen that these two features already constrain the way in which the brain represents the world. For this reason, they at least promise to suggest explanations of *why* certain conscious experiences may be impossible. In every particular case, we will need further, *ad hoc* assumptions, but general background assumptions, such as the ones specified here, will serve as guiding principles.

Such assumptions and guiding principles specify at most *epistemic* explanations, in that they make it likely that certain conscious experiences are impossible. In other words, neither do they logically entail the impossibility of certain types of conscious experience, nor do they specify neural (causal) mechanisms that render certain experiences impossible. But they do specify mechanism *sketches* (cf. Machamer et al., 2000, p. 18; Piccinini & Craver, 2011), and thereby also suggest *ontic* explanations as to why a given candidate for an impossible experience cannot be created by human brains. An ontic explanation of a phenomenon describes causal interactions between parts of a system, in a way that shows how the

[17] In Wiese (2017b) I call this a *dynamic event representation* (see Wiese, 2017b, p. 14).

Wiese, W. (2020). Breaking the self: Radical disruptions of self-consciousness and impossible conscious experiences. *Philosophy and the Mind Sciences, 1*(I), 10.
https://doi.org/10.33735/phimisci.2020.I.32

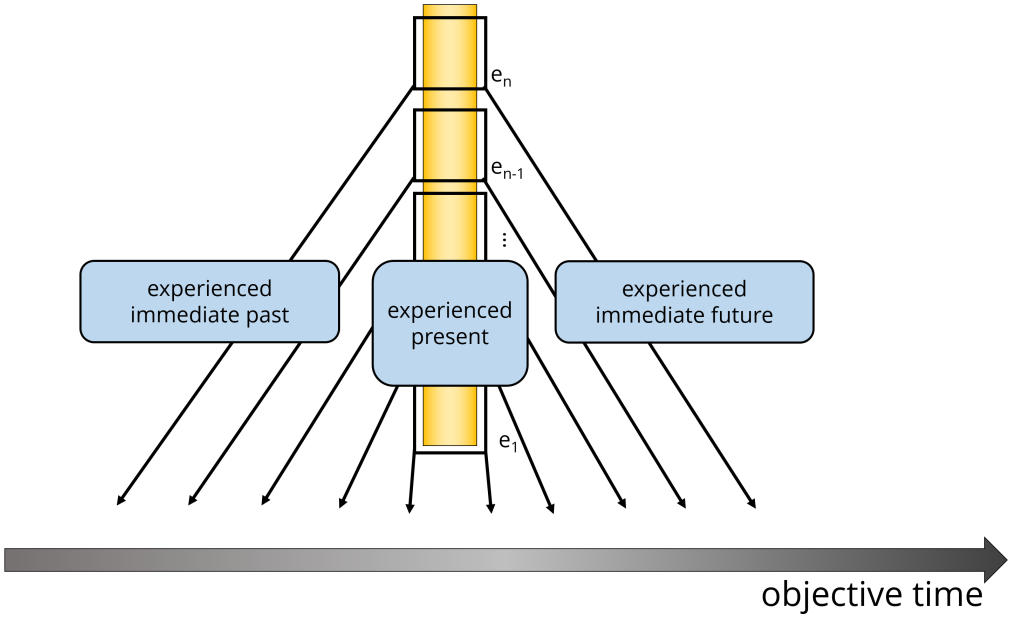

Figure 1: The yellow bar illustrates a dynamic event representation. It is a distributed representation of multiple properties (individual properties are represented by $e_1, \dots, e_{n-1}, e_n$) that jointly specify a temporally extended event: near the top, properties that change slowly and are instantiated over a longer period of time (e.g., hundreds of milliseconds) are represented; near the bottom, properties that change quickly and are instantiated in the immediate present only (e.g., for less than one hundred milliseconds) are represented. The represented period of time during which the different properties are instantiated is illustrated with black arrows (marking the beginning and the end of each represented period during which a property is instantiated). Together, $e_1, \dots, e_{n-1}, e_n$ represent a single, enduring event that we experience "as rising from the recent past, and as continuing into the near future" (Wiese, 2017b, p. 13), because some of its properties remain invariant over a period of time that is slightly longer than the short interval we experience as the present. Together, they represent a *specious* present (James, 1890). (Figure adapted from Wiese, 2017b, p. 12.)

activities of the parts and their organization give rise to the target phenomenon (see Craver & Tabery, 2017, sec. 2). The idea is thus that we can move from descriptions of computational features to descriptions of mechanisms, and this will (hopefully) enable us to derive testable predictions.

5 From radical disruptions of self-consciousness to impossible conscious experiences

5.1 General idea

Recall that I suggest to start with the observation that no positive evidence for the existence of a given type of conscious experience exists. With respect to self-

Wiese, W. (2020). Breaking the self: Radical disruptions of self-consciousness and impossible conscious experiences. *Philosophy and the Mind Sciences, 1*(I), 10.
https://doi.org/10.33735/phimisci.2020.I.32

consciousness, this means that there is a phenomenal configuration (a combination of features associated with different types of self-consciousness), that is not reported in the literature (or in databases such as https://erowid.org). Given the many variations of phenomenal configurations reported in radical disruptions of self-consciousness, there is thus a *prima facie* reason to believe that the lack of evidence for the phenomenal configuration in question is not a mere coincidence.

By itself, this reason is too weak to justify the claim that the phenomenal configuration at hand is indeed nomologically impossible. This is where the various approaches outlined in sections 3 and 4 come into play. As already pointed out, I will pursue the *dependency approach*, which focuses on causal dependencies between the neural underpinnings of phenomenal properties. When two features are *mutually* causally dependent then one cannot be instantiated without the other. When only one feature depends on the other (A is causally necessary for B, but not vice versa), the first cannot be instantiated without the second, but the second can be instantiated without the first. Another type of causal dependency involves a third feature C, such that both A and B are causally dependent on C (but C need not be dependent on A and B). In what follows, I will consider "the sense of boundaries" (Ataria, 2014), which is constituted by different phenomenal properties. Two of them are the sense of agency and the sense of time, and I will suggest that the sense of time is necessary for the sense of agency.

5.2 The sense of boundaries

We usually experience a boundary between ourselves and the world. I experience the world from a first-person perspective: *I* am at the center of my experience, I am on the inside, within the boundaries of my body, and the rest of the world is outside. Yochai Ataria (2014) calls this the *sense of boundaries*. The sense of boundaries corresponds to a bundle of phenomenal properties. It is particularly interesting because it is likely that the phenomenal properties constituting the sense of boundaries are not completely independent: at least some of them cannot be instantiated without also instantiating some other phenomenal properties. Put differently, there are configurations of phenomenal properties (associated with the sense of boundaries) that correspond to (nomologically) impossible experiences – or at least this is what existing research on the sense of boundaries suggests.

But what exactly is the "sense of boundaries" (SB)? First of all, the SB is not identical to a physical boundary. Instead, it is an experienced boundary between one's lived[18] body and the rest of the world. Crucially, since the lived body depends on sensations in different sensory modalities, it can be modulated by changing sensory input (e.g., by closing one's eyes or by entering a hot bath of water[19]), or by

[18]The *lived body* is a term used by phenomenologists to refer to the body, experienced as *your own body*, not experienced as an object.

[19]In some cases this can even affect the experienced boundary between the lived body and the environment. Millière (2019) cites a written interview, in which a deafblind individual describes

Wiese, W. (2020). Breaking the self: Radical disruptions of self-consciousness and impossible conscious experiences. *Philosophy and the Mind Sciences*, *1*(I), 10.
https://doi.org/10.33735/phimisci.2020.I.32

focusing attention on different aspects of one's current experience (cf. Ataria, 2014). Moreover, Ataria (2014) distinguishes between the experience of boundaries and experiencing these boundaries as having determinate spatial locations (cf. Ataria, 2014, p. 1142; Martin, 1995, p. 271).

Since boundaries between oneself and the world can be experienced in different ways, it is best to define the SB as a bundle of phenomenal properties. Ataria et al. (2015) characterize the SB in terms of several phenomenal properties (which they call "categories"), including the sense of agency and the sense of time (see Ataria et al., 2015, pp. 143–144).

As already noted, I shall focus on these two phenomenal properties here. By the "sense of time", one could mean different phenomena, for instance, the experience of temporal properties such as duration or succession, consciously remembering past events or imagining future events, a conscious estimate of how long it has taken to read up to this sentence, or having a sense of which of two temporally extended events was longer.[20] Ataria et al. (2015) do not give a definition of the sense of time. However, quoted reports in which their subject describes how their sense of time changes indicate that the sense of time is regarded as a gradual property and that it can also fail to be instantiated.[21]

The absence of the sense of time is described as "timelessness" in an earlier publication, to which the authors refer in a footnote (see Berkovich-Ohana, Dor-Ziderman, Glicksohn, & Goldstein, 2013). In a more recent publication, the sense of time is characterized as follows:

> Specifically, we are referring to a sense of a past continuing into the future; to the reduction in the sense of duration; and finally, to the sense of continuity itself which also disintegrates. (Dor-Ziderman, Ataria, Fulder, Goldstein, & Berkovich-Ohana, 2016, p. 3)

I submit that the relevant notion of "sense of time" is that of the passage of time on short timescales. That is, it is not primarily about having a sense of the duration of past events, or of events on long timescales. Instead, it is about experiencing temporally extended events that are occurring now, have a duration, and reach into the very recent past and the immediate future (including internal events, such as the ongoing endeavor to regulate attention). Furthermore, the sense of time does not necessarily entail a sense of self, i.e., temporal experience without any form of self-experience may be possible (see Windt, 2015, pp. 18–19).

---

the experience of a loss of spatial boundaries when they take a hot bath (see Millière, 2019, p. 271).

[20] I am grateful to an anonymous reviewer for asking me to be more specific about the sense of time.

[21] "Sense of time is obviously much clearer … yet I can't describe it really" (Ataria et al., 2015, p. 139)

"I would need to stop meditating, totally open my eyes and be in totally ordinary consciousness, only then would the sense of time come back fully" (Ataria et al., 2015, p. 140)

"time is not relevant anymore" (Ataria et al., 2015, p. 142)

Wiese, W. (2020). Breaking the self: Radical disruptions of self-consciousness and impossible conscious experiences. *Philosophy and the Mind Sciences, 1*(I), 10.
https://doi.org/10.33735/phimisci.2020.I.32

Furthermore, Ataria et al. (2015) suggest that there is a causal dependency between the sense of boundaries and the sense of time:

> It seems that the sense of time is a kind of 'mirror reflection' of the SB. Thus any alteration in the level of flexibility is reflected by a proportional adjustment in the sense of time. If we examine this phenomenon more broadly, we may say that the sense of self and the sense of time go hand in hand (Ataria et al., 2015, p. 143)

Unpacking this statement requires a bit reflection, because the sense of time, as characterized by Ataria et al. (2015), is not an additional feature, but actually one of the features that constitute the SB. So saying that the SB changes whenever the sense of time changes would be trivial (just as saying that the state of my computer screen changes when one of the displayed pixels changes). However, the statement does suggest more substantive implications. First, it suggests that any change in the sense of time (i.e., its becoming weaker or stronger) is reflected in corresponding changes in all other features constituting the SB. Secondly, it suggests that any change in one of the other features also goes along with a corresponding change in the sense of time. Perhaps the strongest interpretation would be that all of the features constituting the SB are positively correlated, such that the degree to which one of the features is expressed is predictive of how strongly (or weakly) the other features are expressed.

For the purpose of illustration, let us consider the much weaker hypothesis that the sense of time is *necessary* for the *sense of agency* (which is one of the other properties constituting the SB). In a narrow sense, the sense of agency is the experience of causing a bodily or mental action (Gallagher, 2000, p. 15). In addition to that, there is a more general notion that includes the experienced *ability* to act (cf. Metzinger, 2013, p. 2). Let us call this the *sense of ability* (cf. Slaby, 2012). In what follows, "sense of agency" will refer to the general concept, which subsumes the sense of ability. I will also write "minimal sense of agency" for the sense of ability.

Is it possible to have a sense of agency without a sense of time? If I experience myself as causing or generating an event that is occurring now, I must, by logical necessity, experience that causing as an event that is unfolding now. So a strong sense of agency (involving the experience of causing an action) seems to require a sense of time.[22] This is compatible with the hypothesis that "timeless" experiences

[22] What if it is possible to experience an action that is not unfolding at any particular point in time, but is an "eternal" event? In fact, this might actually be the case in some drug-induced states in which the sense of time is thoroughly disrupted. Many such states may involve "thought loops" in which the same thought consistently recurs with no sense of when it begins and when it ends. A person may have the experience of causing a particular (mental) action, but without having the experience of causing it at any particular point in time, because the action does not have a determinate temporal location. Here, it would be relevant to determine whether an "eternal" event is best conceived as an event that has always existed and will always exist (in that case, it would have temporal properties) – or as an event that simply lacks any temporal properties (in

Wiese, W. (2020). Breaking the self: Radical disruptions of self-consciousness and impossible conscious experiences. *Philosophy and the Mind Sciences*, *1*(I), 10.
https://doi.org/10.33735/phimisci.2020.I.32

exist (in which there is no conscious experience of the passage of time at all). The idea is merely that a sense of time must be there whenever there is a sense of agency.

What about the sense of ability? In its minimal form, the sense of agency only requires an awareness of *possibilities* for action. So the hypothesis derived from this example is that being aware of possibilities for action requires a sense of time. And this is equivalent to the hypothesis that a conscious experience involving a (minimal) sense of agency, but no sense of time, is impossible.

If I merely experience the ability to initiate an action, I do not have to experience the action as unfolding. Still, it seems that I have to experience this possibility as referring to a *future* event, viz. a potential action that can be executed in the future. However, when I experience myself as having the ability to act, potential actions may be represented indeterminately,[23] and they may not explicitly be experienced as *future* actions, but simply as *something* that can be done (or as something *I* can do) – just as I can think about potentially existing objects, without imagining that they exist at any particular point in time. But is it *nomologically* possible to experience myself as an agent, without experiencing the passage of time, given the background assumptions outlined in section 4 above?

According to the ideomotor principle, there is an overlap between the representations involved in early action initiation (planning) and the representations involved in perceiving actual actions.[24] Assuming that the phenomenal representations contributing to the sense of ability include such "ideomotor" representations, we get a heuristic (epistemic) explanation of why one cannot experience oneself as being the potential author of actions, without having a sense of time.

The explanation runs as follows. We are assuming that the sense of ability is underpinned by representations that can be used to initiate actions and to represent actions that are actually unfolding. As such, they must be relatively high in the processing hierarchy. This hierarchy not only reflects differences in levels of abstraction, or differences in spatial grain, but also differences in temporal grain: at higher levels, features that change slowly are represented, which underpins experiences of *endurance*. Those features are relatively abstract and can be common to many different actions. This may explain why the sense of ability is not necessarily specific: I can experience the ability to act, without experiencing the ability to do anything in particular. But if I experience myself as the potential author of an action by experiencing myself as instantiating such an enduring feature, I thereby experience the passage of time, at least in a minimal sense, because the feature is experienced as being instantiated now, in the near future, and in the recent past.

that case, a sense of agency without a sense of time would be possible. I am grateful to the guest editors of this special issue for suggesting this possibility.

[23] Such representations may also underpin existential feelings – see Ratcliffe (2012).

[24] As Bernhard Hommel, one of the proponents of the ideomotor principle puts it, the theory "is meant to provide a framework for understanding linkages between (late) perception and (early) action, or action planning." (Hommel, Müsseler, Aschersleben, & Prinz, 2001, p. 849).

Wiese, W. (2020). Breaking the self: Radical disruptions of self-consciousness and impossible conscious experiences. *Philosophy and the Mind Sciences, 1*(I), 10.
https://doi.org/10.33735/phimisci.2020.I.32

Put more simply, the idea is that the sense of ability inherits experienced temporal properties from representations of potential action. And since potential actions cannot be represented without reference to temporal properties (this follows from the two computational principles), the sense of ability must also involve a sense of time.

This explanation is relatively ad hoc. In line with the dependency approach, empirical evidence must be sought for a causal dependence between the neural realizers of the representations referred to above. This would require making more specific hypotheses about the underlying mechanisms. Since the dependency between the sense of agency and the sense of time mainly serves as an example here, further analyses remain a task for future work. But notice that the dependency approach may yield surprising results. So far, I have talked as if one could always map phenomenal properties to distinct parts of a mechanism, and establish whether there is a causal dependency. However, it may also turn out that what has been described as a single phenomenal property (such as the sense of ability) actually involves different parts of a mechanism (which only sometimes interact and bring about a sense of ability). For instance, perhaps the sense of ability is underpinned by a structure that underpins the subject component and a structure that underpins the "object component" (potential action). The sense of time could be underpinned by a third structure, or it could be intertwined with the first two. Using background assumptions as a framework, some of these possibilities may be more likely than others. And in the end, the dependency between phenomenal properties may even turn out to be more intimate than a causal dependency between different neural events. That is, it could turn out that there is actually a constitutive dependency between different phenomenal properties.

Conclusion

The general hypothesis of this paper is that some phenomenal configurations (combinations of phenomenal properties) are logically possible but nomologically impossible, given further assumptions about how the brain generates conscious experiences. I have argued that research on impossible conscious experiences may provide insights about neural mechanisms underpinning consciousness and is not just interesting for the sake of marking out the terrain of conscious experiences that can be generated by the human brain.

I have described three approaches to finding impossible conscious experiences: the ability approach (which starts from limits on cognitive abilities), the intensity approach (which starts from considerations about *extreme* conscious experiences, involving phenomenal properties that vary on a spectrum, but may not exceed certain thresholds), and the dependency approach (which focuses on causal dependencies between the neural underpinnings of different phenomenal properties).

Radical disruptions of self-consciousness provide a useful test bed for at least the dependency approach, since many types of self-consciousness can be

Wiese, W. (2020). Breaking the self: Radical disruptions of self-consciousness and impossible conscious experiences. *Philosophy and the Mind Sciences, 1*(I), 10.
https://doi.org/10.33735/phimisci.2020.I.32

described in terms of combinations of phenomenal properties (such as the sense of agency or a sense of time). And if even in altered states (or radical disruptions) of self-consciousness certain configurations of phenomenal properties are not experienced, this warrants further explorations of the hypothesis that those configurations are indeed impossible. This does not presuppose that conscious experiences lacking any kind of consciousness of oneself are impossible. Rather, it assumes that there are different types of self-consciousness, and that at least some of them are dependent on other phenomenal properties. I have illustrated this with the hypothesis that the sense of agency requires a sense of time (without there being a logical implication between the two). In addition to investigating the neural underpinnings of actual conscious experiences, thinking about impossible conscious experiences may sharpen the conceptual tools used to describe conscious experiences and lead to deeper explorations of their underlying mechanisms.

Acknowledgments
I am grateful to two anonymous reviewers for constructive criticism and to Thomas Metzinger and Raphaël Millière for their detailed comments on the penultimate draft of this paper. I would like to thank Jennifer Windt for comments on the accepted version of the manuscript, and the reviewers and participants of the Selfless Minds workshop at the Frankfurt Institute for Advanced Studies in 2018 for comments and constructive criticism. I thank Robin Wilson for editing the manuscript.

References

Ataria, Y. (2014). Where do we end and where does the world begin? The case of insight meditation. *Philosophical Psychology*, *28*(8), 1128–1146. https://doi.org/10.1080/09515089.2014.969801

Ataria, Y., Dor-Ziderman, Y., & Berkovich-Ohana, A. (2015). How does it feel to lack a sense of boundaries? A case study of a long-term mindfulness meditator. *Consciousness and Cognition*, *37*, 133–147. https://doi.org/10.1016/j.concog.2015.09.002

Bayne, T. (2014). Replies to commentators. *Analysis*, *74*(3), 520–529. https://doi.org/10.1093/analys/anu066

Bayne, T., & McClelland, T. (2018). Ensemble representation and the contents of visual experience. *Philosophical Studies*, *176*(3), 733–753. https://doi.org/10.1007/s11098-018-1037-3

Berkovich-Ohana, A., Dor-Ziderman, Y., Glicksohn, J., & Goldstein, A. (2013). Alterations in the sense of time, space, and body in the mindfulness-trained brain: A neurophenomenologically-guided MEG study. *Frontiers in Psychology*, *4*, 912. https://doi.org/10.3389/fpsyg.2013.00912

Block, N. (1995). On a confusion about a function of consciousness. *Behavioral and Brain Sciences*, *18*(2), 227–247. https://doi.org/10.1017/S0140525X00038188

Block, N. J., & Fodor, J. A. (1972). What psychological states are not. *The Philosophical Review*, *81*(2), 159–181. https://doi.org/10.2307/2183991

Byrne, A. (2018). Inverted qualia. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Winter 2018). https://plato.stanford.edu/archives/win2018/entries/qualia-inverted/; Metaphysics Research Lab, Stanford University.

Chalmers, D. J. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies*, *2*(3), 200–219. Retrieved from https://www.ingentaconnect.com/contentone/imp/jcs/1995/00000002/00000003/653

Chalmers, D. J. (1996). *The conscious mind: In search of a fundamental theory.* Oxford: Oxford University Press.

Chalmers, D. J. (2004). The representational character of experience. In B. Leiter (Ed.), *The future for philosophy* (pp. 153–181). Oxford: Oxford University Press.

Wiese, W. (2020). Breaking the self: Radical disruptions of self-consciousness and impossible conscious experiences. *Philosophy and the Mind Sciences*, *1*(I), 10. https://doi.org/10.33735/phimisci.2020.I.32

Chudnoff, E. (2013). Gurwitsch's phenomenal holism. *Phenomenology and the Cognitive Sciences*, *12*(3), 559–578. https://doi.org/10.1007/s11097-012-9279-6

Clark, A. (1996). *Sensory qualities*. Oxford: Oxford University Press.

Cowan, N. (2010). The magical mystery four: How is working memory capacity limited, and why? *Current Directions in Psychological Science*, *19*(1), 51–57. https://doi.org/10.1177/0963721409359277

Crane, T. (2009). Intentionalism. In B. McLaughlin, A. Beckermann, & S. Walter (Eds.), *The Oxford handbook of philosophy of mind* (pp. 479–493). Oxford: Oxford University Press.

Craver, C. F. (2014). The ontic account of scientific explanation. In M. I. Kaiser, O. R. Scholz, D. Plenge, & A. Hüttemann (Eds.), *Explanation in the special sciences: The case of biology and history.* (pp. 27–52). https://doi.org/10.1007/978-94-007-7563-3_2

Craver, C., & Tabery, J. (2017). Mechanisms in science. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Spring 2017). Retrieved from https://plato.stanford.edu/archives/spr2017/entries/science-mechanisms/

Dennett, D. C. (1969). *Content and consciousness*. New York; London: Routledge.

Dor-Ziderman, Y., Ataria, Y., Fulder, S., Goldstein, A., & Berkovich-Ohana, A. (2016). Self-specific processing in the meditating brain: A MEG neurophenomenology study. *Neuroscience of Consciousness*, *2016*(1), niw019. https://doi.org/10.1093/nc/niw019

Drayson, Z. (2012). The uses and abuses of the personal/subpersonal distinction. *Philosophical Perspectives*, *26*, 1–18. https://doi.org/10.1111/phpe.12014

Drayson, Z. (2014). The personal/subpersonal distinction. *Philosophy Compass*, *9*(5), 338–346. https://doi.org/10.1111/phc3.12124

Farrell, B. A. (1950). Experience. *Mind*, *59*(234), 170–198. https://doi.org/10.1093/mind/LIX.234.170

Friston, K. (2011). What is optimal about motor control? *Neuron*, *72*(3), 488–498. https://doi.org/10.1016/j.neuron.2011.10.018

Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., & Pezzulo, G. (2017a). Active inference: A process theory. *Neural Computation*, *29*(1), 1–49. https://doi.org/10.1162/NECO_a_00912

Friston, K. J., Parr, T., & Vries, B. de. (2017b). The graphical brain: Belief propagation and active inference. *Network Neuroscience*, *1*(4), 381–414. https://doi.org/10.1162/NETN_a_00018

Friston, K., Samothrakis, S., & Montague, R. (2012). Active inference and agency: Optimal control without cost functions. *Biological Cybernetics*, *106*(8), 523–541. https://doi.org/10.1007/s00422-012-0512-8

Gallagher, S. (2000). Philosophical conceptions of the self: Implications for cognitive science. *Trends in Cognitive Sciences*, *4*(1), 14–21. https://doi.org/10.1016/s1364-6613(99)01417-5

Gallagher, S. (2013). A pattern theory of self. *Frontiers in Human Neuroscience*, *7*, 443. https://doi.org/10.3389/fnhum.2013.00443

Gallagher, S., & Daly, A. (2018). Dynamical relations in the self-pattern. *Frontiers in Psychology*, *9*(664). https://doi.org/10.3389/fpsyg.2018.00664

Herbart, J. F. (1825). *Psychologie als Wissenschaft neu gegründet auf Erfahrung, Metaphysik und Mathematik. Zweiter, analytischer Teil.* Königsberg: Unzer.

Herbart, J. F. (1834). *Lehrbuch zur Psychologie* (2nd ed.). Königsberg: Unzer.

Herbort, O., & Butz, M. V. (2012). Too good to be true? Ideomotor theory from a computational perspective. *Frontiers in Psychology*, *3*, 494. https://doi.org/10.3389/fpsyg.2012.00494

Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). The theory of event coding (TEC): A framework for perception and action planning. *Behavioral and Brain Sciences*, *24*(05), 849–878. https://doi.org/10.1017/S0140525X01000103

James, W. (1890). *The principles of psychology*. New York: Henry Holt.

Kiebel, S. J., Daunizeau, J., Friston, K. J., & Sporns, O. (2008). A hierarchy of time-scales and the brain. *PLoS Computational Biology*, *4*(11), e1000209. https://doi.org/10.1371/journal.pcbi.1000209

Koksvik, O. (2014). Three models of phenomenal unity. *Journal of Consciousness Studies*, *21*(7–8), 105–131. Retrieved from https://www.ingentaconnect.com/content/imp/jcs/2014/00000021/f0020007

Kriegel, U. (2015). *The varieties of consciousness*. Oxford: Oxford University Press.

Lamme, V. (2015). The crack of dawn. In T. K. Metzinger & J. M. Windt (Eds.), *Open MIND*. https://doi.org/10.15502/9783958570092

Letheby, C. (2020). Being for no-one: Psychedelic experience and minimal subjectivity. *Philosophy and the Mind Sciences*, *1*(I), 5. https://doi.org/10.33735/phimisci.2020.I.47

Wiese, W. (2020). Breaking the self: Radical disruptions of self-consciousness and impossible conscious experiences. *Philosophy and the Mind Sciences*, *1*(I), 10. https://doi.org/10.33735/phimisci.2020.I.32

Lohmann, J., Belardinelli, A., & Butz, M. V. (2019). Hands ahead in mind and motion: Active inference in peripersonal hand space. *Vision*, *3*(2), 15. https://doi.org/10.3390/vision3020015

Lycan, W. G. (1996). Layered perceptual representation. *Philosophical Issues*, *7*, 81–100. https://doi.org/10.2307/1522894

Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of Science*, *67*(1), 1–25. Retrieved from https://www.jstor.org/stable/pdf/188611.pdf

Martin, M. G. (1995). Bodily awareness: A sense of ownership. In J. L. Bermúdez, A. Marcel, & N. Eilan (Eds.), *The body and the self* (pp. 267–289). Cambridge, MA: MIT Press.

Metzinger, T. (2013). The myth of cognitive agency: Subpersonal thinking as a cyclically recurring loss of mental autonomy. *Frontiers in Psychology*, *4*, 931. https://doi.org/10.3389/fpsyg.2013.00931

Metzinger, T. (2020). Minimal phenomenal experience: Meditation, tonic alertness, and the phenomenology of "pure" consciousness. *Philosophy and the Mind Sciences*, *1*(I), 7. https://doi.org/10.33735/phimisci.2020.I.46

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81–97.

Millière, R. (2019). Are there degrees of self-consciousness? *Journal of Consciousness Studies*, *26*(3-4), 252–282. Retrieved from https://www.ingentaconnect.com/contentone/imp/jcs/2019/00000026/f0020003/art00012

Millière, R. (2020). The varieties of selflessness. *Philosophy and the Mind Sciences*, *1*(I), 8. https://doi.org/10.33735/phimisci.2020.I.48

Millière, R., Carhart-Harris, R. L., Roseman, L., Trautwein, F.-M., & Berkovich-Ohana, A. (2018). Psychedelics, meditation, and self-consciousness. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.01475

Nagel, T. (1974). What is it like to be a bat? *The Philosophical Review*, *83*(4), 435–450. https://doi.org/10.2307/2183914

Newen, A. (2018). The embodied self, the pattern theory of self, and the predictive mind. *Frontiers in Psychology*, *9*, 2270. https://doi.org/10.3389/fpsyg.2018.02270

Perlis, D. (1997). Consciousness as self-function. *Journal of Consciousness Studies*, *4*(5), 509–525.

Piccinini, G., & Craver, C. (2011). Integrating psychology and neuroscience: Functional analyses as mechanism sketches. *Synthese*, *183*(3), 283–311. https://doi.org/10.1007/s11229-011-9898-4

Prosser, S. (2016). *Experiencing time*. Oxford: Oxford University Press.

Ratcliffe, M. (2012). The phenomenology of existential feeling. In J. Fingerhut (Ed.), *Feelings of being alive* (pp. 23–53). Berlin: de Gruyter.

Revonsuo, A. (1998). How to take consciousness seriously in cognitive neuroscience. *Communication and Cognition: An Interdisciplinary Quarterly Journal*, *30*(3-4), 185–205.

Rosenthal, D. M. (2015). Quality spaces and sensory modalities. In P. Coates & S. Coelman (Eds.), *Phenomenal qualities: Sense, perception, and consciousness* (pp. 33–65). Oxford: Oxford University Press.

Schweizer, P. (2019). Triviality arguments reconsidered. *Minds and Machines*, *29*, 287–308. https://doi.org/10.1007/s11023-019-09501-x

Searle, J. R. (1979). What is an intentional state? *Mind*, *88*(349), 74–92. Retrieved from https://www.jstor.org/stable/pdf/2253065.pdf

Searle, J. R. (1983). *Intentionality: An essay in the philosophy of mind*. Cambridge: Cambridge University Press.

Sebastián, M. Á. (2020). Perspectival self-consciousness and ego-dissolution: An analysis of (some) altered states of consciousness. *Philosophy and the Mind Sciences*, *1*(I), 9. https://doi.org/10.33735/phimisci.2020.I.44

Slaby, J. (2012). Affective self-construal and the sense of ability. *Emotion Review*, *4*(2), 151–156. https://doi.org/10.1177/1754073911430136

Stock, A., & Stock, C. (2004). A short history of ideo-motor action. *Psychological Research*, *68*, 176–188. https://doi.org/10.1007/s00426-003-0154-5

Teng, S., & Whitney, D. (2011). The acuity of echolocation: Spatial resolution in sighted persons compared to the performance of an expert who is blind. *Journal of Visual Impairment & Blindness*, *105*(1), 20–32. https://doi.org/10.1177/0145482X1110500103

Tong, F., Nakayama, K., Vaughan, J. T., & Kanwisher, N. (1998). Binocular rivalry and visual awareness in human extrastriate cortex. *Neuron*, *21*, 753–759. https://doi.org/10.1016/s0896-6273(00)80592-9

Wiese, W. (2017a). Action is enabled by systematic misrepresentations. *Erkenntnis*, *82*(6), 1233–1252. https://doi.org/10.1007/s10670-016-9867-x

Wiese, W. (2017b). Predictive processing and the phenomenology of time consciousness. A hierarchical extension of Rick Grush's trajectory estimation model. In T. Metzinger & W. Wiese (Eds.), *Philosophy and predictive processing*. Frankfurt am Main: MIND Group.

Wiese, W. (2020). Breaking the self: Radical disruptions of self-consciousness and impossible conscious experiences. *Philosophy and the Mind Sciences*, *1*(I), 10. https://doi.org/10.33735/phimisci.2020.I.32

Wiese, W. (2018). Toward a mature science of consciousness. *Frontiers in Psychology*, *9*, 693. https://doi.org/10.3389/fpsyg.2018.00693

Windt, J. M. (2015). Just in time. Dreamless sleep experience as pure subjective temporality. In T. K. Metzinger & J. M. Windt (Eds.), *Open MIND*. Frankfurt am Main: MIND Group.

Wolpert, D. M., Ghahramani, Z., & Flanagan, J. R. (2001). Perspectives and problems in motor learning. *Trends in Cognitive Sciences*, *5*, 487–494.

Zahavi, D. (2018). Consciousness, self-consciousness, selfhood: A reply to some critics. *Review of Philosophy and Psychology*. https://doi.org/10.1007/s13164-018-0403-6

Zahavi, D., & Kriegel, U. (2016). For-me-ness: What it is and what it is not. In D. O. Dahlstrom, A. Elpidorou, & W. Hopp (Eds.), *Philosophy of mind and phenomenology: Conceptual and empirical approaches* (pp. 36–53). New York: Routledge.

Zwickel, J., & Prinz, W. (2012). Assimilation and contrast: The two sides of specific interference between action and perception. *Psychological Research*, *76*, 171–182. https://doi.org/10.1007/s00426-011-0338-3

Wiese, W. (2020). Breaking the self: Radical disruptions of self-consciousness and impossible conscious experiences. *Philosophy and the Mind Sciences*, *1*(I), 10.
https://doi.org/10.33735/phimisci.2020.I.32