Constitutive Self-Consciousness

Raphaël Millière

Macquarie University

Abstract

The claim that consciousness constitutively involves self-consciousness has a long philosophical history, and has received renewed support in recent years. My aim in this paper is to argue that this surprisingly enduring idea is misleading at best, and insufficiently supported at worst. I start by offering an elucidatory account of consciousness, and outlining a number of foundational claims that plausibly follow from it. I subsequently distinguish two notions of self-consciousness: consciousness of oneself and consciousness of one's experience. While 'self-consciousness' is often taken to refer to the former notion, the most common variant of the constitutive claim, on which I focus here, targets the latter. This claim can be further interpreted in two ways: on a deflationary reading, it falls within the scope of foundational claims about consciousness, while on an inflationary reading, it points to determinate aspects of phenomenology that are not acknowledged by the foundational claims as being aspects of all conscious mental states. I argue that the deflationary reading of the constitutive claim is plausible, but should be formulated without using a term as polysemous and suggestive as 'self-consciousness'; by contrast, the inflationary reading is not adequately supported, and ultimately rests on contentious intuitions about phenomenology. I conclude that we should abandon the idea that self-consciousness is constitutive of consciousness.

Keywords: consciousness, self-consciousness, phenomenology, introspection, memory, attention

Many fall into the trap of confusing consciousness with self-consciousness.

Sutherland (1989: 90)

1. Introduction

As conscious creatures, we enjoy a wealth of experiences that provide us with information about the external world. But we also have the capacity to be *self*-conscious: we can be conscious of ourselves and of our very experiences. Self-consciousness is commonly taken to be a pervasive feature of our conscious mental lives (e.g., Chalmers 1996: 10). Many even go so far as to claim that self-consciousness is, in some sense to be further elucidated, *constitutive* of consciousness. Let us call this the *constitutive claim* about self-consciousness.

The constitutive claim has a long history in philosophy.[1] It occupies a central place in the phenomenological tradition, where it has received 'nearly unanimous agreement' (Gallagher and Zahavi 2012: 52). In recent years, it has also risen to considerable prominence in philosophy of mind.[2] For example, Uriah Kriegel writes that consciousness 'constitutively involves self-consciousness, in the sense that the former cannot occur in the absence of the latter' (2004: 182); likewise, according to Dan Zahavi, 'self-consciousness is a constitutive feature of phenomenal consciousness' (2014: 17); and Harry Frankfurt goes so far as to say that 'consciousness *is* self-consciousness' (Frankfurt 1988: 161).

The constitutive claim is intended to elucidate the nature of consciousness. Indeed, its proponents argue that the notion of a subject being *self-conscious* plays a constitutive role in an account of what it is for a mental state of that subject to be a *conscious* mental state. My aim in this paper is to show that the constitutive claim is more likely to obfuscate the notion of consciousness than to illuminate it – unless it is supplemented by extensive qualification that could easily be avoided by omitting the reference to self-consciousness in the first place.

The paper proceeds in two parts. In the first part, I set the stage by clarifying the key notions at play in the constitutive claim. I start by outlining foundational claims that plausibly follow from an elucidatory account of consciousness, but do not appeal to the notion of self-consciousness (§2). These are foundational claims insofar as they delineate minimally sufficient conditions for a mental state to count as a *conscious* mental state. I then distinguish two notions of self-consciousness that the constitutive claim could target (§3). While self-consciousness is generally taken to refer to the subject's consciousness *of herself* (as herself), proponents of the constitutive claim generally have a different notion in mind, namely the subject's consciousness *of her experience*.

---

[1] The full history of various versions of the constitutive claim lies beyond the scope of this paper. See Caston (2002) on an early formulation of the claim in Aristotle; MacKenzie (2015) and Coseru (2016) on the development of a similar claim in the Yogacara school of Buddhist philosophy; and Thiel (2011) and Weinberg (2016) on early modern versions of the claim.

[2] Among many examples, see Legrand (2007), Janzen (2008), Kriegel (2009), Chalmers (2010), Gertler (2010), Sebastián (2012), Siewert (2013), Strawson (2013), Zahavi (2014), Montague (2016), Nida-Rümelin (2018), Horgan (2019), Chaturvedi (2022), Giustina (2022a).

The second part critically reviews prominent versions of the constitutive claim. First, I draw attention to a potentially problematic semantic shift found in discussions of the so-called 'subjective character' or 'for-me-ness' of consciousness, often used to motivate the constitutive claim (§4). I subsequently argue that interpretations of the constitutive claim fall into two categories, focusing on particularly influential examples. In the first category are claims that merely point to aspects of phenomenology already acknowledged by the foundational claims outlined in the first part (§5). While such claims are plausible after further clarification, their appeal to self-consciousness is potentially misleading. In the second category are claims that point to distinctive aspects of phenomenology going *beyond* the foundational claims outlined in the first part (§6). I argue that such claims are under-motivated as substantive claims about the nature of consciousness, after reviewing additional arguments appealing to the putative role of constitutive self-consciousness in episodic memory and attention (§7). The upshot of this analysis is that we need not – and should not – appeal to self-consciousness to elucidate the notion of consciousness.

2. Foundational claims about consciousness

2.1 The Nagelian dictum

To investigate the nature of the relationship between consciousness and self-consciousness, it is crucial to gain some clarity on the meaning of each of these terms. The notion of consciousness that the constitutive claim seeks to elucidate is, in Block (1995)'s terminology, *phenomenal consciousness*. Following proponents of the constitutive claim, we can use Nagel (1974)'s influential account as a starting point to unravel this notion. It has become customary to apply Nagel's formulation to consciousness as a property of mental states, as follows:

(N1) A mental state M of a subject S is a *conscious* mental state of S at time *t* if and only if there is something it is like for S to be in M at *t*.

The phrase 'something it is like' (henceforth, SIL phrase) plays a central role in this account. In his article, Nagel also employs the related phrase 'what it is like' (henceforth, WIL phrase).[3] There is some disagreement about how to interpret the SIL and WIL phrases, and whether they are genuinely helpful to elucidate the notion of consciousness.[4]

It is implausible that the SIL and WIL phrases should be treated as technical terms, as some have suggested.[5] Indeed, they are generally not given a technical exposition – including in Nagel's original article. Moreover, the WIL phrase is widely used in common language and popular

[3] See Nagel (1974: 439). Several philosophers used the WIL phrase to characterize conscious mental states before Nagel, including Russell (1926), Wittgenstein (1947), Farrell (1950), and Sprigge (1971).

[4] Snowdon (2010) argues that the SIL and WIL phrases are ultimately unhelpful to characterize conscious experience. Here I will follow proponents of the constitutive claim in arguing that these phrases are helpful, if they are properly construed.

[5] See, for example, Lewis (1995: 140) and Janzen (2011: 279).

culture to characterize conscious experiences.[6] In turn, the SIL phrase merely involves existential quantification over the property captured by the WIL phrase. (Similarly: if you can meaningfully ask *what* colour *x* is, then you can say without further technical exposition that *there is some colour* such that *x* is that colour.) If the SIL and WIL phrases are not technical terms, one might wonder whether they are used to express comparative statements, given that '*x* is like *y*' commonly means '*x* resembles *y*'. However, there is a broad consensus that this is not what the SIL and WIL phrases are intended to express.[7] Nagel himself cautioned against this interpretation of the WIL phrase:

> [T]he analogical form of the English expression 'what it is *like*' is misleading. It does not mean 'what (in our experience) it *resembles*', but rather 'how it is for the subject himself'. Nagel (1974: 440, fn. 6)

Notice that in this passage Nagel puts a particular emphasis on the phrase 'for the subject himself', just like in his initial exposition of the SIL phrase he puts an emphasis on there being 'something it is like *for* the organism' (Nagel 1974: 436). This emphasis is significant to understand the Nagelian account of consciousness, although it should not be misinterpreted. Following Lormand (2004) and Stoljar (2016), I note that there are two potentially implicit argument places in a sentence such as:

(a) There is something it is like to have a headache.

There is a first argument place in (a) for the covert subject of the infinitive verb 'to have' in the embedded clause 'to have a headache'. For this subject to be overtly mentioned in the sentence, one needs to add a prepositional phrase; let us write this as 'for^{SUBJ} {NP}', where {NP} is a noun phrase. Used in this way, 'for' specifies *who* has a headache. The second potentially implicit argument place in (a) is for the covert indirect object of the impersonal verbal form 'it is' in the main clause 'there is something it is like'. Again, the indirect object can be overtly specified with a prepositional phrase; let is write this as 'for^{OBJ} {NP}'. This specifies *for whom* is it like something (to have a headache).

We can make both argument places in (a) explicit as follows:

(b) There is something it is like for^{OBJ} S, for^{SUBJ} S to have a headache.

In sentences like (b), where the indirect object of 'is like' is a creature, and the object of the infinitive or gerund is a mental state, it is natural to interpret 'for^{OBJ}' as specifying the *psychological subject* whose experience is being talked about. Thus, on Stoljar (2016)'s affective account of the semantics of sentences containing the SIL phrase or the WIL phrase, in stereotypical contexts of use, (b) means in effect:

---

[6] An early example can be found in the novel *It Is Never Too Late to Mend* by Charles Reade (1856), in which a prison guard asks a chaplain who has just tried on himself a torture device used to punish prisoners: "What is it like, sir?"; to which the chaplain replies, anticipating Nagel's philosophical commentary on the WIL phrase: "such knowledge can never be imparted by description; you shall take your turn in the jacket" (Reade 1856: 164). See Hellie (2004) and Farrell (2016) for more recent examples.

[7] See, for example, Hellie (2004), Snowdon (2010), Farrell (2016), and Stoljar (2016); for a defence of the resemblance meaning of the WIL phrase, see Gaskin (2019).

(c) There is *an experiential way* that having a headache affects S.

Accordingly, we can now unpack Nagel's account of consciousness as follows:

(N2) A mental state M of a subject S is a *conscious* mental state of S at time *t* if and only if S's being in M at *t* affects S in some experiential way at *t*.

(N2) is still not specific enough as an elucidation of consciousness, for a subject might be affected *in some experiential way* by a *non-conscious* mental state. For example, Mary's non-conscious desire for ice-cream might cause her to experience hunger. In such a case, (N2) entails that Mary's desire for ice-cream is in fact conscious; but this is not the notion of consciousness that the Nagelian dictum aims to capture. Rather, a conscious mental state must be *constitutively* such that it affects its subject in an experiential way:

(N3) A mental state M of a subject S is a *conscious* mental state of S at time *t* if and only if M is *constitutively* such that S's being in M at *t* affects S in some experiential way at *t*.

The idea of a mental state affecting its subject *in some experiential way* need not be so mysterious. We can supplement the Nagelian elucidation of consciousness as a property of mental states with a bridging principle (SIL), elucidating the meaning of the SIL phrase in terms of a mental state M being *constitutively* such a subject's being in M contributes to the subject's *overall phenomenology*:

(SIL) There is *something it is like* for a subject S to be in a mental state M at time *t* if and only if S's being in M at *t constitutively* contributes to S's overall phenomenology at *t*.

Taken together, (N1) and (SIL) straightforwardly entail:

(N4) A mental state M of a subject S is a conscious mental state of S at time *t* if and only if S's being in M at *t constitutively* contributes to S's overall phenomenology at *t*.

(N4) satisfies plausible requirements for the interpretation of the Nagelian account: it does not rely on special technical terms, and it does not express a comparative statement. As such, it provides the desired elucidatory account of consciousness as a property of mental states.

An elucidation is not an analytical definition – as it has often been remarked, it seems difficult to define phenomenal consciousness in more fundamental terms. However, the Nagelian account does provide a non-technical way to grasp the folk-psychological concept that 'conscious' refers to, when it is predicated of mental states. (N4) further elucidates the nature of the relationship between conscious mental states and psychological subjects, by making it clear that: (a) a mental state's being a *conscious* mental state involves its being appropriately related to a subject; (b) the appropriate way in which a conscious mental state is related to its subject is an *experiential* or *phenomenological* way of affecting its subject; (c) finally, a conscious mental state is *constitutively* such that it affects its subject in an experiential way, that is, it *constitutively* contributes to its subject's overall phenomenology.

2.2 Phenomenality and phenomenal character

In the Nagelian account, the SIL phrase is used to express an existentially quantified claim – there is *something*, rather than *nothing*, that it is like for a subject S to be in a conscious mental

state M. By contrast, the WIL phrase is used to refer to the 'witness' (in the logical sense) of such an existential claim – provided that there is *something*, rather than *nothing*, that it is like for S to be in M, the WIL phrase refers to *what* S's being in M is like. In other words, the WIL phrase does not merely capture the fact that M makes *some* contribution or other to S's phenomenology, but rather *what* M's contribution to S's overall phenomenology is.

By analogy with the bridging principle (SIL), we can elucidate the meaning of the WIL phrase as follows:

(WIL) *What it is like* for a subject S to be in a mental state M at time *t* is what S's being in M at *t* constitutively contributes to S's overall phenomenology at *t*.

The SIL phrase captures the *higher-order* property of making *any* contribution to the subject's overall phenomenology, rather than *none* at all. Let us call this higher-order property 'phenomenality'. When a mental state is constitutively such that there is something it is like for its subject to be in it, that mental state instantiates the higher-order property of phenomenality. In other words, phenomenality refers to the feature that is common across *all* conscious mental states (and that all non-conscious mental states lack).

In turn, the WIL phrase captures the lower-order property of making *some specific* contribution to the subject's overall phenomenology, rather than *some other specific* contribution. This lower-order property is commonly called the 'phenomenal character' of a conscious mental state. The phenomenal character of a conscious mental state is *what* that mental state constitutively contributes to the subject's overall phenomenology. As we shall see, arguments in favour of the constitutive claim often seem to obscure the nature of the relation between phenomenality and phenomenal character.

3. Two notions of self-consciousness

With this elucidatory account of consciousness in place, let us turn to the notion of *self-consciousness* featured in the constitutive claim. The term can have two broad meanings, depending on how the prefix *self-* qualifies 'consciousness'. A first notion of self-consciousness results from the common use of the prefix *self-* to indicate that the subject of the act or attitude denoted by the root of the lexical compound is also the object of that act or attitude. For example, 'self-love' denotes the love that a subject has *for herself*. Understood in such a way, 'self-consciousness' straightforwardly refers to the subject's being conscious *of herself*. Note that unlike 'consciousness', the term only refers to a property of subjects rather than mental states. Accordingly, we can formulate a first variant of the constitutive claim as follows:

(SC1) Necessarily, for any subject S, if S is conscious at *t*, then S is conscious *of S* [SELF] at *t*.

A few authors seem to endorse a claim in the vicinity of (SC1).[8] By and large, however, proponents of the constitutive claim insist that the notion of self-consciousness they deem constitutive of consciousness should not be construed as consciousness *of oneself* (let alone

[8] See, for example, Flanagan (1992: 194); Wider (2006: 77–78); Siewert (2013: 256); Nida-Rümelin (2014); Duncan (2019); Strawson (2022). For a criticism of (SC1), see Peacocke (2014: 30–39).

consciousness *of oneself as oneself*). They have in mind a distinct understanding of self-consciousness, on which the prefix *self-* indicates that the root of the lexical compound is directed at itself or takes itself as its own object. So understood, self-consciousness refers to the consciousness *of consciousness itself*. This somewhat unorthodox use of the prefix *self-* was explicitly introduced by Sartre (1943), who uses the French term '*conscience de soi*' (self-consciousness) as equivalent to '*conscience de conscience*' (consciousness of consciousness).[9]

This second notion of self-consciousness stands in need of further elucidation. At a first pass, one might think that consciousness *of consciousness itself* refers to the subject's consciousness *of her being conscious*, or *of her having a certain conscious experience*. It is natural to understand the phrase '*x* is conscious *of x's being F*' as shorthand for '*x* is conscious *that x is F*'. However, proponents of the claim that consciousness constitutively involves self-consciousness typically deny that this involves a propositional attitude, or requires the use of a concept of self (for this would entail, rather implausibly, that being conscious at all requires bearing a propositional attitude or using a concept of self).[10] Rather, they claim that '[t]o be self-aware [or, equivalently, self-conscious] is... to be conscious *of* one's experience' (Gallagher and Zahavi 2019, my emphasis).

Accordingly, we can formulate the second and main variant of the constitutive claim as follows:

(SC2) Necessarily, for any mental state M of a subject S, if M is a conscious mental state of S at *t*, then S is conscious *of M* at *t*.

The two notions of self-consciousness – consciousness *of oneself* and consciousness *of one's experience* – are not prima facie equivalent. Unfortunately, there is no standard terminology to distinguish them, which has often led commentators to confuse the two resulting variants of the constitutive claim, (SC1) and (SC2).[11] Drawing upon a piece of terminology introduced in English by the phenomenologist Aron Gurwitsch, I shall refer to the first notion of self-consciousness as 'egological self-consciousness', and to the second notion as 'non-egological self-consciousness'. While slightly awkward, these labels emphasize the fact that the former notion, but not the latter notion, involves a reference to the self.[12]

---

[9] See also Brentano (1874: 119) and Husserl (1918: 46–48) for earlier formulations along the same lines.

[10] See, for example, Zahavi (2014: 35) and Kriegel (2009: 301).

[11] Guillot (2017) suggests that this confusion is due to the lack of clarity of various formulations of the constitutive claim, while Zahavi (2018) contends that the confusion only arises from inattentive or uncharitable readings of the relevant formulations. Part of the problem is that some proponents of the constitutive claim suggest themselves that (SC1) and (SC2) are equivalent, after defining the notion of selfhood at play in the former in such a minimal way that it collapses into the latter (e.g., Nida-Rümelin 2017).

[12] See Gurwitsch (1941). This terminology is also adopted by Frank (2007: 154), Kriegel (2009: 177–79), and Zahavi (2014: 48). These labels are sometimes used differently, to mark a contrast between consciousness of oneself involving an "explicit" form of self-representation (a concept of self), and "implicit" consciousness of oneself (e.g., through visual perception in an egocentric frame of reference).

Many philosophers – across both side of the historical divide between phenomenology and analytic philosophy of mind – have explicitly defended a version of (SC2).[13] Consider, for example, the following passage:

> Consciousness *is* self-consciousness... The self-consciousness in question is a sort of *immanent reflexivity* by virtue of which every instance of being conscious grasps not only that of which it is an awareness but also the awareness of it... What I am here referring to as 'self-consciousness' is neither consciousness of a self – a subject or ego – nor consciousness that there is awareness... The reflexivity in question is merely consciousness's awareness of itself. Frankfurt (1988: 161–62)

It is not immediately obvious whether this excerpt really goes beyond the foundational claims outlined in §2. Indeed, it might be taken to highlight that conscious experiences are not merely events that happen to us; they also constitutively contribute to what it is like for us. If the non-egological notion of self-consciousness simply points to the fact that conscious experiences contribute to one's overall phenomenology, then (SC2) plausibly follows from the Nagelian elucidatory account of consciousness. However, if this notion points to a determinate aspect of phenomenology that is not already acknowledged by the Nagelian account as an aspect of all conscious mental states, then (SC2) is a distinct substantive claim that should be independently motivated.

4. From 'for-me-ness' to self-consciousness

As I mentioned at the outset, contemporary proponents of the constitutive claim generally take the Nagelian elucidatory account of consciousness as a starting point. To get from this account of consciousness to (SC2), they place a potentially distorting emphasis on some of its features.

Recall that I have characterized what is common to *all* conscious mental states as 'phenomenality', the higher-order property of there being *something* (rather than *nothing*) that it is like to be in a mental state. This notion has been glossed with a more suggestive choice of terminology, including 'subjectivity' (McGinn 1991: 29; Levine 2001: 7), 'subjective character' (Kriegel 2009: 1), and 'for-me-ness' (Zahavi and Kriegel 2016).[14] These terms are generally introduced via an emphasis on the role of 'for *x*' in the WIL and SIL phrases:[15]

> For every possible experience we have, each of us can say: whatever it is like for me to have this experience, it is *for me* that it is like that to have it... Although I live through various different experiences, there is consequently something experiential that remains the same, namely, their first-personal character. All the different experiences are characterized by a dimension of *mineness*, or *for-me-ness*. Zahavi (2014: 19)

---

[13] See, among others, Gurwitsch (1941), Sartre (1943), Goldman (1970: 70), Frankfurt (1988: 161–62), Zahavi (1999, 2005, 2014), Thompson (2007: 285), Janzen (2008, 2011), Kriegel (2009), Strawson (2013), and Montague (2016: 41).

[14] Nagel himself introduced the notion of "subjective character" to denote what I have called *phenomenal character*, rather than *phenomenality* (Nagel 1974), p. 445, fn. 11; see also p. 439).

[15] See also Levine (2001: 7), Kriegel (2009: 1), Zahavi and Kriegel (2016: 36), and Guillot (2017: 24).

This exposition calls for a few remarks. First, the use of the first-person pronoun in 'for-*me*-ness' might suggest that it refers to a property of *a particular subject*'s conscious mental states rather than *any subject*'s conscious mental states. However, phenomenality is not simply what *my* conscious mental states have in common (*qua* conscious mental states *of mine*), but what *all* conscious mental states have in common (*qua* conscious mental states). Second, as we have seen, both the SIL and the WIL phrases can be completed with 'for *x*'. Consequently, the emphasis on '*for me*' does not properly capture the distinction between phenomenality (there being something, rather than nothing, that it is like *for me*) and phenomenal character (what, specifically, it is like *for me*).

Moreover, it often suggested that the 'subjective character' or 'for-me-ness' of an experience is a *component* of its phenomenal character, rather than the higher-order property of its having some phenomenal character or other.[16] It is important not to mischaracterize the distinction between phenomenality and phenomenal character. Any mental state that instantiates the first-order property of having a certain phenomenal character instantiates *ipso facto* the second-order property of phenomenality, as the latter merely involves existential quantification over the former. Consequently, it would seem like double-counting to treat the second-order property of phenomenality as if it were a *further* first-order property to be specified as part of a mental state's phenomenal character. As a matter of logic, rather than substantive philosophical theory, a specification of *what* it is like to be in a mental state M *guarantees* that there is *something* it is like to be in M; conversely, the fact that there is *something* it is like to be in M *guarantees* that there is some answer to the question of *what* it is like to be in M. If there is *nothing* it is like to be in M, then the question of *what* it is like to be in M does not arise.

Consider by analogy the case of coloured objects. Some worldly objects are coloured – they have some colour or other, rather than none. We could use the term 'chromaticity' for the second-order property of having *some* colour(s) or other, rather than none; and 'chromatic character' for the first-order property of having *some specific* colour(s), rather than any other. Accordingly, all coloured objects instantiate chromaticity, although they may differ with respect to their chromatic character. However, we cannot treat chromaticity as if it were a *further* first-order property of coloured objects to be specified as part of the object's chromatic character.

The emphasis on the 'for *x*' phrase in the Nagelian account leads proponents of the constitutive claim to offer the following gloss: there is something it is like *for* a subject to be in a mental state M only if S is conscious (or aware) *of M*. Thus, Zahavi contends that the notion of for-me-ness 'was introduced in order to capture the special awareness we have of our ongoing experiences' (2018: 706), and 'is simply a question of being pre-reflectively aware of one's own consciousness' (2014: 24). Likewise, Kriegel argues that 'a mental state has subjective character just in case it is *for* the subject, in the sense that the subject has a certain awareness of it' (2009: 38).

Once we get to such formulations of the constitutive claim, through successive glosses of the notions of 'subjective character' and 'for-me-ness', one starts to wonder how far we have wandered away from the Nagelian elucidatory account of consciousness. With this question in mind, I will examine in more detail the main versions of the constitutive claim to determine

---

[16] See, for example, Kriegel (2009: 45) and Kriegel (forthcoming).

whether they point to any aspect of phenomenology that is not already acknowledged by the foundational claims outlined in §2.

5. The deflationary interpretation of the constitutive claim

A number of proponents of (SC2) put a great deal of emphasis on the idea that the self-consciousness they deem constitutive of consciousness is distinctly 'thin', 'minimal', or 'flat' (e.g., Thomasson 2000; Strawson 2013; Zahavi 2014, 2017; O'Conaill 2017; Frank 2022). Echoing similar remarks by Husserl and Sartre, they specifically insist that this minimal form of self-consciousness is 'pre-reflective': it is not a matter of *reflecting* upon one's conscious mental state. Rather, it refers to the fact that 'experience is given [to the subject], not as an object, but precisely as subjectively lived through' (Zahavi 2014: 16).

This minimalist characterization of pre-reflective self-consciousness is meant to distinguish it from a higher-order state. Higher-order theories (HOT) of consciousness are committed to a claim that seems very similar to (SC2): the so-called *transitivity principle*, according to which a conscious mental state is a mental state whose subject is, in some way, aware of (Rosenthal 1997). This awareness is understood in terms of a hierarchical relation between mental states: a mental state is conscious, rather than unconscious, if and only if it is suitably represented by a higher-order mental state (Rosenthal 1997). However, HOT theorists often state that the higher-order mental state through which a subject is aware of a lower-order conscious mental state is not, itself, a *conscious* mental state. In other words, the higher-order mental state does not make a *distinct* contribution to the subject's phenomenology, over and above what it is like for the subject to be aware of the lower-order mental state. Insofar as HOT theorists are proponents of (SC2), then, they seem endorse a deflationary interpretation of the claim that does not point to aspects of phenomenology beyond those acknowledged by foundational claims about consciousness.[17]

It is not immediately obvious that this deflationary interpretation is also favoured by proponents of (SC2) who, like Zahavi, reject the higher-order view. Their understanding of pre-reflective self-consciousness is often explicitly related to the 'for *x*' phrase in the Nagelian dictum, through the notion of 'for-me-ness':

> The for-me-ness... of experience simply refer[s] to the subjectivity of experience, to the fact that [one's] experiences are pre-reflectively self-conscious and thereby present in a distinctly subjective manner... Zahavi (2014: 41)

How are we to understand the claim that experiences are 'present in a distinctly subjective manner'? To shed light on his understanding of pre-reflective self-consciousness or for-me-ness,

---

[17] Some HOT theorists do occasionally suggest that unconscious higher-order thoughts can make a distinct contribution to a subject's phenomenology. For example, the acquisition of refined gustatory concepts instantiated in a higher-order thought can affect one's first-order experience of a wine's taste (Rosenthal 2005: 187–88). However, given that the higher-order thought is not itself conscious, this is not so different from Mary's non-conscious desire for ice cream causing her to experience hunger. If that is the case, the HOT theorist should remain committed to the deflationary version of (SC2). (I am grateful to an anonymous referee for raising that point.)

Zahavi considers the following thought experiment (Zahavi 2014), pp. 22-23; see also (Zahavi 2005), p. 127):

> PHENOMENAL TWINS
> Two individuals, Mick and Mack, are physically and psychologically type-identical. Mick and Mack's respective token experiences, although numerically distinct, have exactly the same type of phenomenal character.

For the sake of clarity, consider Mick and Mack's experiences at a specific time *t*; call Mick's experience at *t* 'e_1' and Mack's experience at *t* 'e_2'. From a third-person perspective, Zahavi claims, there is no 'relevant qualitative difference between the two [experiences]' (*ibid.*, p. 22). Having said that, he prompts the reader to abandon the third-person perspective and consider what it is like for Mick to have e_1 at *t*. According to Zahavi, there is for Mick 'a crucial difference between [e_1 and e_2], a difference that would prevent any kind of conflation [between e_1 and e_2]' (*ibid.*, p. 22). This difference consists in the fact that only e_1, and not e_2, is 'given first-personally to [Mick] at all, and therefore part of [Mick's] experiential life' (*ibid.*, p. 22). To use Zahavi's alternative terminology, Mick is pre-reflectively conscious of e_1, but *not* of e_2.

In accordance with the elucidatory account of consciousness offered in §2, we can express the relevant difference as follows: while there is (constitutively) something it is like for^{OBJ} Mick, for^{SUBJ} Mick to have e_1, there is (constitutively) nothing it is like for^{OBJ} Mick, for^{SUBJ} Mack to have e_2. Importantly, this does not entail that pre-reflective self-consciousness or 'for-me-ness' is a special qualitative property of the experience that is somehow unique to the experience's subject, i.e., that Mick's experience has a distinctive flavor of 'for-*Mick*-ness'. One can see how, in this respect, Zahavi's suggestion that 'for Mick, his experience will be quite unlike Mack's experience (and vice versa)' (*ibid.*, p. 24) could be misunderstood: given that Mick and Mack's experiences are type-identical (i.e., they share the exact same type of phenomenal character), what it is like for^{OBJ} Mick, for^{SUBJ} Mick to have e_1 is *exactly the same* as what it is like for^{OBJ} Mack, for^{SUBJ} Mack to have e_2.

The purpose of PHENONEMAL TWINS, as I understand it, is to stress the fact that being the subject of a conscious mental state is not *merely* a matter of instantiating a property like any other. In particular, it differs substantially from instantiating a property like mass. Mick's having a mass of 80kg is a different token instantiation from Mack's having a mass of 80kg, but there is a mere numerical distinction between the two token instantiations. Mick's having e_1 is also a different token instantiation from Mack's having e_2, but there is a sense in which the difference goes further than this metaphysical fact. Indeed, there is something it is like for^{OBJ} Mick, for^{SUBJ} Mick to undergo e_1, and there is something it is like for^{OBJ} Mack, for^{SUBJ} Mack to undergo e_2. By contrast, there is nothing it is like for^{OBJ} Mick, for^{SUBJ} Mick to have a mass of 80kg, and vice versa.

The second lesson to be drawn from the thought experiment is that each conscious experience constitutively contributes to the overall phenomenology of *one and only one* subject of experience. This is what is occasionally called (somewhat contentiously) the 'privacy' of experience (e.g., Sprigge 1969). Zahavi's emphasis on this point is apparent from his elucidation of 'for-me-ness' as pointing to 'the fact that our acquaintance with our own experiential life differs from the acquaintance we have with the experiential life of others and vice versa' (Zahavi 2014: 24). Thus, on this deflationary account, to say that I am pre-reflectively conscious of a

mental state M – or, equivalently, to say that M has 'for-me-ness' – is to say (a) that M constitutively contributes to my overall phenomenology, and (b) that M does *not* constitutively contribute to any other subject's overall phenomenology. Given this minimal understanding of pre-reflective self-consciousness or 'for-me-ness', the claim that a mental state is a conscious mental state if and only its subject is pre-reflectively conscious of it – that is, if and only if it has 'for-me-ness' – strikes me as true, because it does not go beyond the foundational claims about consciousness outlined in §2.

Zahavi himself emphasizes that one should not understand the notion of 'pre-reflective self-consciousness' on an act-object model, similarly to the notion of 'consciousness *of an object*': 'the experience [of an apple] is not itself an object on a par with the apple, but instead constitutes the very access to the appearing apple' (Zahavi 2014: 35). The relevant notion of pre-reflective self-consciousness must be understood in a minimal way, so that what it is for S to be (pre-reflectively) conscious of mental state M is for M to be a conscious mental state of which S is the subject; or, equivalently, for there to be something it is like for^{OBJ} S, for^{SUBJ} S to be in M; or, also equivalently, for M to constitutively contribute to S's overall phenomenology (to the exclusion of anyone else's).

Nonetheless, one can see how the notion of pre-reflective self-consciousness can be mistaken for a less minimal notion. Using the term 'self-consciousness' and its cognates to point to notions that can be constructed within the scope of the Nagelian account of consciousness requires a considerable amount of exposition and qualification (e.g., with adjectives such as 'pre-reflective', as well as lengthy glosses). Moreover, despite repeated clarification from proponents of the deflationary interpretation of (SC2) about what they intend by 'self-consciousness', this view has often been misconstrued as pointing to a more substantive notion.[18] Given these two observations, it would be preferable – both for proponents of the deflationary interpretation and for their readers – to avoid the terminology of 'self-consciousness' altogether when discussing foundational claims about consciousness.

6. The inflationary interpretation of the constitutive claim

If all proponents of (SC2) simply endorsed a deflationary interpretation of the claim, the discussion of constitutive self-consciousness would be, at worst, likely to invite misunderstanding. However, many appear to reject such interpretation, favouring instead a more inflationary view (e.g., Kriegel 2003; Janzen 2008; Kriegel 2009, forthcoming; Montague 2017; Chaturvedi 2022; Giustina 2022a, 2022b). Let us call these *inflationists* about constitutive self-consciousness.

Inflationists also generally introduce their view through a discussion of the 'subjective character' or 'for-me-ness' of conscious mental states. For example, Uriah Kriegel – one of the most influential inflationists – claims that 'it is central to subjective character that it enables an epistemic or mental relation between the subject and her experience' (Kriegel 2009: 105). Like HOT theorists, inflationists characterize this epistemic or mental relation is the relation of *awareness*. To distinguish the subject's awareness of worldly objects from the subject's

[18] See Zahavi (2017, 2018) for a defense against misunderstandings of his minimal interpretation of (SC2).

awareness of her experience itself, Kriegel calls the former *outer awareness* and the latter *inner awareness*. The specific version of (SC2) to be elucidated is what has come to be known as 'Ubiquity of Inner Awareness Thesis' (UIA):

(UIA) For any mental state M of a subject S, if M is conscious at *t*, then (a) S is aware of M at *t* and (b) S's awareness of M is part of S's overall phenomenology at *t*.[19]

(UIA)'s clause (b) introduces the contentious idea that the subject's awareness of her conscious mental state is itself part of the subject's overall phenomenology, suggesting that it is a further *component* of the state's phenomenal character. On this view, the relation between inner awareness or 'for-me-ness' and phenomenal character is not exhausted by the idea that the former is a determinable of which the latter is a determinate. Thus, Kriegel compares the relation between inner awareness and the subject's phenomenology to the relation between a keystone and a thirteen-stone masonry arch:

> As a mere commonality and yet a substantive one, for-me-ness serves a double function as both (i) a component among others in a conscious state's overall phenomenal character and (ii) a precondition for the existence of all other phenomenal components (*as* phenomenal components). Compare the keystone of a thirteen-stone masonry arch. On the one hand, it is a stone among others composing the arch, as intrinsically 'beefy' as the other twelve. On the other hand, if we remove it the whole arch collapses, and to that extent it is a precondition for there being any other arch-component. Kriegel (forthcoming)

By analogy with the keystone, your (inner) awareness of your (outer) awareness of an external object is taken to both (a) enable your (outer) awareness of that object to make any contribution *at all* to your phenomenology, and (b) make its own, *independent* contribution to your phenomenology. The resulting version of the constitutive claim is clearly substantive, and does not plausibly follow from the Nagelian elucidatory account of consciousness. It stands in need of additional support. Inflationists insist that that their view is not only phenomenologically plausible, but best supported by an appeal to phenomenology (e.g., Zahavi 2005: 24; Janzen 2008: 101; Kriegel 2009, pp. 50, 182, 175; Montague 2017: 378; Giustina 2022b: 346). In fact, many take their view to be 'so self-evident as to *allow* no cogent argument that would derive it from truths even more obvious than it' (Kriegel 2019: 144). However, there is no shortage of philosophers attesting that they do not share this intuition.[20] If inner awareness were indeed a ubiquitous and phenomenologically manifest feature of consciousness, one might expect a broader agreement on its existence.

To alleviate concerns about the elusiveness of inner awareness, inflationists typically emphasize that it is normally 'peripheral', or 'inattentive', and that it can never be introspected (e.g., Kriegel 2009; Giustina and Kriegel 2017; Montague 2017). Thus, on Kriegel's 'attention-shift' model of introspection, introspection consists in shifting attention from external objects to one's awareness

---

[19] Kriegel (2009: 181). See also Chaturvedi (2022) and Giustina (2022a; 2022b).

[20] See Dretske (1993), Siewert (1998), Smith (2004); Gennaro (2008), Gertler (2012), Mehta (2013), McClelland (2015), Howell and Thompson (2017), and Lang (2021).

of these objects, through a conversion of *peripheral* inner awareness into *focal* inner awareness. Introspection, on this view, is nothing but an inbuilt inner awareness of one's experience to which attentional resources have been allocated. It follows that inner awareness itself cannot be introspected. The reason why critics deny the ubiquity of inner awareness is that they fail to notice such awareness upon introspecting their experience – but this is precisely what the attention-shift model of introspection predicts: 'introspecting cannot reveal peripheral inner awareness because it *annihilates* it (by supplanting it)' (Kriegel 2009: 184).

This move allows inflationists to characterize their view as compatible with the so-called transparency of experience. The traditional formulation of the transparency thesis states that when one attempts to attend to one's experience, one can only attend to features of the worldly objects presented by one's experience (Harman 1990). In other words, attempting to introspect one's experience does not reveal properties of the experience itself, but only properties of its objects – properties of what it is an experience *of*. Note that this formulation of the transparency thesis is not strictly incompatible with (UIA), because (UIA) does not explicitly say that inner awareness is ever *focal* rather than *peripheral*. However, inflationists do need to give some account of what happens when we attend to our occurrent experience, that is, when our *peripheral* inner awareness of that experience is supplanted by the *focal* inner awareness of it.

Kriegel denies that his view is incompatible with the transparency thesis, which he formulates in the following way:

(TE) For any experience *e* and any feature F, if F is an introspectible feature of *e*, then F is (part of) the first-order representational content of *e*.[21]

The notion of first-order representational content refers here to content that does not involve properties of the representational vehicle, but only involves worldly objects and their properties. In other words, (TE) states that the only features of one's experience that are available to introspection are those that are part of the experience's world-directed representational content. Kriegel goes on to say that his view is compatible with (TE), because the attention-shift model predicts that *peripheral* inner awareness cannot be introspected. Upon introspecting, peripheral inner awareness turns into focal inner awareness, and such awareness is directed to the first-order representational content of the experience.

This treatment of the transparency of experience raises several concerns. Firstly, (TE) does not reflect traditional formulations of the transparency thesis, which precisely deny that one can introspect properties of one's experience: any attempt to attend to your experience of the blue sky would result in attending to what your experience represents, namely the blue sky itself.[22] Thus, the following would be a more accurate statement of the transparency thesis:

(TE*) For any experience *e*, upon attempting to attend to a property of *e*, one can only attend to objects and properties that are part of the worldly scene that *e* represents.

---

[21] Kriegel (2009: 181); see also Kriegel (forthcoming).

[22] Interestingly, Kriegel himself initially introduces the transparency thesis as the claim that "whenever we try to introspect one of our experiences, we can become acquainted only with what it is an experience of – not with the experiencing itself" (Kriegel 2009: 69). This is presumably not equivalent to (TE).

The formulation of the transparency thesis as (TE*) suggests that it is in fact *incompatible* with the inflationist interpretation of (SC2). If introspection is just focal inner awareness, and focal inner awareness is just peripheral inner awareness plus attention, then it would seem that anything of which a subject has *peripheral* inner awareness is available to introspection – given that it is a possible target for *focal* inner awareness. It follows that there must be something that is available to introspection over and beyond the experience's first-order representational content. If not, then (UIA) would become vacuous, for there would be no difference between outer awareness and inner awareness. Because the focal inner awareness that constitutes introspection just is the same in-built inner awareness that happens to be peripheral in the non-introspective case, introspecting one's experience cannot be the same thing as attending to the object and properties it represents. Thus, the inflationary interpretation of (SC2) does not appear to be compatible with the traditional version of the transparency thesis, Kriegel himself deems phenomenologically plausible (e.g., Kriegel 2009: 69–71). In fact, this incompatibility is acknowledged by other inflationists who explicitly reject the transparency thesis.[23]

Kriegel admits that his inflationary interpretation of (UIA) is 'a substantive claim' (Kriegel 2009: 47, fn. 38), and is not 'trivial or uncontroversial' (Kriegel 2009: 50). Nonetheless, he maintains not only that this view is phenomenologically plausible, but also that his rebuttal of potential objections (including the objection that it is incompatible with the transparency thesis) should lead us to conclude that 'we have *no good reason* not to accept that the overall phenomenology of a conscious subject at a time always and necessarily includes an element of inner awareness' (*ibid.*, p. 196). I have argued that we do, in fact, have good reasons not to accept that view without additional justification. The transition from the Nagelian dictum to the inflationary interpretation of (UIA), through the emphasis on the 'for *x*' phrase, seems to involve a form of double-counting that equivocates on the meaning of subjective character or for-me-ness.

Inflationists insist that their view, aside from being phenomenologically plausible, has explanatory purchase in providing an account of 'the substantive commonality among conscious states' (Kriegel forthcoming). However, on the alternative proposal that actually flows from the Nagelian elucidation of consciousness, the substantive commonality among conscious mental states is merely the second-order property of phenomenality: all conscious mental states are constitutively such that they make *some* contribution *or other* to their subject's overall phenomenology. This property is not a *component* of the first-order property of phenomenal character in any meaningful sense. To treat it as such is not only phenomenologically implausible, but perhaps also metaphysically suspect – in so far as a second-order property obtained by existential quantification over a first-order property cannot be a component of the first-order property.

The keystone analogy illustrates this point. As Kriegel notes, a keystone is both (i) a genuine part of any thirteen-stone masonry arch, and (ii) required – at least as a matter of nomological necessity – for any such arch to have parts at all. Indeed, the keystone makes a distinctive contribution to the 'architectural character' of the arch, as it were, in addition to securing its structural integrity, thus enabling it to exist *qua* arch. By analogy, it is at least conceivable that

[23] See Chaturvedi (2022: 9–10), Giustina (2022b: 343), Janzen (2008: ch. 7), Montague (2016: ch. 4).

some type of phenomenal property be both (i) a genuine part of any mental state's phenomenal character, and (ii) required – at least as a matter of nomological necessity – for any mental state to have phenomenal character at all. For example, there could be a phenomenal property of 'pure awareness' that is always present in the background of any conscious experience.[24] Instantiating this special phenomenal property could be a (nomologically) necessary condition for a mental state to instantiate phenomenality. However, this would not entail that the two properties are *one and the same* (by analogy: the keystone is not identical with the arch's structural integrity, even if it is a necessary condition for it).

In summary, the inflationary version of (SC2) is a substantive claim that does not plausibly follow from foundational claims about consciousness, and appears to rest on contentious intuitions about phenomenology. As such, it stands in need of further support as a claim about what is constitutive of all conscious mental states. Additionally, it appears to be premised on a potentially misleading treatment of phenomenality as a further *component* of phenomenal character.

7. Arguments from memory and attention

Some inflationists have proposed to supplement appeals to phenomenology or conceptual analysis with additional arguments to support their interpretation of the constitutive claim, appealing instead to the purported role of self-consciousness in memory and attention. In light of the foregoing discussion, I will briefly argue that these supplemental arguments do not provide adequate independent support to the inflationist view.[25]

The argument from memory proceeds as follows (adapting from Kriegel 2019 and Giustina 2022a, 2022b):

(P1) For any conscious mental state M of a subject S at time t_1, there is a time t_2 such that S can episodically remember M at t_2 (where $t_2 > t_1$);

(P2) For any subject S and event E occurring at time t_1, S can episodically remember E at time t_2 (where $t_2 > t_1$) only if S is aware of E at t_1;

(C) Therefore, for any conscious mental state M of S at t_1, S is aware of M at t_1.

Assuming that mental states are events (i.e., instantiations of phenomenal properties by a subject at a time), the argument is valid. Note, however, that nothing in this argument strictly requires the subject's awareness of the conscious mental states she subsequently recalls to be phenomenologically manifest. The relevant awareness may just as well be construed in the deflationary sense outlined earlier, namely, in terms of the mental state constitutively contributing to the subject's overall phenomenology.

---

[24] For proposals along these lines, see e.g. Albahari (2009) and Metzinger (2020). This is, of course, a substantive and controversial claim in its own right.

[25] See also Schear (2009) and Stoljar (2021) for critical discussion.

Inflationists might insist that 'When I recall yesterday night's concert, I recall not only the concert... but also my experience of it' (Giustina 2022a: 7). But there is a lingering ambiguity in this response. Certainly, my perceptual state while watching a concert makes a contribution to my phenomenology; so does my subsequent state of episodically remembering the concert. Furthermore, part of what it is like to episodically remember the concert plausibly involves a sense of ownership over the memory – feeling like it originates in my own past experience (Klein and Nichols 2012). Recent work on episodic memory suggests that this sense of ownership is grounded in source monitoring mechanisms, whose goal is to determine whether the information conveyed by the memory has been acquired first-hand or not (Michaelian 2016; Mahr and Csibra 2018; Boyle 2019). These mechanisms employ various content-based markers to determine whether a memory originates in experience, such as the level of detail of the memory. Importantly, the existence of such content-based markers does not depend on the occurrence of a phenomenologically manifest inner awareness of the experience when it unfolds. Accordingly, the fact that one typically remembers episodic memories as originating in one's own experience does not vindicate the inflationist view of inner awareness.

One might object that the second premise of the argument is intended to hold not just for experiences, but for all events. The awareness that a subject is supposed to have in order to remember an event is taken to be the kind of awareness we can have not only of our experience of a concert, but of the concert itself. However, it is not obvious that we can be aware of a concert in the deflationary sense outlined above; that is, it might seem a bit odd to say that the concert itself – as opposed to my experience of it – makes a distinct contribution to my overall phenomenology. Thus, the objection goes, it is not clear that we can understand (P2) as involving awareness in only the deflationary sense.[26]

This objection highlights an ambiguity in the argument from memory. Suppose you attended a performance of Richard Wagner's *Tristan und Isolde* for the first time. After the performance, you recall the famous dissonant chord first heard in the third bar of the opera's overture. According to (P2), if you can remember the chord, then you were aware of it when you heard it; and if you can remember your *experience* of the chord, then you were aware of that experience itself when you had it. But is there really a difference between remembering the chord as you heard it, and remembering your auditory experience of the chord? It is unclear what that difference might be; it seems that these are two ways of describing the same act of remembering. It would be misguided to say, on the basis of (P2), that you were aware of two events at once when you heard the chord: the chord as you heard it, and your auditory experience of the chord. Rather, you were aware of the chord *through* your auditory experience of it; that is, your auditory experience of the chord made a distinct contribution to your overall phenomenology.

It is worth noting that the argument appears to presuppose an archival account of episodic memory, on which remembering merely consists in retrieving a discrete memory trace encoding an event one was aware of in the past. Most modern accounts of episodic memory agree, by contrast, that it involves a generative component similar to imagination (Michaelian 2011; Robins 2016). For example, remembering the chord from *Tristan*'s overture plausibly requires 'hearing' it in the mind's ear by making use of auditory imagination to fill in the details. If

[26] I am grateful to an anonymous referee for raising this objection.

episodic memory is at least partly generative, then not all aspects of events we remember – as we remember them – need be aspects we were aware of when the events occurred. When you remember what the auditory experience of *Tristan*'s chord felt like, it is not because you are able to retrieve a phenomenologically manifest inner awareness of that experience that was encoded in a memory trace, in addition to the sound of the chord itself. A more plausible story might go as follows: you consciously heard the sound of the chord in the opera house; hearing that sound made a distinct contribution to your auditory phenomenology; as a result, the sound was encoded in an episodic gist; by retrieving the gist and filling in the details with auditory imagination, you can 'hear' the sound in your mind's ear as you remember it. Nothing in this account suggests that you could only recall hearing *Tristan*'s chord if you had an inner awareness of your auditory experience of it that made a distinct contribution to your phenomenology, over and above whatever contribution the first-order experience made to your auditory phenomenology.[27]

The argument from attention proceeds along somewhat similar lines (adapting from Giustina 2022a):

(P1) For any conscious mental state M of a subject S, it is possible for S to *consciously* attend to M;

(P2) If it is possible for S to *consciously* attend to M, then S is aware of M;

(C) Therefore, for any conscious mental state M of a subject S, S is aware of M.

Once again, nothing in this argument mandates an inflationary reading of the subject's awareness of her conscious mental state. If I consciously see a blue sky, I may be in a position to attend to what my visual experience of the blue sky contributes to my overall phenomenology. This is only possible insofar as my visual experience does contribute to my phenomenology. It does not entail, however, that my visual experience making some specific contribution to my phenomenology makes an *additional*, *independent* contribution to my phenomenology – beyond what the experience itself contributes to it. In other words, it does not entail that I have a phenomenologically manifest inner awareness of my visual experience. Perhaps I do have such manifest awareness once I consciously attend to the experience, which is consistent with deflationary accounts of (SC2).[28] However, the claim that this manifest awareness is *ubiquitous* in conscious experience, let alone *constitutive* of consciousness, is not adequately supported by the argument from attention.

8. Conclusion

I have argued that none of the foundational claims that plausibly follow from an elucidatory account of consciousness invite the idea that consciousness constitutively involves self-

[27] See also Millière & Newen (2022) for a detailed account of the role that different forms of self-representation play in episodic memory.

[28] HOT theorists, for example, hold that introspecting a conscious experience involves tokening a *third-order* non-conscious mental states representing the second-order mental state through which the first-order mental state is conscious, thereby making the second-order mental state conscious as well.

consciousness. While proponents of this constitutive claim emphasize that the relevant notion of self-consciousness should be understood in a non-egological sense (consciousness *of one's experience*), it is ultimately dubious that appealing to this notion does much to elucidate consciousness without extensive and potentially misleading qualifications.

On the deflationary interpretation of the constitutive claim, the non-egological notion of self-consciousness merely points to aspects of phenomenology also acknowledged by the foundational claims. Specifically, it highlights the fact that all conscious mental states are constitutively such that they make some contribution or other to one (and only one) subject's phenomenology. Once it has been properly understood, this deflationary interpretation of the claim is certainly plausible; but it is preferable to formulate foundational claims about consciousness without using a term as polysemous as 'self-consciousness', which is likely to invite a stronger reading and has in fact generated much confusion in the literature. On the more inflationary interpretation of the constitutive claim, the non-egological notion of self-consciousness is treated both as the second-order property of phenomenality and as a first-order component of the phenomenal character of mental states. The resulting claim stands in need of further support, and, more problematically, appears to rest on contentious conceptual grounds.

The idea of constitutive self-consciousness may prove more confusing than illuminating. At best, the appeal to self-consciousness to elucidate the nature of consciousness is an unnecessary lexical detour that ultimately circles back to less suggestive terms. At worst, it can result in a substantive claim that remains under-motivated as an account of what is constitutive of consciousness. Self-consciousness in an important feature of our conscious mental life; but it is not constitutive of it.

Acknowledgements

I am particularly grateful to William Child, Martin Davies, Christopher Peacocke, Frédérique de Vignemont, and Dan Zahavi for their insightful feedback on earlier versions of this paper. I also benefited from constructive comments received during presentations at the University of Oxford, the Institut Jean Nicod, and the Mind Group at New York University, notably from Simon Brown, David Chalmers, Robert von Gulick, Thomas Nagel, Luke Roelofs, and Bénédicte Veillet.

References

Albahari, Miri (2009) 'Witness-Consciousness: Its Definition, Appearance and Reality', *Journal of Consciousness Studies* **16**: 62–84.

Block, Ned (1995) 'On a Confusion About a Function of Consciousness', *Brain and Behavioral Sciences* **18**: 227–47. doi:10.1017/S0140525X00038188.

Boyle, Alexandria (2019) 'The Impure Phenomenology of Episodic Memory', *Mind & Language* **n/a**: 1–20. doi:10.1111/mila.12261.

Brentano, Franz (1874/1995) *Psychology from an Empirical Standpoint*, International Library of Philosophy, Oskar Kraus and Linda L McAlister, eds. Routledge.

Caston, Victor (2002) ‘Aristotle on Consciousness’, *Mind* **111**: 751–815. doi:10.1093/mind/111.444.751.

Chalmers, David J (1996) *The Conscious Mind: In Search of a Fundamental Theory*. Oxford University Press.

Chalmers, David J (2010) *The Character of Consciousness*. Oxford University Press.

Chaturvedi, Amit (2022) ‘Attentional Structuring, Subjectivity, and the Ubiquity of Reflexive Inner Awareness’, *Inquiry* **0**: 1–40. doi:10.1080/0020174X.2022.2032324.

Coseru, Christian (2016) ‘Dignaga and Dharmakirti on Perception and Self-Awareness’, in John Powers, ed., *The Buddhist World*: 526–37. London and New York: Routledge.

Dretske, Fred (1993) ‘Conscious Experience’, *Mind* **102**: 263–83.

Duncan, Matt (2019) ‘The Self Shows Up in Experience’, *Review of Philosophy and Psychology* **10**: 299–318. doi:10.1007/s13164-017-0355-2.

Farrell, B A (1950) ‘Experience’, *Mind* **59**: 170–98.

Farrell, J (2016) ‘’What It Is Like’ Talk Is Not Technical Talk’, *Journal of Consciousness Studies* **23**: 50–65.

Flanagan, Owen J (1992) *Consciousness Reconsidered*. MIT Press.

Frank, Manfred (2007) ‘Non-Objectal Subjectivity’, *Journal of Consciousness Studies* **14**: 152–73.

Frank, Manfred (2022) ‘In Defence of Pre-Reflective Self-Consciousness: The Heidelberg View’, *Review of Philosophy and Psychology*. doi:10.1007/s13164-022-00619-z.

Frankfurt, Harry G (1988) *The Importance of What We Care About: Philosophical Essays*. Cambridge University Press.

Gallagher, Shaun and Dan Zahavi (2012) *The Phenomenological Mind*. Routledge.

Gallagher, Shaun and Dan Zahavi (2019) ‘Phenomenological Approaches to Self-Consciousness’, in Edward N Zalta, ed., *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University.

Gaskin, Richard (2019) ‘A Defence of the Resemblance Meaning of ‘What It’s Like’’, *Mind* **128**: 673–98. doi:10.1093/mind/fzx023.

Gennaro, Rocco J (2008) ‘Representationalism, Peripheral Awareness, and the Transparency of Experience’, *Philosophical Studies* **139**: 39–56. doi:10.1007/s11098-007-9101-4.

Gertler, Brie (2010) *Self-Knowledge*. Routledge.

Gertler, Brie (2012) ‘Conscious States as Objects of Awareness: On Uriah Kriegel, Subjective Consciousness: A Self-Representational Theory’, *Philosophical Studies* **159**: 447–55. doi:10.1007/s11098-011-9763-9.

Giustina, Anna (2022b) ‘A Defense of Inner Awareness: The Memory Argument Revisited’, *Review of Philosophy and Psychology*. doi:10.1007/s13164-021-00602-0.

Giustina, Anna (2022a) ‘Nature Does Not Yet Say No to Inner Awareness: Reply to Stoljar’, *Erkenntnis*. doi:10.1007/s10670-022-00557-3.

Giustina, Anna and Uriah Kriegel (2017) ‘Fact-Introspection, Thing-Introspection, and Inner Awareness’, *Review of Philosophy and Psychology* **8**: 143–64. doi:10.1007/s13164-016-0304-5.

Goldman, Alvin I (1970) *A Theory of Human Action*. Princeton University Press.

Guillot, Marie (2017) ‘I Me Mine: On a Confusion Concerning the Subjective Character of Experience’, *Review of Philosophy and Psychology* **8**: 23–53. doi:10.1007/s13164-016-0313-4.

Gurwitsch, Aron (1941) ‘A Non-Egological Conception of Consciousness’, *Philosophy and Phenomenological Research* **1**: 325–38. doi:10.2307/2102762.

Harman, Gilbert (1990) ‘The Intrinsic Quality of Experience’, *Philosophical Perspectives* **4**: 31–52.

Hellie, Benj (2004) ‘Inexpressible Truths and the Allure of the Knowledge Argument’, in Yujin Nagasawa, Peter Ludlow and Daniel Stoljar, eds., *There’s Something About Mary*: 333. MIT Press.

Horgan, Terry (2019) ‘Pre-Reflective Vs. Reflexive Self-Awareness’, *ProtoSociology* **36**: 298–315. doi:10.5840/protosociology20193611.

Howell, Robert J and Brad Thompson (2017) ‘Phenomenally Mine: In Search of the Subjective Character of Consciousness’, *Review of Philosophy and Psychology* **8**: 103–27. doi:10.1007/s13164-016-0309-0.

Husserl, Edmund (1918/2001) *Die Bernauer Manuskripte Über das Zeitbewusstsein (1917/18)*, Husserliana: Edmund Husserl Gesammelte Werke, no. XXXIII, Rudolf Bernet and Dieter Lohmar, eds. Springer Netherlands.

Janzen, Greg (2008) *The Reflexive Nature of Consciousness*, vol. 72, Advances in Consciousness Research. John Benjamins Publishing Company.

Janzen, Greg (2011) ‘In Defense of the What-It-Is-Likeness of Experience’, *The Southern Journal of Philosophy* **49**: 271–93. doi:10.1111/j.2041-6962.2011.00074.x.

Klein, Stanley B and Shaun Nichols (2012) ‘Memory and the Sense of Personal Identity’, *Mind* **121**: 677–702. doi:10.1093/mind/fzs080.

Kriegel, Uriah (2003) ‘Consciousness as Intransitive Self-Consciousness: Two Views and an Argument’, *Canadian Journal of Philosophy* **33**: 103–32.

Kriegel, Uriah (2004) ‘Consciousness and Self-Consciousness’, *The Monist* **87**: 182–205.

Kriegel, Uriah (2009) *Subjective Consciousness: A Self-Representational Theory*. Oxford University Press.

Kriegel, Uriah (2019) ‘Dignāga’s Argument for the Awareness Principle: An Analytic Refinement’, *Philosophy East and West* **69**: 143–55. doi:10.1353/pew.2019.0003.

Kriegel, Uriah (forthcoming) ‘The Three Circles of Consciousness’, in M Guillot and M Garcia-Carpintero, eds., *The Sense of Mineness*. Oxford University Press.

Lang, Stefan (2021) ‘A Methodological Objection to a Phenomenological Justification of the Ubiquity of Inner Awareness’, *ProtoSociology* **38**: 59–73. doi:10.5840/protosociology2021384.

Legrand, Dorothée (2007) ‘Pre-Reflective Self-as-Subject from Experiential and Empirical Perspectives’, *Consciousness and Cognition* **16**: 583–99. doi:10.1016/j.concog.2007.04.002.

Levine, Joseph (2001) *Purple Haze: The Puzzle of Consciousness*. Oxford University Press.

Lewis, David (1995) ‘Should a Materialist Believe in Qualia?’, *Australasian Journal of Philosophy* **73**: 140–44. doi:10.1080/00048409512346451.

Lormand, Eric (2004) ‘The Explanatory Stopgap’, *Philosophical Review* **113**: 303–57. doi:10.1215/00318108-113-3-303.

MacKenzie, Matthew (2015) ‘Reflexivity, Subjectivity, and the Constructed Self: A Buddhist Model’, *Asian Philosophy* **25**: 275–92. doi:10.1080/09552367.2015.1078140.

Mahr, Johannes B and Gergely Csibra (2018) ‘Why Do We Remember? The Communicative Function of Episodic Memory’, *The Behavioral and Brain Sciences* **41**. doi:10.1017/S0140525X17000012.

McClelland, Tom (2015) ‘Affording Introspection: An Alternative Model of Inner Awareness’, *Philosophical Studies* **172**: 2469–92. doi:10.1007/s11098-014-0421-x.

McGinn, Colin (1991) *The Problem of Consciousness: Essays Toward a Resolution*. Blackwell.

Mehta, Neil (2013) ‘Is There a Phenomenological Argument for Higher-Order Representationalism?’, *Philosophical Studies* **164**: 357–70. doi:10.1007/s11098-012-9859-x.

Metzinger, Thomas (2020) ‘Minimal Phenomenal Experience: Meditation, Tonic Alertness, and the Phenomenology of ‘Pure’ Consciousness’, *Philosophy and the Mind Sciences* **1**: 7. doi:10.33735/phimisci.2020.I.46.

Michaelian, Kourken (2011) ‘Generative Memory’, *Philosophical Psychology* **24**: 323–42. doi:10.1080/09515089.2011.559623.

Michaelian, Kourken (2016) *Mental Time Travel: Episodic Memory and Our Knowledge of the Personal Past*. MIT Press.

Millière, Raphaël and Albert Newen (2022) ‘Selfless Memories’, *Erkenntnis*. doi:10.1007/s10670-022-00562-6.

Montague, Michelle (2016) *The Given: Experience and Its Content*. Oxford University Press.

Montague, Michelle (2017) ‘What Kind of Awareness Is Awareness of Awareness?’, *Grazer Philosophische Studien* **94**: 359–80. doi:10.1163/18756735-09403004.

Nagel, Thomas (1974) ‘What Is It Like to Be a Bat?’, *The Philosophical Review* **83**: 435–50. doi:10.2307/2183914.

Nida-Rümelin, Martine (2014) ‘Basic Intentionality, Primitive Awareness and Awareness of Oneself’, in Anne Reboul, ed., *Mind, Values, and Metaphysics*: 261–90. Springer International Publishing.

Nida-Rümelin, Martine (2017) ‘Self-Awareness’, *Review of Philosophy and Psychology* **8**: 55–82. doi:10.1007/s13164-016-0328-x.

Nida-Rümelin, Martine (2018) ‘The Experience Property Frame Work: A Misleading Paradigm’, *Synthese* **195**: 3361–87. doi:10.1007/s11229-016-1121-1.

O’Conaill, Donnchadh (2017) ‘Subjectivity and Mineness’, *Erkenntnis* 1–17. doi:10.1007/s10670-017-9960-9.

Peacocke, Christopher (2014) *The Mirror of the World: Subjects, Consciousness, and Self-Consciousness*. Oxford University Press.

Reade, Charles (1856) *‘It Is Never Too Late to Mend’: A Matter of Fact Romance*, vol. 1. Bernard Tauchnitz.

Robins, Sarah K (2016) ‘Misremembering’, *Philosophical Psychology* **29**: 432–47. doi:10.1080/09515089.2015.1113245.

Rosenthal, David M (1997) ‘A Theory of Consciousness’, in Ned Block, Owen J Flanagan and Guven Guzeldere, eds., *The Nature of Consciousness*. MIT Press.

Rosenthal, David M (2005) *Consciousness and Mind*. Oxford: Clarendon Press.

Russell, Bertrand (1926) ‘Philosophical Consequences of Relativity’.

Sartre, Jean-Paul (1943/1948) *Being and Nothingness*, Hazel E Barnes, trans. Philosophical Library.

Schear, Joseph K (2009) ‘Experience and Self-Consciousness’, *Philosophical Studies* **144**: 95–105. doi:10.1007/s11098-009-9381-y.

Sebastián, Miguel Ángel (2012) ‘Experiential Awareness: Do You Prefer ‘It’ to ‘Me’ ?’, *Philosophical Topics* **40**: 155–77.

Siewert, Charles (1998) *The Significance of Consciousness*. Princeton University Press.

Siewert, Charles (2013) ‘Phenomenality and Self-Consciousness’, in Uriah Kriegel, ed., *Phenomenal Intentionality*: 235–57. Oxford University Press.

Smith, David Woodruff (2004) *Mind World: Essays in Phenomenology and Ontology*. Cambridge University Press.

Snowdon, Paul (2010) ‘On the What-It-Is-Like-Ness of Experience’, *The Southern Journal of Philosophy* **48**: 8–27. doi:10.1111/j.2041-6962.2010.01003.x.

Sprigge, Timothy L S (1969) ‘The Privacy of Experience’, *Mind* **78**: 512–21.

Sprigge, Timothy L S (1971) ‘Final Causes’, *Proceedings of the Aristotelian Society, Supplementary Volumes* **45**: 149–92.

Stoljar, Daniel (2016) ‘The Semantics of ‘What It’s Like’ and the Nature of Consciousness’, *Mind* **125**: 1161–98. doi:10.1093/mind/fzv179.

Stoljar, Daniel (2021) ‘Is There a Persuasive Argument for an Inner Awareness Theory of Consciousness?’, *Erkenntnis*. doi:10.1007/s10670-021-00415-8.

Strawson, Galen (2013) ‘‘Self-intimation’’, *Phenomenology and the Cognitive Sciences* **14**: 1–31. doi:10.1007/s11097-013-9339-6.

Strawson, Galen (2022) ‘Self-Awareness: Acquaintance, Intentionality, Representation, Relation’, *Review of Philosophy and Psychology*. doi:10.1007/s13164-022-00639-9.

Sutherland, Norman Stuart, ed. (1989) *The International Dictionary of Psychology*. Continuum.

Thiel, Udo (2011) *The Early Modern Subject: Self-Consciousness and Personal Identity from Descartes to Hume*. Oxford University Press.

Thomasson, Amie L (2000) ‘After Brentano: A One-Level Theory of Consciousness’, *European Journal of Philosophy* **8**: 190–210.

Thompson, Evan (2007) *Mind in Life: Biology, Phenomenology, and the Sciences of Mind*. Harvard University Press.

Weinberg, Shelley (2016) *Consciousness in Locke*. Oxford University Press.

Wider, Kathleen (2006) ‘Emotion and Self-Consciousness’, in Uriah Kriegel and Kenneth Williford, eds., *Self-Representational Approaches to Consciousness*: 63–87. MIT Press.

Wittgenstein, Ludwig (1947/1980) *Remarks on the Philosophy of Psychology*, vol. 1, G E M Anscombe and G H von Wright, eds.; G E M Anscombe, trans. Blackwell.

Zahavi, Dan (1999) *Self-Awareness and Alterity: A Phenomenological Investigation*. Northwestern University Press.

Zahavi, Dan (2005) *Subjectivity and Selfhood: Investigating the First-Person Perspective*. MIT Press.

Zahavi, Dan (2014) *Self and Other: Exploring Subjectivity, Empathy, and Shame*. Oxford University Press.

Zahavi, Dan (2017) 'Thin, Thinner, Thinnest: Defining the Minimal Self', in Christoph Durt, Thomas Fuchs and Christian Tewes, eds., *Embodiment, Enaction, and Culture: Investigating the Constitution of the Shared World*: 193–200. MIT Press.

Zahavi, Dan (2018) 'Consciousness, Self-Consciousness, Selfhood: A Reply to Some Critics', *Review of Philosophy and Psychology* **9**: 703–18. doi:10.1007/s13164-018-0403-6.

Zahavi, Dan and Uriah Kriegel (2016) 'For-Me-Ness: What It Is and What It Is Not', in Daniel O Dahlstrom, Andreas Elpidorou and Walter Hopp, eds., *Philosophy of Mind and Phenomenology: Conceptual and Empirical Approaches*. Routledge.