Published in Review of Philosophy and Psychology https://doi.org/10.1007/s13164-018-0403-6

Please quote from published version.

Consciousness, Self-consciousness, Selfhood: A reply to some critics

Abstract
Review of Philosophy and Psychology has lately published a number of papers that in various ways take issue with and criticize my work on the link between consciousness, self-consciousness and selfhood. In the following contribution, I reply directly to this new set of objections and argue that while some of them highlight ambiguities in my (earlier) work that ought to be clarified, others can only be characterized as misreadings.

1. Setting the stage

My work on the connection between consciousness, self-consciousness and selfhood has spanned more than 20 years and has so far resulted in 3 monographs: *Self-awareness and Alterity* from 1999, *Subjectivity and Selfhood* from 2005, and *Self and Other* from 2014. In *Self-awareness and Alterity*, I defended the view that our experiential life is characterized by a form of self-consciousness that is more primitive and more fundamental than the reflective form of self-consciousness that one, for instance, finds exemplified in introspection.[1] In arguing for this claim, I drew on ideas from analytic

[1] In the following I will be using the terms 'self-consciousness' and 'self-awareness' interchangeably, as I have also done in previous writings. I don't think there is any consensus in the philosophical literature concerning their distinction. Perhaps it would have been preferable to simply stick to one of the terms, but since the authors I am going to discuss tend to use either one or the other, I will do so as well, and typically adopt the term of their choice when discussing their work.

philosophy of language, post-Kantian German philosophy and phenomenology. In subsequent writings that primarily engaged with debates in phenomenology and analytic philosophy of mind, I went on to argue that a theory of consciousness that wishes to take the subjective dimension of our experiential life seriously also needs to operate with a (minimal) notion of self. A defense of this particular claim can be found in chapter 5 of *Subjectivity and Selfhood*. My most comprehensive discussion of the relation between consciousness and selfhood to date is, however, to be found in the first part of *Self and Other*. There I outline an experience-based approach to selfhood according to which the self is a built-in feature of experiential life. I distinguish and contrast this notion of self with a variety of other notions, and then proceed to engage with and reply to a number of objections that has been or might be raised against this notion. The objections I consider include

1) the view that subjectivity rather than being a fundamental feature of experience is the outcome of a meta-cognitive operation involving conceptual and linguistic resources;

2) the claim that we as a result of the fundamental transparency of experience are never directly acquainted with our own experiences, neither reflectively nor pre-reflectively;

3) the claim that neuro- and psychopathology offer cases that constitute relevant exceptions to the claim that all experiential episodes are first-personal;

4) the view that subjectivity although being necessary for selfhood is not yet sufficient, and that a proper notion of self rather than simply being experiential must be located and situated within a space of normativity.

By engaging with and addressing these various objections, I not only sought to further clarify my own position, but also to adjust and modify it in light of the incoming criticism. Recently, however, *Review of Philosophy and Psychology* has published a new wave of critical papers. In what follows, I will not offer a summary of the arguments found in the three books mentioned above, but instead reply directly to these recent objections, which differ widely in range and character. Whereas some of them highlight ambiguities in my (earlier) work that ought to be addressed, others can only be characterized as misreadings that must be rectified.

2. For-me-ness, me-ness, and mineness

Let me start with an article that has already, and quite justifiably, had quite some traction. Guillot's point of departure is her dissatisfaction with the increasing proliferation of technical terms used to capture and describe the subjective character of consciousness; terms of art such as *subjectivity*, *for-me-ness*, *me-ishness*, *me-ness*, *myness*, *mineness*, *first-personal character*, *pre-reflective self-awareness*, *sense of self* and *sense of ownership*. According to Guillot, the indiscriminate use of these terms has introduced considerable confusion into the debate, since the terms are far from being conceptually equivalent (Guillot 2017: 26). To illustrate her point, and with the aim of offering a more fine-grained tripartite analysis of the subjective character of consciousness, Guillot proceeds to distinguish the notions of *for-me-ness*, *me-ness*, and *mineness*.

On Guillot's reading, *for-me-ness* is best understood as a label for the special awareness that the subject has of her own occurrent experience. On many construals, this special awareness comes about as a result of the experience possessing a special non-inferential inner awareness, such that in addition to being aware of its ordinary (external) object, it also has itself as (a secondary) object of

awareness. In short, it is by being aware of itself, that the experience possesses a subjective character that makes it be "for me" (Guillot 2017: 28-29).

Me-ness, by contrast, is when the subject of experience rather than simply being aware of the external object (and of the experience of the object) is also aware of herself. Me-ness, in short, is when the subject figures in experience as "an object of phenomenal awareness" (Guillot 2017: 35). It is at this point, that it becomes permissible to speak of phenomenal self-consciousness.

Mineness, finally, is when the experience is phenomenally given as mine. On this reading, mineness is the more complex notion, since it not only requires that the subject is aware of her experience, and aware of herself, but also aware of the possessive relation between herself and the experience, i.e., aware that she is owning the experience (Guillot 2017: 31, 43).

As Guillot then points out, there is *prima facie* a clear distinction to be drawn between an awareness of an experience, an awareness of an experiencer, and an awareness of the experience as owned by the experiencer, and it is neither obvious that the three notions are co-extensive nor that they stand in relations of mutual entailment (Guillot 2017: 32). Since the experience and the subject of experience are normally taken to be distinct particulars, one cannot without further ado argue that for-me-ness (an awareness of the experience) necessarily entails me-ness (an awareness of the experiencer) (Guillot 2017: 34), nor can one automatically move from the "familiar point that a subject is aware of her present experience in a way that others are not" to the claim that "what makes this 'way of being aware' special is that it encompasses […] the subject of awareness, the object of her awareness and their relation" (Guillot 2017: 34).

Guillot is surely right in saying that there is a difference between letting the self figure as the dative of experience, i.e., as the subject of experience, and letting it figure in the accusative as an object of awareness, and that this again is different from being explicitly aware of the possessive relation between the subject and its experience (Guillot 2017: 34-35). But have these differences

really been overlooked in the previous debate? One cannot infer from the fact that certain authors have used, say, the notions of *for-me-ness* and *mineness* interchangeably to the fact that their arguments thereby trade on unwarranted equations, since that would only be the case if the authors in question had defined the terms in the same way as Guillot. Since that is not the case, part of the dispute is merely verbal. Consider, for instance, Rowlands' use of the term *mineness*, where mineness is understood in adverbial terms as the way or mode in which the intentional objects of my experience are presented to me. When I experience objects, I have them *minely*, that is, the objects are given *for-me* (Rowlands 2015: 117). Or take what O'Conaill calls the deflationary reading of mineness, where the mineness of experience simply refers to the first-personal givenness of experience, to the fact that my experiences are given to me in a unique way (O'Conaill 2017: 3). In both cases, mineness is simply used as a synonym for for-me-ness. In my own work, I have also used the two terms interchangeably and synonymously (Zahavi 2011: 58-59, 2014: 19, 22), and repeatedly emphasized that they rather than referring to a distinctive quale are intended as labels for the distinct perspectival givenness or first-personal presence of experience.

What is of primary importance here, however, is not the choice of terms, but the distinctions and the question of their mutual interdependence. According to Guillot, *for-me-ness* is the most basic feature. It is a feature that is present and in place whenever and wherever there is phenomenal consciousness. She doesn't think, however, that for-me-ness necessarily entails me-ness and mineness. Moreover, whereas all three features on her account co-occur in the ordinary experiences of normal subjects (Guillot 2017: 45), there are pathological cases (to be discussed later) where they come apart. This view is clearly offered as a more deflationary alternative to the inflationary view Guillot ascribes to me. Here is the first oddity. Given how Guillot is defining me-ness and mineness, I would also dispute that for-me-ness entails either. Being aware of one's experiences when they occur is neither tantamount to being aware of oneself as a (secondary) object, nor equivalent to being

thematically aware of the experiences *as* one's own. In contrast to Guillot, I consequently do not think that the phenomenal character of a normal experience includes for-me-ness, me-ness, and mineness. Rather, I think they only co-occur quite rarely, namely when we reflect. On that background, one might turn the table and argue that I am defending a more deflationary account than Guillot.

The real issue of controversy, however, is arguably different. It concerns the question of how much one might pack into basic *for-me-ness*. Does for-me-ness entail some sense of self, some form of self-awareness? Guillot denies both. The fact that the experience manifests itself to me does not entail that I am thereby aware of myself in any way. Likewise, the fact that experiences are given first-personally to the subject does not entail that the experiencing subject is thereby self-aware (Guillot 2017: 48-49). In fact, to claim that for-me-ness involves self-awareness is, according to Guillot, to make the mistake of thinking that for-me-ness necessarily involves me-ness understood as an awareness of oneself as an object (Guillot 2017: 48).

What we see here, however, is rather Guillot making the same move once more. She arrives with her own definition of self-awareness and then assumes that other authors must be operating with the same definition, whereby it is then easy for her to accuse them of conceptual equivocations. This is particularly odd, given that Guillot herself, earlier in the article, acknowledged that self-awareness can have several meanings. It can mean the awareness that an experience has of itself, but it can also mean the awareness a self has of herself (Guillot 2017: 38). Guillot quickly side-lines the first option, and then focuses on the second. But what would happen if we stuck with the first? On this reading, sometimes called a *non-egological account of self-awareness*, we are dealing with self-awareness when consciousness is aware of or acquainted with itself. Being aware of an ongoing mental state (in contrast to simply being aware of an external object) consequently involves self-awareness. It exemplifies the mind's reflexive capacity; its ability to disclose or reveal itself to itself. If this is how self-awareness is defined, it should be obvious that it is something that by necessity characterizes for-

me-ness. After all, the latter term was introduced in order to capture the special awareness we have of our ongoing experiences. This way of discussing self-awareness is not only one that can be found in the history of philosophy (and more about that later), it is also one that has been popular in recent discussions between defenders and critics of higher-order representationalism concerning the difference between conscious and non-conscious mental states. Whereas higher-order representationalists have typically argued that "there is a strong intuitive sense that the consciousness of mental states is somehow reflexive or self-referential" (Rosenthal 2005: 36), and that the difference in question rests upon the presence or absence of a relevant meta-mental state, and therefore claimed that it is "the addition of the relevant meta-intentional self-awareness that transforms a nonconscious mental state into a conscious one" (Van Gulick 2000: 276), defenders of one-level alternatives have argued that conscious mental states possess an inherent pre-reflective self-awareness. In either case, however, the claim has been that there is a constitutive link between phenomenal consciousness and self-awareness. One might disagree of course, but then one should engage with the theories in question and either deny the difference between conscious and non-conscious mental states or offer an alternative account of their difference. To present matters as if defenders of such views have offered no arguments and are simply trading on conceptual equivocations is disingenuous.

Now, one possible retort would be to argue that even if for-me-ness does in fact entail a kind of self-consciousness, namely a kind of state-self-consciousness, where a mental state is aware of itself (which obviously doesn't necessarily entail that it is also aware of itself *as* a mental state), it remains the wrong kind of self-consciousness, it isn't subject-self-consciousness, it isn't a consciousness of self. And as Guillot insists, since the self and the experience are distinct particulars, an awareness of the latter does not automatically involve an awareness of the former (Guillot 2017: 34). Using this line of argument, Guillot explicitly takes issue with my position and objects to the proposal that the for-me-ness of experience entails that I am thereby aware of myself in any way

(Guillot 2017: 49). She further argues that my transition from the subjective character of experience to an awareness of the self in experience is a leap that I am only able to make because I fail to distinguish properly between subjective character understood as for-me-ness, as me-ness, and as mineness (Guillot 2017: 50).

At one point in the article, however, Guillot does acknowledge that her objection lacks purchase against those who opt for a deflationary or thin notion of self, and who accordingly deny that the self and the experience are distinct particulars, but she claims that the latter denial is controversial and in need of independent arguments (Guillot 2017: 37). That such arguments can be found in the literature and have been provided by the very authors Guillot is criticizing is, however, indisputable (cf. Strawson 2009, 2011, Zahavi 2005, 2011, 2014).

Among various arguments heralded by Strawson, we can, for instance, find the following: Since awareness is a property of a subject of awareness, and since awareness of a property of x is *ipso facto* awareness of x, any awareness, A1, of any awareness, A2, entails awareness of the subject of A2 (Strawson 2011: 280-281).

In my own work, I have sought to identify a thin and minimalist notion of self with the very subjectivity of experience. On this account, the self is present in experience, not as an additional experiential object or as an extra experiential ingredient (say as some kind of self-quale) but as the very first-personal givenness of experience. This notion of self is consequently not intended to refer to some persisting and abiding experience-transcendent self-entity, but is rather meant to target and capture the ineliminable subjective and perspectival givenness of consciousness. One immediate implication of this view is that one thereby no longer conceives of the self and the stream of experience as distinct particulars that in principle could be encountered in separation from each other.

I have contrasted and compared this experiential notion of self to other more robust notions of self.[2] I have discussed the relation between them, just as I have discussed the diachronic extension of the experiential self.[3]

Another implication of this view is that the very distinction between an egological and a non-egological self-awareness must eventually be called into question, which obviously makes the dialectics of the discussion a bit complicated. I think it is important to bear the distinction in mind for various systematic and historical reasons, but ultimately, I also think the distinction is motivated by a too narrow and too robust conception of selfhood. When realizing that a thinner and more deflationary notion is available, one should also come to see that no self-awareness strictly speaking is non-egological, but that the self might rather figure differently in different types of self-awareness. Being enthralled by a movie, enjoying a nip of Yellow Spot, feeling humiliated, and evaluating your life goals are four different experiences that are self-conscious and self-involving in different ways.

---

[2] Let me emphasize that the minimal account of self currently under consideration is in no way intended as an exhaustive account of selfhood. Indeed, the label minimal (or thin) is partially employed in order to highlight how limited the notion is and how much more has to be said in order to account for the full-fledged human self (Zahavi 2014: 50).

[3] There are obvious similarities between my own position and that of Strawson. One important difference, however, concerns the issue of diachronic persistency. Whereas Strawson has argued that each distinct experience has its own experiencer (Strawson 2009: 276), such that one and same human organism over its lifetime can be inhabited by a vast multitude of ontologically distinct short-lived selves, I have argued that the experiential self when understood as the ubiquitous first-personal character of experience is not identical or reducible to any specific experience, but rather something that can be shared by a multitude of changing experiences (Zahavi 2014: 72-77).

Of course, some might find these arguments unconvincing, but the proper task would then be to engage with them critically, rather than simply to ignore them. There might be a price to pay for insisting on a very tight constitutive link between self and experience, but the same holds true for any theory that insists that self and experience are distinct and independent particulars. There is something counterintuitive to the claim that the subject, self or I is entirely non-experiential, such that I would remain a self even if I were zombified and ceased having experiences, simply because a brain, a body or a living organism continued to exist. In fact, if self and experience are separated, it is unclear how self-experience would ever be possible, and how a certain object (be it a brain, a body, or a living organism) could ever be singled out and identified as myself.

It should not come as a surprise that some Buddhist philosophers after having insisted upon the difference between self and consciousness, and after having defined the self as a separate possessor or owner of consciousness, argue that there is no such thing and that it can easily be eliminated without loss (Garfield 2015: 106, 129, cf. Siderits, Thompson, Zahavi 2011). In response to my claim, that one can (and ought to) reject such a reified notion of self while still defending and retaining a more minimalist, experiential, notion, the reply has occasionally been that my minimalist account is so deflationary that it ends up being quite similar to a no-self doctrine (Albahari 2009: 80).[4]

In the end, it is hard not to conclude that a significant part of the discussion between no-selfers and pro-selfers is verbal rather than substantial and centred around the question of how deflationary a notion of self it makes sense to operate with. For a variety of reasons, I think it is appropriate to retain the reference to self, and that such a reference will be less misleading than any talk of experiences as being anonymous, impersonal or unowned, but what ultimately matters is not the self-

[4] For a reply, see Zahavi 2011.

label, but the fact that our experiential life is as such and from the beginning characterized by pre-reflective self-consciousness, subjectivity, and for-me-ness. When assessing my claim concerning the constitutive link between consciousness, self-consciousness, and selfhood, it is in any case important to keep my definition of self in mind, and not to assess the proposal on the basis of an imported and imposed definition of selfhood that differs from the one to which I subscribe.

3. Self-awareness, sense of ownership and thought insertion

Guillot's article appeared in a special issue edited by Farrell and McClelland. In their introduction, the two editors repeatedly refer to and adopt her tripartite distinction between for-me-ness, me-ness (which they render "me-ishness") and mineness. They provide their own slant on the definitions, however, and, for instance, describe the me-ness of experience in terms of the subject figuring in experience as a "thing-that-appears" (Farrell and McClelland 2017: 3). Given such a definition, the claim that me-ness is typically present in the ordinary experiences of normal subject seems even more implausible.

Guillot's distinctions are also appropriated by López-Silva, though he also decides to change the spelling of me-ness and opts for my-ness (2017: 6). López-Silva describes his contribution as supplementing Guillot's criticism. One significant difference between the two articles, however, is that López-Silva's article is devoted to an extensive attack on my work, which he repeatedly describes as flawed, unclear, misleading, unwarranted, implausible, untenable, confused, and replete with conceptual and phenomenological inaccuracies.

It will be impossible (and pointless) to address all of López-Silva's criticisms, but one central point of contention concerns the link between phenomenal consciousness and self-awareness. According to López-Silva, I conflate phenomenal character and subjective character, and equate

subjective character and self-awareness and consequently arrive at the unfounded and unwarranted claim that phenomenal consciousness involves self-awareness. As he writes (in one of many passages replete with typos): "One thing is to say that phenomenal conscious [sic] might lead to different degrees of self-awareness, but quite another is to propose that mere phenomenal access to experiences entails phenomenal awareness of the subject of experience As [sic] Zahavi does" (López-Silva 2017: 4). The fact that phenomenal consciousness and self-awareness are distinct is according to López-Silva evident from the fact that they can be instantiated independently. That this is so, can be gathered not only from the fact that we can imagine a possible world where zombie-like entities enjoy phenomenal experiences (can we really imagine that?) without enjoying self-awareness, but also from the fact that we can think of animals who possess phenomenal consciousness without possessing self-awareness, or at least, the kind of self-awareness that features centrally in my work (López-Silva 2017: 4). What kind of self-awareness is that? López-Silva writes that it is not easy to grasp what I have in mind, but that "self-awareness is commonly taken as a state that represents a (phenomenal) self" (López-Silva 2017: 4). Given such a definition, he then concludes that simply possessing phenomenal consciousness in the sense of having phenomenal access to a certain experience does not amount to self-awareness.

As should be obvious, this objection fails for the same reason as Guillot's objection failed. It ignores the fact that the term self-awareness (or self-consciousness) can be used and has been used to refer to the self-directed character of consciousness. Consider, for instance, Dharmakīrti's definition of self-awareness (*svasaṃvedana*) as the awareness that mental states have of themselves, or Śāntarakṣita's discussion of a non-intentional self-awareness that accompanies all object-cognition and makes consciousness conscious, or Dignāga's account of how self-awareness constitutes an immediate, non-conceptual access to our mental life (Kellner 2010: 204, 206, 228). But this way of talking about self-awareness is not restricted to Buddhist defenders of a no-self doctrine. There is a

scholarly debate about whether already Plato and Aristotle were committed to similar views (Plato 1961: 169d-170a, Aristotle 1984: 425b15-17, 1074b35, Gloy 1998, Caston 2002). If we move forward in time, one can find similar ideas in the work of the phenomenologists (for an overview see Zahavi 1999). Outside of phenomenology, the existence of a non-egological form of self-consciousness has been promoted by a group of German philosophers comprising Henrich, Cramer, Pothast, and Frank and known as the *Heidelberg School* (see, e.g., Henrich 1971, Frank 1991, Zahavi 1999). And more recently, and as already pointed out, figures in analytic philosophy of mind have discussed self-awareness, not simply in the context of an engagement with issues related to selfhood or personal identity, but also in connection with an investigation of the difference between conscious and non-conscious mental states. Consider, for example, the following quote from Frankfurt:

> An instance of exclusively primary and unreflexive consciousness would not be an instance of what we ordinarily think of as consciousness at all. For what would it be like to be conscious of something without being aware of this consciousness? It would mean having an experience with no awareness whatever of its occurrence. This would be, precisely, a case of unconscious experience. It appears, then, that being conscious is identical with being self-conscious. Consciousness *is* self-consciousness.
>
> The claim that waking consciousness is self-consciousness does not mean that consciousness is invariably dual in the sense that every instance of it involves both a primary awareness and another instance of consciousness which is somehow distinct and separable from the first and which has the first as its object. That would threaten an intolerably infinite proliferation of instances of consciousness. Rather, 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. It is like a source of light which,

in addition to illuminating whatever other things fall within its scope, renders itself visible as well (Frankfurt 1988: 161-162).

Given this non-egological definition of self-awareness and given that López-Silva repeatedly talks about how phenomenal consciousness is characterized by for-me-ness, which he defines as involving a direct phenomenal access to the very state one is undergoing – i.e., a non-inferential awareness of the experience one is having – it follows that phenomenal consciousness does involve self-awareness.

López-Silva's objection seems premised on his unfamiliarity with the discussions in question. He is, however, familiar with the distinction between reflective and pre-reflective self-awareness and he also knows that I have been defending the existence of the latter. What kind of arguments have I provided? Here López-Silva and Farrell and McClelland converge in their misreadings. According to López-Silva, my "main argumentation" amounts to the following: when asked what we are doing or experiencing, we can usually respond immediately, without the need for any inference or observation. The reason we can do so is then claimed to be due to the presence of pre-reflective self-consciousness (López-Silva 2017: 5). As for Farrell and McClelland, they write that "Zahavi suggests that accurate phenomenological description is 'the best argument to be found' for thinking that non-reflective experience is characterized by 'pre-reflective self-consciousness' of that very experience" (Farrell and McClelland 2017: 10). Whereas López-Silva refers to page 21 of *Subjectivity and Selfhood* in support of his reading, Farrell and McClelland base their interpretation on a passage taken from page 24 in the same book. The problem in both cases is that I am not presenting my own view. Whereas the passage from page 21 is taken from a place where I am expounding on Sartre's position in *Being and Nothingness*, the passage on page 24 is taken from my summary of an argumentative strategy pursued by other phenomenologists. As I write in immediate continuation of the passage

quoted by Farrell and McClelland, however, although I have quite a lot of sympathy for that kind of answer, I think a more theoretical argument is needed, which is what I then provide over the following six pages (Zahavi 2005: 24-29). The argument is a form of transcendental argument according to which reflective self-consciousness (the existence of which few people are prepared to deny) is seen as a form of object cognition, and where object cognition, which by definition involves a contrast between the subject and the object of cognition, is then shown not to be able to result in self-consciousness (facing either a vicious circle or an infinite regress) unless it is supported by a more fundamental non-objectifying pre-reflective form of self-consciousness. I will not recapitulate the argument here, which draws on and incorporates ideas from analytic philosophy of language (Wittgenstein, Castañeda, Shoemaker), post-Kantian German philosophy (Fichte, Henrich and Frank) and phenomenology, but it is an argument that figures centrally in my earlier book *Self-awareness and Alterity*, where it is also developed more extensively (Zahavi 1999: 14-38, 49-62).

Why ignore this central argument? In the case of Farrell and McClelland, I suspect the reason is as follows: I have always identified my position and theoretical background as phenomenological. For some, however, what this boils down to is the idea that theoretical claims ought to be based on careful experiential descriptions. But people can have conflicting phenomenological intuitions, and as Farrell and McClelland observe, such disagreements can be hard to resolve, since neither side can provide compelling reasons for why those on the other side are in error (Farrell and McClelland 2017: 10). Farrell and McClelland's suggestion of how to get out of this impasse is the following: We should recognize that our phenomenological intuitions are often embedded in a wider net of theoretical commitments and assumptions, and then relocate the discussion such as to focus on the arguments offered for and against those very commitments and assumptions.

This is a reasonable suggestion, but it isn't new. It is a view, I have been arguing for in all my books. The real disagreement is situated elsewhere. It concerns the interpretation of what

phenomenology and phenomenological philosophy amount to. I will not rehearse the arguments and textual evidence here (but see, for instance, Zahavi 2003, Zahavi 2007, Gallagher & Zahavi 2008, Zahavi 2017). It is enough to say that one has to be quite ignorant of phenomenology to think that it is primarily in the business of offering refined descriptions of inner experience.

Let me briefly return to López-Silva. The main claim of his article is that cases of thought insertion demonstrate that mineness (and ownership) is no essential part of phenomenal consciousness, and he therefore argues that the existence of thought insertion "undermines Zahavi's argumentation" (López-Silva 2017: 7).

Is this a novel challenge? Not really. In the past, Metzinger (2003), Lane (2012), Guillot (2017) and many others have made similar claims. Is it a convincing challenge? Not surprisingly, it all depends on what notions of mineness and ownership we are operating with.

According to what Guillot calls the simple account of thought insertion, such experiential episodes simply lack a sense of ownership (see Metzinger 2003: 334, 382, 445-446). This simple account has been criticized by people who insist that since the patient continues to have first-personal access to the inserted thoughts some sense of ownership is retained. Guillot now argues that this criticism is based on a failure to distinguish different notions of subjective character. She then proposes a more sophisticated version of the simple account, according to which episodes of thought insertion are experiences which lack mineness but retain for-me-ness and me-ness (Guillot 2017: 43) and then argues that the idea that a sense of ownership is retained simply because of the presence of first-person access "trades on an unwarranted equation between for-me-ness and mineness" (Guillot 2017: 47). A somewhat similar criticism can be found in Farrell and McClelland who write that many would argue that "an honest examination" of the relevant pathologies have shown the claim that a sense of ownership is retained to be false (Farrell and McClelland 2017: 6).

I don't know whether Farrell and McClelland are thereby implying that my own account is dishonest, but the underlying premise in both of these criticisms is that the notion *sense of ownership* is univocal, and it isn't. In previous writings, I have distinguished several notions of ownership and argued that one of these is retained in thought insertion (e.g. Grünbaum & Zahavi 2013, Zahavi 2014). Consider for instance the distinction, originally introduced by Albahari, between *personal ownership* and *perspectival ownership.* On her account, for a subject to own something in a perspectival sense is simply for the experience, thought, or action in question to present itself in a distinctive manner to the subject whose experience, thought, or action it is. The reason I can be said to own my thoughts or perceptions perspectivally is consequently because they appear to me in a manner that is different from how they can appear to anybody else (Albahari 2006: 53-54). Given this definition of ownership, episodes of thought-insertion do not lack ownership.[5]

[5] In earlier writings, I also suggested that episodes of thought insertion might involve a lack of sense of agency and a misattribution of agency to someone or something else (Zahavi 2005: 144). This proposal has been criticized by López-Silva and by Farrell and McClelland. López-Silva argues that the proposal that thought insertion merely involves a lack of a sense of agency is unable to distinguish the particular phenomenology characterizing episodes of thought insertions from the phenomenology characterizing other experiences with a disrupted sense of agency such as unbidden thoughts or obsessive thoughts (López-Silva 2017: 9). Likewise, Farrell and McClelland write that the idea that thought insertion might involve a lack of agency has been comprehensively rebutted since such an interpretation fails to differentiate cases of thought insertion from other cases where the subject supposedly lacks a sense of agency, such as cases of unbidden thoughts (Farrell and McClelland 2017: 6). What is odd about these criticisms is that they obviously misrepresent the proposal in question. The proposal was not that thought insertion is distinctly characterized by a lack of a sense

As for López-Silva, like Guillot he accepts that episodes of thought insertion are characterized by for-me-ness and my-ness (or sense of subjectivity) (López-Silva 2017: 8, 12, 15). But what is it then that they lack? They lack mineness, since patients experiencing episodes of thought insertion are no longer aware of the experiences *as* their own. Does this affect my proposal? Obviously not, since I am not defining mineness in the way López-Silva does, but rather using it as a synonym for for-me-ness.

As should have become clear by now, substantial parts of the recent criticism are based on a quite superficial reading of my work. Consider, for example, three passages on mineness that all

---

of agency. The proposal was that thought insertion is characterized by a lack of a sense of agency *and* by a misattribution of agency to someone or something else. I assume the latter is sufficient to distinguish thought insertion from unbidden thoughts or obsessive thoughts. Just for the record, I no longer think that thought insertion can be understood simply as a case involving a disorder of sense of agency. I think this is too simplistic an account. But the aim of my discussion was never to offer a positive account of thought insertion, but simply to rule out what I took to be a mistaken account, namely the claim that thought insertions provided prima facie evidence for the existence of phenomenal states that lack for-me-ness. Pathological experiences continue to be characterized by a subjective presence and a what-it-is-likeness that make them utterly unlike public objects that in principle are accessible in the same way to a plurality of subjects. Regardless of how alienated or distanced the patients feel vis-à-vis the experiences, the experiences do not manifest themselves entirely in the public domain – whatever the patients might be claiming. This is what most fundamentally makes the experiences first-personal, and this is why even these pathological experiences retain their for-me-ness. Since this is a claim that Guillot and López-Silva are also endorsing, one might again ask what exactly they are objecting to.

contradict interpretations of my view found in the critical papers discussed so far. In *Subjectivity and Selfhood*, I write that mineness "must be distinguished from any explicit I-consciousness. I am not (yet) confronted with a thematic or explicit awareness of the experience as being owned by or belonging to myself" (Zahavi 2005: 124). In *Self and Other*, I continue:

> More specifically, and contrary to what seems to be assumed by the critics, the mineness of experience is not some specific feeling or determinate quale. It is not a quality or datum of experience on a par with, say, the scent of crushed mint leaves or the taste of chocolate […]. Rather, the mineness refers to the distinct manner, or how, of experiencing. It refers to the first-personal presence of all my experiential content; it refers to the experiential perspectivalness of phenomenal consciousness. It refers to the fact that the experiences I am living through present themselves differently (but not necessarily better) to me than to anybody else (Zahavi 2014: 22).

After having discussed various pathological cases, I then conclude the chapter by writing

> I think we need to distinguish two different phenomenological claims: a minimalist one and a more robust one. On the minimalist reading, the for-me-ness and mineness of experience simply refer to the subjectivity of experience, to the fact that the experiences are pre-reflectively self-conscious and thereby present in a distinctly subjective manner, a manner that is not available to anybody else. I take it that this feature is preserved in all the cases I have discussed. On a slightly more robust reading, the for-me-ness and mineness of experience can refer to a sense of endorsement and self-familiarity, to the quality of 'warmth and intimacy' that William James claimed characterizes our own present thoughts (James

1890: 239). If this is what is meant by for-me-ness and mineness, I think it can be disturbed and perhaps even be completely absent (Zahavi 2014: 41).

Given that my main aim in discussing thought insertion was to show that the latter does not constitute a case where phenomenal consciousness lacks for-me-ness, it is hard to see what the disagreement is all about. As was also the case with Guillot, one might again wonder who defends the more inflationary view. According to López-Silva, in most normal cases, experiences involve for-me-ness, my-ness, and a sense of mineness. Given how he is defining the terms, I once again find such a claim implausible and too strong. When running to catch the bus, my focus is on the bus, and I am not aware of my experiences *as* being owned by *me*.

4. Transparency and the power of reflection

The special issue edited by Farrell and McClelland also contains a paper by Howell and Thompson entitled "Phenomenally Mine: In Search of the Subjective Character of Consciousness". In their paper, Howell and Thompson are not out to dispute that it is a metaphysical fact and conceptual truth that experiences are owned in the sense of necessarily being someone's experiences. Nor are they disputing that individuals have a privileged access to their own experiential states in the sense that they enjoy a special kind of first-person authority vis-à-vis these states, which they lack when it comes to the experiential states of others. Their target is *phenomenal me-ness*. As they concede, this is something that goes by many different names including for-me-ness, mineness, and experiential subjectivity, but for the sake of simplicity, they stick to me-ness. How do they define it? For Howell and Thompson, phenomenal me-ness cannot merely be a metaphysical feature of experience but must be something that contributes to the overall phenomenal character of the experience. In addition, not

everything that is phenomenally pervasive would qualify as a candidate for me-ness. In order to deserve its name, it would have to be something that in some sense were self-involving or self-referential (Howell and Thompson 2017: 106).

In the course of assessing various proposals, Howell and Thompson also take up my view and discuss two long quotes from *Subjectivity and Selfhood*. In the passages in question, I am asking the reader to consider a sequence of experiences like the following: First you see an orange, then you see an apple, and then you remember the apple. If we compare the initial situation with the final situation, neither the object nor the act type remains the same. I then ask whether this means that everything has changed and reply in the negative. There is something that remains the same between the first and the last experience, something that makes the relation between those two experiences utterly unlike the relation between two experiences belonging to distinct streams of consciousness. What is it that the two former experiences have in common? My claim is that they are both characterized by the same first-personal character, by the same for-me-ness. Is this a mere metaphysical claim, one with no impact on phenomenality? I deny this and argue that the for-me-ness in question refers to the distinct perspectival givenness or first-personal presence of experience (Zahavi 2005: 59).

Howell and Thompson's assessment of this argument is that it conflates an epistemic point with a phenomenal one (Howell and Thompson 2017: 113). Indeed, they explicitly question whether the epistemic feature "is phenomenally manifest, or […] constitutes a feature of phenomenal character" (Howell & Thompson 2017: 111).

I find this an odd claim to make. Consider the difference between my access to my own feeling of joy (as it is subjectively lived through) and the access I have to your feeling of joy (as it is displayed in your facial expression and verbal reports). Is that a difference with no phenomenal impact? Is there not an experiential, i.e., phenomenal, difference between feeling joyful yourself and observing somebody else's joy? In reply, however, Howell and Thompson might appeal to what they consider

a Sartrean line of argument; one they take to be needed if one is to preserve the transparency of experience. On this account, which they call the *Unreflective Naive Transparency Thesis*, our experiential life is completely oblivious to itself prior to reflection. Prior to reflection, our experiences make no appearance, there is no manifestation of subjectivity whatsoever. When pre-reflectively perceiving or recollecting a certain event everything one is aware of is completely objective. The fact that one is aware of it, makes no difference. It would be precisely the same if one weren't aware of it. Indeed, as Dretske argues in his defence of this kind of phenomenal externalism, "everything you are aware of would be the same if you were a zombie" (Dretske 2003: 1).

Whereas Dretske thinks this holds true even in the case of introspection, Howell and Thompson disagree. On their account, everything changes the moment we start to reflect. Through reflection, through the employment of various meta-cognitive operations, and through the imposition of a theoretical framework, we can lay claim to and appropriate the experience, and thereby bring it to givenness. The sense of self is precisely a product of rather than a condition for such appropriation (Howell and Thompson 2017: 123). By contrast, to suggest that experiences are always given, always characterized by me-ness, is in their view to fall prey to the so-called *refrigerator fallacy*, i.e., thinking that the light is always on, simply because it is always on whenever we open the door of the refrigerator (Howell and Thompson 2017: 114, cf. Schear 2009). Howell and Thompson agree, of course, that very few are inclined to deny, upon reflection, that their experiences are theirs, but on their view, there is nothing phenomenal that motivates this appropriation (Howell and Thompson 2017: 114). It is consequently not the conscious episode itself that provides part of the justification for the subsequent self-ascription of the episode in question. But if one denies that a reflective self-ascription such as "I am happy" is based on experiential evidence, if one insists that it entirely lacks experiential grounding and is in no way answerable to experiential facts, it is difficult to see how one can at the same time preserve and accommodate something like first-person authority.

A related difficulty confronts defenders of a Lichtenbergian anonymity or impersonality thesis (cf. O'Conaill 2017). If experience is at bottom conscious but impersonal, how can we then preserve the veracity of first-person reports? To say 'I have a headache' rather than 'There is a headache' would be to say too much, would be to falsify rather than to articulate the experience in question.

Indeed, this is precisely the reason why Guillot favours the claim that for-me-ness, me-ness and mineness are conjointly present in the ordinary experiences of normal subjects. On her account, it is their presence in simple experience that explains why the mere having of an experience makes me justified in judging that the experience happens, that it is mine, and that I exist. All three judgments are supported by something about the experience, something intrinsic to it. As Guillot writes "This I take to be at least a *prima facie* reason to think that we typically have *experiential* access to the experience, to ourselves, and to the fact that the experience is ours; or, in my terminology, that the phenomenal character of a normal experience includes for-me-ness, me-ness, and mineness" (Guillot 2017: 46).

Whereas I agree with the former part of her claim, I would dispute the latter part. As we have already seen, for Guillot, the self has to be given in the accusative, i.e., at the very least as a marginal object, if it is to be experientially present. I think this is a mistake, and that the self (i.e. subject) cannot for principled reasons in the first instance figure as an object, just as first-personal self-reference cannot be grounded on an identification of a certain object as oneself (Shoemaker 1968). In both cases, the problem is the same: In order for me to recognize a certain object as myself, I need to hold something true of it that I already know to be true of myself. To block the infinite regress, I would instead appeal to a prior non-objectifying and non-dual self-acquaintance as that which necessarily precedes and enables any awareness of oneself as an object and any thematic awareness of one's experiences *as* one's own.

Had our pre-reflective experiences really been as anonymous, impersonal, and invisible as Howell and Thompson suggest, it is hard to understand how we could ever start to target them in reflection, let alone appropriate them as ours, i.e., imbue them with first-personal character. This is also why the obvious reply to the *refrigerator objection* is that it leaves it quite mysterious how our reflective gaze or monitoring stance could possibly have that kind of illuminating effect. What is surprising is that Howell and Thompson seem to recognize this. They admit that a mental state cannot be imbued with me-ness simply as a result of being the object of a further mental state. Rather, if awareness of awareness is to give rise to me-ness, "the first order state" must already be "imbued with some phenomenally apparent quality of mine-ness" (Howell and Thompson 2017: 119). I think this is exactly right. This is precisely the view I was arguing for in *Self-awareness and Alterity*. It is difficult to see, however, why this shouldn't affect their own proposal. If there is no phenomenal me-ness on the pre-reflective level, it is quite unclear how such a sense can arise in and through reflection.[6]

[6] Howell and Thompson present their view as Sartrean view. They note that Sartre not only denied that pre-reflective consciousness involves any awareness of or reference to a self or an ego, but also take him to be defending the same unreflective naïve transparency view as themselves, according to which experience makes no appearance on the pre-reflective level (Howell and Thompson 2017: 109, 111-112). This is however a misinterpretation, especially if one also considers Sartre's position in *Being and Nothingness.* Not only did Sartre argue that self-consciousness is "*the only mode of existence which is possible for a consciousness of something*" (Sartre 2003: 10). He also argued that "pre-reflective consciousness is self-consciousness. It is this same notion of self which must be studied, for it defines the very being of consciousness" (Sartre 2003: 100). Indeed, as he points out in the chapter "The self and the circuit of selfness" in *Being and Nothingness,* consciousness is by no

5. Conclusion

It is one thing to introduce conceptual distinctions and to claim that such a regimentation is required if one is to avoid talking at cross purposes. Given the number of misinterpretations that my own work has generated, the need for some regimentation is apparent. It is something else, however, to accuse people of being confused and of employing invalid arguments that trade on conceptual conflations simply because they don't employ the distinctions in the same way as oneself.

One problem with the recent critical papers that have been published in *Review of Philosophy and Psychology* is that none of them references *Self and Other* which was published in 2014. Had they done so, I suspect many of the misunderstandings might have been avoided, since I in that book already anticipated and offered replies to most of the recent objections.

Another problem with a good part of the recent criticism is that it has had a far too myopic focus and failed to properly situate my arguments in the wide range and diversity of theoretical discussions and traditions that I have drawn and relied on. When arguing for the interdependence of consciousness, self-consciousness and selfhood, when arguing that phenomenal consciousness is characterized by for-me-ness and that this amounts to a minimal notion of self, I am not primarily making a descriptive claim such that people who disagreed with me could then be accused of having failed to attend sufficiently carefully to their own experiential life. Rather my claim is based on

---

means impersonal when pre-reflectively lived through. Rather it is characterized by a "fundamental selfness" (Sartre 2003: 127), precisely because of its ubiquitous self-consciousness. As I read Sartre, his proposal is that rather than starting with a preconceived notion of self, we should let our understanding of what it means to be a self, arise out of our analysis of self-consciousness.

various considerations concerning the difference between conscious and non-conscious mental states, concerning the nature of first-person authority and the possibility of first-personal self-reference, concerning the temporal unity of the stream of consciousness, the nature of epistemic asymmetry and social cognition, etc. In making these claims, I have drawn on discussions found not only in contemporary analytic philosophy of mind, but also in Kant, German Idealism, the Brentano school, neo-Kantianism, phenomenology, analytic philosophy of language and the Heidelberg School. If one wants to engage in a proper theoretical discussion of for-me-ness, one should engage with these more overarching theoretical discussions and not pretend that the issue can be settled simply by new conceptual stipulations or more detailed experiential descriptions.[7]

REFERENCES

Albahari, M. 2006. *Analytical Buddhism: The Two-Tiered Illusion of Self.* New York: Palgrave Macmillan.

Albahari, M. 2009. Witness-Consciousness: Its Definition, Appearance and Reality. *Journal of Consciousness Studies* 16/1: 62–84.

Aristotle. 1984. *The Complete Works of Aristotle I-II*, ed. J. Barnes. Princeton: Princeton University Press.

Caston, V. 2002. Aristotle on Consciousness. *Mind* 111/444: 751–815.

[7] Thanks to Adrian Alsmith, Felipe León, Raphaël Millière, Matthew Ratcliffe and Galen Strawson for comments to an earlier version of this reply.

Dretske, F. 2003. How Do You Know You Are Not a Zombie? In B. Gertler (ed.), *Privileged Access: Philosophical Accounts of Self-Knowledge.* Aldershot: Ashgate.

Farrell, J. & McClelland, T. 2017. Editorial: Consciousness and Inner Awareness. *Review of Philosophy and Psychology* 8: 1-22.

Frank, M. 1991. Fragmente einer Geschichte der Selbstbewußtseins-Theorie von Kant bis Sartre. In M. Frank (ed.), *Selbstbewußtseinstheorien von Fichte bis Sartre*. Frankfurt am Main: Suhrkamp.

Frankfurt, H. 1988. *The Importance of What We Care About: Philosophical Essays*. Cambridge: Cambridge University Press.

Gallagher, S., & Zahavi, D. 2008. *The Phenomenological Mind: An Introduction to Philosophy of Mind and Cognitive Science.* London: Routledge.

Garfield, J. L. 2015. *Engaging Buddhism: Why It Matters to Philosophy*. New York: Oxford University Press.

Gloy, K. 1998. *Bewusstseinstheorien. Zur Problematik und Problemgeschichte des Bewusstseins und Selbstbewusstseins.* Freiburg: Alber.

Grünbaum, T., & Zahavi, D. 2013. Varieties of Self-Awareness. In K. W. M. Fulford, M. Davies, R. Gipps, G. Graham, J. Sadler, G. Stanghellini, & T. Thornton (eds.), *The Oxford Handbook of*

Philosophy and Psychiatry. Oxford: Oxford University Press.

Guillot, M. 2017. *I Me Mine*: on a Confusion Concerning the Subjective Character of Experience. *Review of Philosophy and Psychology* 8, 23-53.

Henrich, D. 1971. Self-Consciousness, a Critical Introduction to a Theory. *Man and World* 4, 3-28.

Howell, R. J. & Thompson, B. 2017. Phenomenally Mine: In Search of the Subjective Character of Consciousness. *Review of Philosophy and Psychology* 8, 103-127.

Kellner, B. 2010. Self-Awareness (*svasaṃvedana)* in Dignāga's Pramāṇasamuccaya and -vṛtti: A Close Reading. *Journal of Indian Philosophy* 38: 203-231.

Lane, T. 2012. Toward an Explanatory Framework for Mental Ownership. *Phenomenology and the Cognitive Sciences* 11/2: 251–86.

López-Silva, P. 2017. Me and I Are Not Friends, Just Aquaintances [sic]: on Thought Insertion and Self-Awareness. *Review of Philosophy and Psychology*. Online first. https://doi.org/10.1007/s13164-017-0366-z

Metzinger, T. 2003. *Being No One.* Cambridge, MA: MIT Press.

O'Conaill, D. 2017. Subjectivity and Mineness. *Erkenntnis*, online first. https://doi.org/10.1007/s10670-017-9960-9

Plato. 1961. *The Collected Dialogues of Plato*, ed. E. Hamilton and H. Cairns. Princeton: Princeton University Press.

Rosenthal, D. 2005. *Consciousness and Mind*. Oxford: Oxford University Press.

Rowlands, M. 2015. Sartre on pre-reflective consciousness: The adverbial interpretation. In S. Miguens, G. Preyer, C. B. Morando (eds.), *Pre-reflective consciousness. Sartre and contemporary Philosophy of mind*. London: Routledge.

Sartre, J.-P. 2003. *Being and Nothingness: An Essay in Phenomenological Ontology*, trans. H. E. Barnes. London and New York: Routledge.

Schear, J. K. 2009. Experience and Self-Consciousness, *Philosophical Studies 144*(1), 95–105.

Shoemaker, S. 1968. Self-Reference and Self-Awareness. *Journal of Philosophy* 65, 556–579.

Siderits, M., Thompson, E., Zahavi, D. (eds.) 2011. *Self, No Self? Perspectives from Analytical, Phenomenological, and Indian Traditions.* Oxford: Oxford University Press.

Strawson, G. 2009. *Selves: An Essay in Revisionary Metaphysics*. Oxford: Oxford University Press.

Strawson, G. 2011. Radical self-awareness. In Siderits, M., Thompson, E., Zahavi, D. (eds.), *Self, No Self? Perspectives from Analytical, Phenomenological, and Indian Traditions.* Oxford: Oxford

University Press.

Van Gulick, R. 2000. Inward and Upward: Reflection, Introspection, and Self-Awareness. *Philosophical Topics* 28/2, 275–305.

Zahavi, D. 1999. *Self-Awareness and Alterity: A Phenomenological Investigation.* Evanston: Northwestern University Press.

Zahavi, D. 2003. *Husserl's Phenomenology.* Stanford: Stanford University Press.

Zahavi, D. 2005. *Subjectivity and Selfhood: Investigating the First-Person Perspective.* Cambridge, MA: The MIT Press.

Zahavi, D. 2007. Killing the Straw Man: Dennett and Phenomenology. *Phenomenology and the Cognitive Sciences* 6/1–2: 21–43.

Zahavi, D. 2011. The Experiential Self: Objections and Clarifications. In Siderits, M., Thompson, E., Zahavi, D. (eds.), *Self, No Self? Perspectives from Analytical, Phenomenological, & Indian Traditions*. Oxford: Oxford University Press.

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

Zahavi. D. 2017. *Husserl's Legacy: Phenomenology, Metaphysics, and Transcendental Philosophy.* Oxford: Oxford University Press.