

9 Does consciousness necessitate self-awareness?

Consciousness and self-awareness in Sartre's *The Transcendence of the Ego*

Daniel R. Rodríguez Navas

Introduction

Sartre's *The Transcendence of the Ego* has been misunderstood. The misunderstanding turns on the thesis that consciousness necessarily involves at least some form of self-awareness. According to an increasingly standard, but in my view incorrect, reading, that is one of Sartre's central claims in the essay. Consider, for instance, the following passage of Zahavi's *Subjectivity and Selfhood*:

Sartre, probably the best-known defender of a phenomenological theory of self-consciousness, considered consciousness to be essentially characterized by intentionality. He also claimed, however, that each intentional experience is characterized by self-consciousness [i.e., what I have called "self-awareness"]. Thus, Sartre took self-consciousness to constitute a necessary condition of being conscious of something.

(Zahavi 2008, p. 12)

In this chapter I will present my reasons for thinking that the standard reading is wrong, and explain why this should matter to us.

As I just pointed out, the standard reading differs from mine over the thesis that consciousness necessarily involves at least some form of self-awareness. Let's begin by clarifying this thesis and its philosophical implications.

Consciousness is a familiar but notoriously difficult phenomenon to define. We can think of it as the capacity to have experiences; or as the capacity to be in states that there is something that it is like to be in, to be in states that "have a phenomenology"; or as the capacity to be in mental states that involve a perspective or a point of view; or as the capacity to be aware of things in general (regardless of what it is that one is aware of, or of the way in which one is aware of it).

Though these ways of thinking about consciousness do not quite capture "the nature of" this enigmatic phenomenon, they can work as ostensive definitions that enable us to identify the thing that we're talking about, and that suffices for present purposes.

In the same way we can think of consciousness as the capacity to be aware of things in general, we may think of self-awareness as the capacity that a conscious creature may have to be aware of itself as such. But what is it for a creature to be aware of itself as such? Here, again, we come before an infamously difficult question, the difficulty lying in unpacking the “as such,” in giving an account of the capacity to have thoughts about the self or “the I.”

Perhaps because of this difficulty, self-awareness is sometimes treated as a *sui generis* phenomenon, a phenomenon that is elusive, hard, perhaps even impossible to explain, but which is familiar to everyone because everyone knows what it is like to have thoughts about themselves.

However, as in the case of consciousness, identifying the phenomenon of self-awareness does not require that there be a priorly available account of “the nature of” self-awareness or of the concept of “self” at our disposal. It is that capacity, whatever it is, that makes it possible for you to use words like “I,” “yo,” “je,” “ich,” etc.; the capacity to be in first-personal states; the capacity to be in I* states (in Castañeda’s sense). It is the capacity, in brief, that you have to be in mental states that are about yourself, and to understand all the while both that you are the subject (or bearer, or agent) of those states, and that those states are about you.

That thing, the self, whatever it may be, the thing that you associate with the “sense of ownership” or “feeling of mineness” that characterizes your mental life, the thing about which you think when you think of yourself, is the topic of Sartre’s *The Transcendence of the Ego*, as well as the topic of this chapter.¹

Let’s now return to the thesis on which the standard reading and the one that I propose differ. The thesis bears on the relation between the capacity to be aware of anything at all, and the capacity to be aware of oneself *as oneself*.

According to the standard reading, Sartre argues that the ability to be aware of anything at all is intrinsically bound up with the ability to be aware of oneself as oneself, bound up in such a way that one cannot have the former without having the latter. According to the reading I propose, Sartre argues for the opposite point: the ability to be aware of things in general does not necessarily involve the ability to be aware of oneself as oneself; it is possible for an agent to be capable of being aware of a range of things, but altogether lack the ability to be aware of itself as itself.

Now, note that the thesis that consciousness necessarily involves at least some form of self-awareness is, so formulated, ambiguous. It admits of at least two readings, a weak and a strong one, depending on whether it is heard as a claim about types of conscious entities or as a claim about conscious states.

Weak: All conscious entities are self-aware.

Strong: All conscious states are states of self-awareness.

The reason I call them weak and strong is, as you might suspect, that while the latter implies the former, the converse does not hold. That all conscious states are states of self-awareness implies that all conscious creatures are self-aware. But since it is possible for a creature to be self-aware without always being in states of self-awareness, the converse does not hold.

In the passage from Zahavi's *Subjectivity and Selfhood* cited above, he attributes the strong version of the thesis to Sartre:

Sartre . . . claimed, however, that each intentional experience is characterized by self-consciousness [i.e., what I am calling "self-awareness"]. Thus, Sartre took self-consciousness to constitute a necessary condition of being conscious of something.

(Zahavi 2008, p. 12)

Yet I will not argue only against Zahavi's version of the standard reading. I will not only argue, that is, that Sartre rejects the strong version of the thesis. I will argue that he rejects even the weak version of the thesis. On my reading, in *The Transcendence of the Ego*, in addition to defending the view that conscious states are not necessarily states of self-awareness, Sartre maintains the possibility that there be conscious entities that completely lack the capacity to be aware of themselves as such. In other words, he argues that, at least in principle, it is possible that there be creatures that are at no point of their lives aware of themselves as such.

There is a further source of unclarity in the thesis, as I have formulated it. The thesis says that consciousness necessarily involves some form of self-awareness. The source of unclarity lies in the implied idea that there are various forms or types of self-awareness. The best way to clarify this point and its stakes is perhaps by describing the mental lives of four different types of hypothetical creatures capable of being in conscious states:

- 1 *Narcissus*, or the hyper-self-aware creature: Narcissus is a creature that is always actually thinking about itself. It's not just that the "I think" can accompany all its representations; it's that it in fact does. Narcissus does not think: "This croissant is delicious," but thinks "I am thinking that this croissant is delicious." Its occurrent mental life is crowded by its own presence. It's the kind of creature of which we might correctly say that it has an Ego that is almost the size of its entire universe.
- 2 *Tinnitus*, or the semi-hyper-self-aware creature: This is a creature that is always thinking about itself. All of its states are partly about itself. Yet most of the time, and this is what distinguishes it from Narcissus, Tinnitus' attention is not focused on itself, but on whatever else it happens to be thinking about. Thus when Tinnitus thinks about the delicious croissant that it is eating, it's almost entirely absorbed by the deliciousness of the croissant. It is still aware of itself as such. But only in a very minimal sense. It is merely peripherally aware of itself, you

might say. Or you may invoke the phenomenological idea of an intentional horizon, and say that Tinnitus is always there for itself in its intentional horizon, sort of waving at itself in the distance, rather than as a prominent object in the intentional field. Or you may also say, if you like this way of speaking, that Tinnitus is always tacitly, or implicitly, or primitively aware of itself. However you may like to think about it, what is distinctive about Tinnitus' mental life is that all its states are states of self-awareness (even if, to insist, most of them are states of self-awareness in a minimal sense). So Tinnitus is always aware of itself as such, it is permanently present to itself or has an ongoing sense of self, but its self-awareness is much like constant background noise, like the constant ringing noise that people who suffer from tinnitus hear all the time. Hence its name.²

- 3 *Marie Antoinette*, or the occasionally-self-aware creature: This is a creature that can sometimes think about herself as such, but is not always doing it. In fact, during much of her waking life Marie Antoinette is not thinking about herself as such, at all. Not even minimally. When she goes to the gym to play squash, she focuses on the game. When she watches movies she gets so absorbed in them that she literally forgets herself. When she's listening to her friends talk, she really is listening to them, and not thinking about herself at all. She can of course always think about herself. And she does it as often as whatever activity she is engaged in requires it. But the rest of the time she really is not present to herself at all.

In fact, her thoughts about herself are in this sense much like her thoughts about the back of her neck: they are able to, but do not, accompany all her representations. At any given time she can add, to any given occurrent thought, an awareness of the back of her neck. But that her awareness of herself and of the back of her neck are able to accompany all her representations does not imply that they in fact do, not even in the minimal sense in which Tinnitus' awareness of itself as such accompanies all its occurrent mental states. This is of course not to say that her awareness of herself as such plays a functionally analogous role in her mental life to the role therein played by her awareness of the back of her neck; it is not to say that the explanation of her ability to add back-of-the-neck awareness to any of her thoughts is the same as the explanation of her ability to add an "I think" to all her thoughts. It is just to point out that, much like Marie Antoinette is only aware of the back of her neck on particular occasions, when that is relevant, and just as the rest of the time she does not have a single thought about it, so it is with her awareness of herself as such.

- 4 *Trump*, or the never-ever-self-aware creature: Trump has never had a single thought about itself as such. It simply lacks the capacity for it. Every now and then it has the unpleasant experience of seeing an image of a creature in the mirror, a creature that, unbeknownst to it,

happens to be itself. But that “unnoticed reflectivity” is as close as it ever gets to self-awareness. Trump just lacks a conception of itself as such that it could identify anything to.

Note that one of the principles of differentiation between these four types of creatures is a distinction between minimal and higher forms of self-awareness. For the higher type, we can think of what would be the reference of phrases like “fully first-personal states,” “states of explicit self-awareness,” “states of maximal self-awareness,” “fully reflective states,” and so forth. These are all meant to refer to the kind of state that one is in when having the type of thought that could only be expressed by using the first-person pronoun as the grammatical subject. Evidently, the use of such phrases goes in hand with the idea that there are other, weaker forms of self-awareness. Thus we find in the literature the notions of implicit or tacit self-awareness, minimal self-awareness, peripheral self-awareness, and so forth. I shall henceforth refer to these two broad forms of self-awareness, respectively, as higher forms of self-awareness and minimal forms of self-awareness. In using them, I shall be thinking of them as umbrella terms that can replace the terms of any distinction between types of self-awareness involving the claim that one type is primitive, or basic, or minimal, etc., relative to the other one.

Bearing in mind this distinction, we may characterize our four creatures as follows: (1) Narcissus is always highly self-aware; (2) Tinnitus is always at least minimally self-aware (the question of whether it is ever “highly” self-aware is left open); (3) Marie Antoinette is at least sometimes not self-aware at all, though at other times she is self-aware (whether in these cases she is minimally or highly self-aware is left open); and (4) Trump is never self-aware, not even minimally.

The descriptions of these types of creatures have purely heuristic value: my argument does not require that these creatures actually exist, or even that their existence be a live possibility.

In addition to helping bring out the idea that one may (as many do) draw a contrast between two putative forms of self-awareness, minimal and high, the idea of these four types of creatures can also help illustrate some of the implications of the thesis that consciousness necessarily involves at least some form of self-awareness.

There were, you may remember, two versions of that thesis. According to the weak version, all conscious *entities* are self-aware; according to the strong version, all conscious *states* are states of self-awareness. The weak version rules out the existence of creatures like Trump, but allows for the possibility of the other three types of creatures. The strong version, which Zahavi attributes to Sartre, in ruling out the existence of any conscious state that is not a state of self-awareness, also rules out the existence of creatures like Trump and Marie Antoinette. Since on my reading Sartre denies even the weak version of the thesis, it is also part of my reading that he allows for the possibility that even creatures like Trump exist.

Leaving historical and exegetical concerns aside for a moment: What is the philosophical import of this thesis? What could a creature like Trump actually be like? Why might it matter whether creatures of that kind might in principle exist?

Whether or not you think that creatures like Trump can in principle exist, whether or not, in other words, you think that to be conscious is ipso facto to be aware of oneself as such, will determine the kind of constraints that you think an account of self-awareness must satisfy.

If, for instance, you think that consciousness necessarily involves at least a minimal form of self-awareness (if you think that all conscious creatures are at least like Tinnitus), then you may think that all that is required in order to explain fully the capacity to be aware of oneself as such is an explanation of how higher forms of self-awareness can be attained on the basis of minimal self-awareness. Depending on how you think of minimal self-awareness, the specific form of your account will vary. The key point for our purposes is that the idea that there is a minimal form of self-awareness necessitated by consciousness suggests that the task of accounting for higher forms of self-awareness is the task of addressing the question: What cognitive resources must be at work in order for a creature that has a capacity of being in states of minimal self-awareness to be in higher states of self-awareness? Or: How can higher forms of self-awareness be attained on the basis of minimal forms of self-awareness?

If, on the other hand, you do not think that consciousness necessarily involves even minimal forms of self-awareness, then you'll think that even minimal forms of self-awareness stand in need of explanation. And, for someone who does not think that consciousness necessarily involves self-awareness, pending such an explanation of minimal forms of self-awareness, an account of higher forms of self-awareness on the basis of minimal forms of self-awareness fails to fulfill its explanatory goal. It amounts to saying that creatures that are capable of higher forms of self-awareness are capable thereof because they're capable of minimal forms of self-awareness, without explaining how they're capable of minimal forms of self-awareness. And indeed, there is a tendency in the literature to model minimal self-awareness on perception (on the idea of a sense of self, of a background feeling of mineness, of a non-conceptual first-personal content), and to think that being in higher states of self-awareness consists in coming to grasp, in a more robust, or direct, or explicit, or thematic way, what was an essentially first-personal content that was in any case always already available in a minimal form.

To put it in yet a different way, if the question is: *How is self-awareness possible?* the family of answers that go in hand with the idea that consciousness necessarily involves some form of self-awareness often takes a form like the following one: being aware of anything at all involves being minimally aware of oneself as such, and being aware of oneself as such in the more robust sense that is our ultimate explanatory target just involves transforming that

minimal form of self-awareness into a stronger form of self-awareness (by making explicit what was implicit or tacit, through the conceptual uptake of a non-conceptual first-personal content, by thematizing as the object of an intentional act something about which one was already minimally aware, by turning one's attention to what one was only peripherally aware of before). And yet, if self-awareness is not a necessary condition of consciousness, this answer is not a good one. For, while all the explanatory work is done by the notion of a minimal form of self-awareness, the latter is not one that is accounted for, but is merely postulated on the basis of the thesis that consciousness implies some form of self-awareness.

Thus, what lends interest to the exegetical controversy between the standard reading and the one that I propose is that, if the preceding argument is correct, and if the thesis that consciousness requires self-awareness is incorrect, the position that the standard reading attributes to Sartre is one that would commit him, and anyone who might endorse the view that he advocates in *The Transcendence of the Ego*, to a misconception of the form that an adequate explanation of self-awareness ought to take.

Sartre's conception of the relationship between consciousness and self-awareness

In the previous section, I described the distinction between what I have been calling the standard reading of Sartre's *The Transcendence of the Ego* and an alternative. As we saw, that difference turns on the thesis that self-awareness is a necessary condition of consciousness: if the standard reading is correct, Sartre endorsed a strong version of that thesis, according to which all conscious states are ipso facto states of self-awareness. My main aim in this chapter is to show that that reading is incorrect, and to bring into view the philosophical importance of the matter.

The standard reading often relies on a passage in the first part of the essay, in the course of which Sartre makes claims that seem to provide unequivocal evidence in its favor, claims to the effect, for instance, that "*the type of existence of consciousness is to be consciousness of itself*" (Sartre 2003, p. 97 (23–24); emphasis added). The purpose of this section is to show that, in spite of what an initial reading of that passage may suggest, the standard reading is textually problematic. I will present three arguments for this claim. First, the standard reading renders Sartre's goal in *The Transcendence of the Ego* at worst unintelligible, and at best trivial. Second, there are passages of Sartre's text in which he seems explicitly to reject the thesis that the standard reading ascribes to him, namely that all conscious states are states of self-awareness. Third, while, as I just noted, there are statements in Sartre's text that might on a first approximation seem to work as conclusive evidence of the correctness of the standard reading, the interpretation that is required in order for those claims indeed to serve as evidence for the standard reading is ruled out by other claims that he makes in the

context in which they occur. In addition, in the final part of the section, I will provide a positive reading of those claims, one that accounts for Sartre's motivation for making them while showing that, correctly understood, they do not lend support to the standard reading.

***The Ego is “outside” consciousness: Sartre’s goal in
The Transcendence of the Ego***

Sartre announces his goal in the first paragraph of the essay:

For the majority of philosophers the Ego is an “inhabitant” of consciousness. Some assert its formal presence within the “*Erlebnisse*” (the experiences), as an empty principle of unification. Others—for the most part psychologists—think they can identify its material presence, as a center of desires and acts, in each moment of our psychic life. We would like to show that the Ego is neither formally nor materially inside³ consciousness: it is outside, in the world; it is a being of the world, like the Ego of others.

(Sartre 2003, p. 93 (13))

What might it mean to say that the Ego is not inside consciousness, that it is outside, in the world? Within Sartre's conceptual framework, the notion of the Ego is the notion of the self; it is conceived as the agent of actions and as the bearer of states and qualities. He writes:

The I is the Ego as the unity of actions. The Me is the Ego as the unity of states and qualities. The distinction that one establishes between these two aspects of a single reality appears to us purely functional, not to say grammatical.

(Sartre 2003, p. 107 (44))

Now, in a footnote early on in the text, Sartre explains that he uses “consciousness” in the triple sense of the totality of consciousness, of monad, and of each single state or moment of consciousness (Sartre 2003, p. 95 (16)). Of these three senses of “consciousness,” the one that is primarily at work in the goal-defining claim is the sense of consciousness as a single state or moment of consciousness. That this is the case can be seen from the fact that Sartre's arguments for the view that the Ego is outside consciousness are presented in the form of answers to the question whether the “I think” does in fact accompany all our representations (Sartre 2003, p. 94 (15)). In addressing this question, Sartre is addressing the question whether the Ego is in fact present, in some form or another, in each of our representations, in each of our conscious states.

Accordingly, Sartre's goal of showing that the Ego is outside consciousness is tantamount to showing that the Ego is not in fact part of all our

conscious states. But insofar as the notion of the Ego is the notion of the self as the agent of actions and the bearer of states and qualities, the view that the Ego is not “inside” each of our conscious states amounts to the view that not all conscious states are states of self-awareness. Contrary to what the standard reading suggests, Sartre’s goal in the essay is therefore to defend a claim whose immediate implication is that consciousness does not require self-awareness (i.e. that we are like neither Narcissus or Tinnitus).

In the face of this objection, it is open to the standard reader to suggest that “Egoless” is not the same as “lacking any form of self-awareness.” The idea would be that, when Sartre talks about the Ego, what he has in mind are higher forms of self-awareness of the kind that you and I have when, say, we’re deliberating about important, potentially life-changing decisions. Correspondingly, when Sartre raises the question of the relation between the Ego and consciousness, what he has in mind would be a question about the relationship between being conscious and being capable of such higher forms of self-awareness. Thus—the standard reader’s story would go—Sartre’s view would be that, although consciousness necessitates a minimal form of self-awareness, it does not necessitate a higher form of self-awareness, and what he announces in the opening paragraph of the essay as his goal would be the defense of the second part of that view: it would be just the idea that consciousness does not necessitate higher forms of self-awareness. In terms of the forms of consciousness described in the previous section, the standard reader’s position would be that the goal announced by Sartre in the opening paragraph of his paper would be to show that our mental life is not like that of Narcissus, but it would also be part of Sartre’s view that our mental lives is like the mental life of Tinnitus.⁴

But while this move is indeed open to the partisan of the standard reading, it does not much advance her cause. For it is problematic on two counts. First, the view that not all conscious states involve higher forms of self-awareness (the view that our mental lives and indeed the mental lives of any conscious creature is not like that of Narcissus) is uncontroversial, so this interpretive strategy renders Sartre’s goal in *The Transcendence of the Ego* perfectly trivial. Second, on such an interpretive strategy, understanding Sartre’s goal in the essay requires mobilizing a distinction between higher and minimal forms of self-awareness, a distinction that, as I shall endeavor to show, is absent from the whole essay.

Although these considerations do not conclusively show the incorrectness of the standard reading, they have at least the following upshot: since the standard reading generates a number of exegetical puzzles, we can begin to see that the textual evidence for the standard reading is not as conclusive as it may originally have seemed. Endorsing the standard reading comes at a high exegetical cost, and an alternative interpretation that does not impose such an exegetical burden would seem by default preferable.

Consciousness does not require an I; furthermore, the I would be the death of consciousness: not all conscious states are states of self-awareness

As I mentioned in the previous section, in the first part of Sartre's essay, he argues that the Ego is "outside consciousness," and he does this by raising and addressing the question whether the "I think" does in fact accompany all of our representations.

By the end of section 1(A) of the essay, Sartre concludes that the "I think" does not in fact accompany all our representations:

[T]he phenomenological conception of consciousness renders the unifying and individualizing role of the I (i.e. the active face of the Ego) totally useless. On the contrary it is consciousness that makes possible the unity and the personality of my I (i.e. Ego). The transcendental I (the pure Ego), therefore, has no reason for being.

(Sartre 2003, p. 97 (23))

He then proceeds to argue that for the "I think" to accompany all our representations would be "the death of consciousness." Thus, by the end of section 1(A), he concludes:

all the results of phenomenology are threatened by ruin if the I is not, in the same way that the world, a relative existent [i.e. a transcendent one], that is to say, an object for consciousness.

(Sartre 2003, p. 99 (26))

Now, earlier in that first section, Sartre had listed the implications of those claims:

- 1 the transcendental field becomes impersonal, or, if one prefers, pre-personal, it is without I;
- 2 the I only appears on the level of humanity [of the person], and it is only a face of the Me, the active face [of the Me];
- 3 that the "I think" can accompany our representations because it appears on the surface of a unity toward the creation of which it has not contributed, and that it is on the contrary this prior unity that makes it possible;
- 4 that it is possible to wonder whether personality (even the abstract personality of an I) is a necessary accompaniment of a consciousness, or whether it is not possible to conceive of entirely impersonal consciousnesses.

(Sartre 2003, p. 96 (19))

These passages suggest that the standard reading is incorrect. For, as we saw, the standard reading is one that rules out the possibility of the existence of creatures like Trump, creatures who lack all forms of self-awareness

and who consequently have only a purely impersonal form of consciousness. But, when Sartre writes about consciousness as a pre-personal or impersonal field, when he says that it is without I, when he suggests that we ought to wonder whether it is not possible to consider entirely impersonal consciousnesses, he is suggesting not only that consciousness does not necessarily involve self-awareness in the sense that not all conscious states are states of self-awareness; he is in fact making the stronger claim that nothing in the nature of consciousness rules out the possibility that there be Trump-like creatures.

Of course these passages do not conclusively establish the incorrectness of the standard reading. As I already mentioned in passing, and as we will see in more detail in the next section, there are passages in which Sartre seems to endorse the thesis directly that consciousness necessarily involves self-awareness. In light of this, the conclusion we're entitled to draw from this section is only that even the passages that lend support to the standard reading are, taken on their own, inconclusive. No local claim in *The Transcendence of the Ego* is sufficient, on its own, to establish the correctness or incorrectness of the standard reading.

“Consciousness is consciousness of itself. . . .”

As I just mentioned, Sartre argues for the view that the presence of the I inside consciousness would “be harmful.” However, he makes a series of statements that seem to commit him to the view that consciousness essentially involves some form of self-awareness. Since this is the strongest textual evidence for the standard reading, it is worth considering them and the passage in which they occur in some detail:

- (i) But, furthermore, this superfluous I is harmful. If it existed it would tear consciousness from itself, it would divide it, it would slide in every consciousness like an opaque blade. The transcendental I is the death of consciousness.
- (ii) Indeed, the existence of consciousness is an absolute because consciousness is consciousness of itself. That is to say that the type of existence of consciousness is to be consciousness of itself. It has consciousness of itself insofar as it is consciousness of a transcendent object.
- (iii) Everything is therefore clear and lucid in consciousness: the object is in front of it with its characteristic opacity, but consciousness, it is purely and simply consciousness of being consciousness of this object, that is the law of its existence.
- (iv) One must add that this consciousness of consciousness—outside of the case of reflective consciousness, on which we'll insist later on—is not positional, that is to say that consciousness is not for itself its object. Its object is outside of it by nature and it is for this reason that in a single act it posits it and grasps it.

- (v) It does not know itself except as an absolute interiority. We will call such a consciousness: first degree consciousness or ir-reflective consciousness.
- (vi) We ask: is there a place for an I in this consciousness? The answer is clear: obviously not.
- (vii) Indeed, this I is neither the object (since *ex hypothesi* it is interior [i.e. it would be in consciousness]), nor is it of consciousness, since it is something for consciousness, not a translucent quality of consciousness, but, in a certain sense, its inhabitant. Indeed, the I, with its personality, is, however formal, however abstract we take it to be, like a center of opacity [. . .].
- (viii) Thus if one introduces this opacity inside consciousness, one thereby destroyed the very fertile definition [of consciousness] that we provided earlier, one fixes it, one obscures it, it is no longer a spontaneity, it carries within itself a germ of opacity.
(Sartre 2003, p. 98 (23–25))

What could Sartre mean by claims by (ii) and (iii), if not that the type of existence of consciousness, the law of its existence, is to require self-awareness? In the remainder of this section, I will provide three textual arguments and a positive reading of that claim in order to show that its intended meaning could not have been that consciousness requires self-awareness.

Consciousness necessarily involves self-awareness . . . therefore, if the I were immanent it would be the death of consciousness?

The first argument bears on the inconsistency between, on the one hand, the argumentative role of claims (ii) and (iii) to the effect that “consciousness is consciousness of itself” and, on the other hand, the standard reading, according to which they are to be read as the idea that consciousness necessarily involves some form of self-awareness.

The idea is simple: Sartre makes these claims as part of an argument whose conclusion is that the I cannot be inside of, or inherent to, part of, consciousness. Yet if we read the claim that consciousness is consciousness of itself as the claim that consciousness always involves self-awareness, we are saddled with the paradox that, in order to argue that the I is not inherent to consciousness, in order to argue that the I is not part of the content of every conscious state, Sartre would be relying on the claim that consciousness necessarily involves self-awareness, that the self is part of the content of every conscious state.

Consciousness necessarily involves self-awareness . . . of a non-positional kind?

On (iv), Sartre argues that consciousness is non-positionally consciousness of itself. He writes, once again:

[T]his consciousness of consciousness . . . is not positional, that is to say that consciousness is not for itself its object. Its object is outside of it by nature and it is for this reason that in a single act it posits it and grasps it.

(Sartre 2003, p. 98 (24))

But note that, if the “consciousness of” in “consciousness of itself” is the “of” of “intentionality,” then the idea of a non-positional consciousness, the idea of a consciousness of something that is not its object, is incoherent. This is not to deny that classical phenomenology contains the resources to accommodate the idea of peripheral forms of awareness, where an agent is simultaneously aware of various things at the same time, albeit of some more intensely or attentively than of some others. But in these cases, the objects of awareness are still, as such, *objects* of awareness; they are still within the intentional horizon of the relevant mental episodes. Furthermore, to the extent that they are objects of awareness, they are posited as objects. Indeed if, as Sartre says in the same statement, *positing* and *grasping* an object are a “single act,” then where there is no positing of an object, there is no grasping it. Thus, that the “consciousness of itself” of which he is speaking is non-positional means that it is a “consciousness of itself” that does not “grasp itself.” But a “consciousness of itself” that does not grasp itself is a “consciousness of itself” that is not aware of itself. Consequently, the claim that consciousness is non-positionally conscious of itself cannot be interpreted in the way that the standard reading requires, as the claim that consciousness necessarily involves self-awareness.

Ir-reflective . . . self-awareness?

On (v), Sartre characterizes the type of consciousness of which he says that it is *consciousness of itself* as *ir-reflective*. Now, on the standard reading, the claim that consciousness is consciousness of itself is the claim that consciousness involves self-awareness. But a consciousness that involves self-awareness just is a reflective consciousness. So the standard reading renders Sartre’s ir-reflectivity claim absurd.

To put it differently, the standard reading, which requires interpreting claims (ii) and (iii) as claims to the effect that consciousness necessarily involves some form of self-awareness, renders the ir-reflectivity claim (v) unintelligible. For the idea that consciousness necessarily involves even a minimal form of self-awareness is the idea that consciousness necessarily involves not only the minimal type of reflectivity that an agent can have when having a thought about itself without recognizing itself as such (e.g., the kind of reflectivity that animals that fail the mirror test have in seeing themselves in the mirror), but even the much more substantive kind of reflectivity at work in states that involves the recognition of oneself as such.

On the reading that I advocate, on the other hand, the ir-reflectivity claim can be easily accounted for. When Sartre writes about ir-reflective forms of consciousness, his point is exactly what it seems to be: that such states do not involve even a minimal form of self-awareness; they do not involve any form of reflectivity at all.⁵

What does Sartre mean when he says that consciousness is consciousness of itself?

But how can we reconcile the claim that the type of existence of consciousness is to be conscious of itself with the idea that there are ir-reflective, non-positional conscious states? If the claim that the type of existence of consciousness is to be consciousness of itself does not mean that consciousness is necessarily aware of itself, then what does it mean?

Let's begin by trying to understand the argument that Sartre is sketching in that context. The upshot of the argument is meant to be that, if the I is "inside" consciousness, if it is part of consciousness in any way other than as its object, then it would be harmful. The argument itself is formulated in terms of the metaphor of light. Consciousness is conceived like light that shines on its intentional objects. It is thus described as essentially luminous or translucent, and its objects are described as opaque. The idea is that the I cannot be part of consciousness because "however formal, however abstract we may suppose it to be, in order for it to be a part of consciousness it would have to act like the center of opacity" (see (vii) above) and thus the light of consciousness would not be able to shine on to its objects: if being aware of transcendent objects required being aware of representations thereof and of ourselves as thinkers of those representations, we would not be able to reach "all the way out" to the objects themselves. The light would shine on our opaque representations of ourselves and the contents of our minds, and by the same token would fail to reach transcendent objects.

As an argument, this is of course too metaphorical to be particularly helpful. But it does give us an exegetical clue for understanding what Sartre means when he says that the type of existence of consciousness is to be conscious of itself. That and similar claims are meant by Sartre to be not only consistent with, but consequent upon, the fact that "everything is clear and lucid inside consciousness." They are meant to be, in contemporary terms, consequent upon the transparency of consciousness, understood as the view that consciousness is first and foremost world-directed, that mental content is not a primary object of consciousness, that being aware of the world and worldly things does not in any way require being aware of mental content as such. This works as an exegetical clue insofar as it sets constraints on the kind of interpretation that can be given to the claim that the type of existence of consciousness is to be consciousness of itself: it rules out any interpretation that would suggest that being conscious requires being

aware of the contents of consciousness as such, or of the self as being aware of those contents.

A second clue to understanding Sartre's claim that the type of existence of consciousness is to be conscious of itself is the ambiguity of the term "consciousness." As we saw, there is a distinction to be drawn between consciousness in the sense of "conscious state or episode," in the sense of "conscious entity," and in the sense of "field of consciousness" (Sartre 2003, p. 94 (15)). As we have also seen, the interpretation according to which Sartre's idea is that any state of consciousness involves awareness of that state is ruled out (it contradicts the ir-reflectivity, the non-positionality, and the transparency claims). However, the interpretation according to which Sartre's idea is that conscious entities are in some sense conscious of "the field of consciousness" is not ruled out. The only constraint is that the sense in which it is conscious of the field of consciousness not be the sense in which they are conscious of the intentional, posited, objects of those states. What sense of consciousness is that?

It helps to bring in and develop the spatial metaphor. As I've been saying, we can think of consciousness as a field.⁶ There's nothing peculiar in the idea of perceptual experiences in which one is not conscious of space as such, in which one is not actively, or explicitly, or thematically thinking about space, even though one is, in a very different sense, quite conscious or aware of space. Consider, for instance, the following situation. While you wait for the pedestrian light to signal that it is your turn to cross the street, you go over the list of things you need to buy at the grocery store. The light changes. You begin to cross, mindful of the environing traffic, mindful of the other pedestrians walking in your direction, careful to avoid stepping on the tiny little beast (is that a dog?) that someone's proudly walking on a leash. All the while you continue to go over your shopping list. All the while you are conscious of your environing space. And yet, at no point do you have a single thought about space as such, even about your environing space. You see things to be avoided, pathways to walk through, and so forth, but you don't have a single thought about space as such at all. In fact, even if you entirely lacked, as the little dog that you crossed paths with probably does, the capacity to think of space *as such*, it would still make sense to say of you, as it does of the little dog, that you were "conscious" or "aware of space." Although crucially, the sense in which you and the little dog could be said to be "conscious of" space in such a situation is very different from the sense in which you can be said to be "conscious of," say, the taste of the coffee that you're actively savoring right now. The latter is the "of" of intentionality, of awareness of whatever object one is occurrently thinking about (in broad senses of "object" and "thinking"). The former isn't.

As it is with space in these cases, we may think, so it is with consciousness. In being conscious of an object, consciousness is conscious of what is within "its field." On these grounds it can be said, speaking very loosely, to be consciousness of itself: it is conscious of the objects present in the field of

consciousness, and in that sense may be said to be “conscious” of “the field of consciousness.” But it does not grasp itself; it is not conscious of itself thematically, or positionally, or as an object; it is not reflectively conscious of itself; the relevant conscious individual is not aware of itself.

On this way of hearing the claim that consciousness is consciousness of itself, the claim is not a very informative one. In the context of an argument for the transparency of consciousness, Sartre’s point is merely that to be conscious is to be conscious of what is given “in” consciousness. The point is not that conscious creatures, insofar as they are conscious of things in the world, can be said to be conscious of what is “in” their minds, and thereby conscious “of their minds.” To insist, in the context of an argument whose conclusion is that there is “no place for the I inside consciousness” that the presence of the I inside consciousness would be, as Sartre dramatically puts it, the death of consciousness, the emphasis is rather meant to be on the fact that conscious creatures are conscious, positionally conscious, of the things of which they are conscious, and that they can only be said to be conscious of themselves insofar as they are “conscious of” what is given “in” consciousness, in a sense of “conscious of” that is not, once again, that of intentionality, and which therefore does not warrant the idea that consciousness always involves self-awareness.

Conclusion: Sartre on consciousness and self-awareness

In the last section, I provided a series of arguments against the view, characteristic of the standard reading, that it is part of Sartre’s position in *The Transcendence of the Ego* that all conscious states involve at least a minimal form of self-awareness.

First, I argued that Sartre’s project in that essay, the project of showing that the Ego is outside consciousness, is the project of showing that consciousness does not necessarily involve self-awareness. In the face of this fact, it is open to the standard reader to argue that there is a distinction between minimal (Egoless) and higher (Ego-involving) forms of self-awareness, and to argue that Sartre’s explicit goal is only to defend the view that consciousness does not necessarily involve *higher* forms of self-awareness (that not all conscious creatures are like Narcissus), although it is also part of his view that consciousness essentially involves minimal forms thereof (that all conscious creatures are like Tinnitus). The problem with this way of reading Sartre’s essay is twofold: first, since the view that consciousness essentially involves higher forms of self-awareness (the view that all conscious creatures are like Narcissus) is implausible, this way of reading Sartre renders his goal in the essay trivial; second, it involves drawing a distinction between higher and lower forms of self-awareness that is absent from Sartre’s text.

The partisan of the standard reading might attempt to insist that, contrary to what I just suggested, this distinction is not absent from Sartre's text. Contrary to what I suggest, Sartre's idea of an Egoless consciousness would not be the idea of a consciousness that lacks self-awareness, but the idea of a consciousness that lacks a higher form of self-awareness. The reference to an ir-reflective consciousness whose "type of existence" is to be "*consciousness of itself being consciousness of an object*" would be precisely a reference to an Egoless form of consciousness that involves a minimal type of self-awareness (like the type of consciousness of Tinnitus). Yet this interpretation, I have argued, is also rendered implausible by the texts. The reason is, once again, twofold: first, in the course of arguing that the I is outside consciousness, Sartre explains that one of the implications of that claim is that one cannot rule out the possibility of purely impersonal forms of consciousness that do not even have the abstract personality of an I; second, considered in context, Sartre's claim that the type of existence of consciousness is to be consciousness of itself cannot be interpreted as the claim that consciousness involves self-awareness, even of a minimal kind, since Sartre uses that claim as a premise for an argument to the effect that the I, however formal, cannot be an intrinsic part of consciousness, and since he characterizes this most basic type of consciousness as both a non-positional consciousness of itself and as ir-reflective.

Finally, I have provided a positive interpretation of Sartre's claim to the effect that consciousness is consciousness of itself. In making that claim, Sartre merely means to be acknowledging the fact that being conscious of anything at all involves being conscious of what is given "in" consciousness, much like one can say of any conscious creature capable of perception and motion that it is conscious of the space it moves in. In neither case does the relevant sense of "being conscious of" imply that the relevant agent is conscious of what it is thus conscious of (consciousness/space) *as such*. Sartre's point, to insist, is not that consciousness necessarily involves some form of self-awareness.⁷

*

In the first section, I suggested that Sartre's thesis, according to which consciousness does not necessarily involve self-awareness, has an important philosophical implication: self-awareness is a cognitive achievement. So long as the possibility for purely impersonal forms of consciousness (like the form of consciousness characteristic of Trump) is not ruled out, the starting point of an account of higher forms of self-awareness must be the capacity to be aware of objects in general, not the capacity to be in conscious states that involve minimal forms of self-awareness. The reason is that so long as such a possibility (i.e., of Trump-like creatures) is not ruled out, even minimal forms of self-awareness, if such there be, stand in just as much need of explanation as higher forms of self-awareness do. Consequently,

their possession by conscious creatures cannot be presupposed or taken for granted within the context of an explanation of self-awareness. So long as the possibility of impersonal forms of consciousness is not ruled out, accounting for self-awareness in terms of more basic or minimal forms of self-awareness amounts to displacing rather than accomplishing the important explanatory task, the task of accounting for the first-personal or “self-involving” character of states of self-awareness.

*

There are three important implications of the view that self-awareness is a cognitive achievement, implications that it would be the next major task to explore and develop. First, the idea that there is a minimal form of self-awareness that is intrinsic to consciousness can suggest a conception of self-awareness as a unitary phenomenon, a picture according to which for any conscious state to be a state of self-awareness is for it to involve awareness of a certain type of object, “a self,” or for it to involve the actualization of a single, general capacity for self-awareness. On the contrary, the view that self-awareness is a cognitive achievement suggests a conception of self-awareness as a non-unitary phenomenon, a picture according to which there may be a range of distinct forms of self-awareness (e.g., proprioceptive self-awareness, “practical” self-awareness, psychological self-awareness), and which requires that, to the extent that it is possible, each of these forms of self-awareness be given an independent account.⁸ A second implication, as Sartre himself notes in the conclusion of his essay, is the exclusion of a conception of our ability to represent minds that leads to solipsism (Sartre 2003, p. 130 (84–85)): the idea that consciousness necessarily involves at least a minimal form of self-awareness suggests that our awareness of other minded creatures as such has the form of the recognition that they are like us, the recognition that, in addition to having observable bodies, they also have unobservable minds. On the contrary, the idea that consciousness does not necessitate self-awareness leaves open the possibility that there be no priority between our self-understanding and our understanding of others as thinking creatures; it leaves open the possibility that the cognitive resources required to represent others as minded and to be aware of oneself as the subject of psychological states are the same. Thirdly, the view that self-awareness is a cognitive achievement also has important epistemological implications: it opens up the possibility that our awareness of much of our mental lives does not involve the type of immediacy and infallibility that a traditional, Cartesian conception of the mind would suggest, the possibility that, as Sartre himself suggests, our knowledge of ourselves is much more similar to our knowledge of others than has been traditionally accepted. These are not issues that can be pursued here, but further, in-depth work on Sartre’s views about consciousness and self-awareness has the potential to illuminate them.

Notes

- 1 Note, lastly, that I just switched from talking about our topic as a capacity, to talking about it as a “thing” of sorts; from the capacity that I have been calling “self-awareness,” to the thing that I have been calling “self.” Which one is our topic, properly speaking? I am partial to thinking primarily in terms of the capacities that certain organisms have, rather than to think of “selves” or “Egos.” The latter, it seems to me, are not distinct types of things, but to borrow an expression from Castañeda, certain types of things regarded under a certain guise. The reason that I am partial to thinking in terms of capacities is that I think it can keep us clear of various metaphysical and philosophical pitfalls (think, for instance, of Anscombe’s claim that “I” does not refer). Sartre, however, tends not to think in terms of capacities but in terms of entities, not in terms of self-awareness but in terms of the Ego (or the *cogito*, the I and the me, consciousness and the self); he writes of “things” or “entities.” Since, accordingly, I cannot fully avoid this reifying way of speaking, I will continue to switch back and forth between talk of consciousness and self-awareness as capacities and to talk about them as things.
- 2 There are two ways in which we may conceive of the clinical condition called “tinnitus.” Since which one of them we favor determines how we conceive this case, it is important to distinguish between them and to indicate which one should serve as the model here. We may think that people who suffer from tinnitus are always hearing the ringing in their ear, not just in the sense that whenever they pay attention to it they will be able to hear it, but also in the sense that it is impossible for them not to hear it, even if they choose not to pay attention to it. On the alternative model, the condition is such that people can hear it whenever they pay attention to it, but whenever they are sufficiently absorbed in any activity or whenever there are sufficiently strong competing aural stimuli, the ringing sound simply disappears completely from the person’s field of consciousness. The way of thinking about the condition which is meant to serve as model for the type of creature in question is the former case, that is, the one in which the person is always conscious of a ringing sound.
- 3 The emphasis is Sartre’s.
- 4 This revised version of the standard reading corresponds to the position defended (and attributed to Sartre) by Zahavi in *Self-Awareness and Alterity*, especially chapter 2 (Zahavi, 1999).
- 5 The partisan of the standard reading might be tempted to respond that the ir-reflectivity claim is about the question whether an *act* of reflection is involved in the type of consciousness that Sartre is discussing, rather than about the question whether such a type of consciousness has *the formal property* of reflectivity. In this case, there would be no incoherence in the idea of a form of consciousness that is reflective in the sense that it involves *self-awareness*, but ir-reflective in that it does not involve an *act* of reflection. The problem with this response is that it rests on the assumption that a state of consciousness can be reflective in the formal sense without involving an act of reflection. But, as we shall see in the next section, this is ruled out by Sartre’s conception of consciousness, and more specifically by the thesis that consciousness is “translucid.”
- 6 In the technical phenomenological sense, the concept of consciousness as a field is the concept of consciousness as the region of being that is the residuum of

the phenomenological reduction, and which constitutes the phenomenologist's field of investigation. See, for instance, Husserl (1982 §33, p. 66 [59] and §50, pp. 112–114 [93–95]).

- 7 A supplementary argument, but one whose exposition would require a paper of its own, is that the conception of consciousness that the standard reading attributes to Sartre can be shown to be much closer to Husserl's 1913s views than to his 1901 views in this regard. Since one of Sartre's goals is to defend the 1901 view against the 1913 view, this also suggests that the standard reading misrepresents Sartre's views about consciousness. See, in particular, Husserl (1982, §57).
- 8 Thus, for instance, in the second part of the essay, where Sartre describes the constitution of the Ego (i.e., how Ego-involving thoughts are possible), he provides relatively independent accounts of the kind of awareness that we have of our actions and thoughts on the one hand, and of our emotions and dispositions on the other.

References

- Husserl, E. (1982). *Ideas Pertaining to a Pure Phenomenology and to a Phenomenological Philosophy*, vol. I. (Kersten, E., trans.). Dordrecht: Kluwer.
- Sartre, J.-P. (2003). *La Transcendance de l'ego*. Paris: VRIN.
- Zahavi, D. (1999). *Self-Awareness and Alterity: A Phenomenological Investigation*. Evanston, IL: Northwestern University Press.
- Zahavi, D. (2008). *Subjectivity and Selfhood: Investigating the First-person Perspective*. Cambridge, MA: MIT Press.