

# A DEFENSE OF INNER AWARENESS: THE MEMORY ARGUMENT REVISITED

Anna Giustina

Forthcoming in *Review of Philosophy and Psychology*

**Abstract.** The psychological reality of an inner awareness built into conscious experience has traditionally been a central element of philosophy of consciousness, from Aristotle, to Descartes, Brentano, the phenomenological tradition, and early and contemporary analytic philosophy. Its existence, however, has recently been called into question, especially by defenders of so-called transparency of experience and first-order representationalists about phenomenal consciousness. In this paper, I put forward a defense of inner awareness based on an argument from memory. Roughly, the idea is that since we can only recall something if we were aware of it at the time of its occurrence, and since we can recall our own experiences, we must be aware of our own experiences at the time of their occurrence. The argument is far from new: it goes back to the Buddhist tradition and has been revived more recently in Buddhist Scholarship but also in contemporary analytic philosophy of mind, in particular by Uriah Kriegel. However, I believe that, since it is the best extant argument for inner awareness, it deserves more extensive treatment. My goal is to strengthen the memory argument by (i) making some conceptual distinctions as to the exact thesis about inner awareness the argument is supposed to support, (ii) considering different ways the argument may be reconstructed depending on the exact thesis to be supported, and (iii) defending the argument from a new objection, raised very recently by Daniel Stoljar.

## 1. Inner awareness and its theses

Inner awareness is awareness of one's own current conscious experience. Awareness of experience may take different forms: one may be aware of one's own experience by *attending* to it (as when the doctor asks you to describe the pain you feel in your stomach and you need to bring your attention to the pain sensation to reply); or one may be aware of one's own experience by *thinking* about it (as when you think: "I should avoid spicy food if I want *this burning sensation in my stomach* to end"). However, those who theorize about inner awareness typically focus on a different kind of awareness, more minimal and less cognitively demanding, one that does not imply thinking about or attending to the experience. Inner awareness, in this stricter sense, is often characterized as *pre-reflective*, where this implies its being *non-attentive* and *non-conceptual* (thus not thought-like). The idea is that even when your attention is focused on something else, say, a lively conversation or a difficult task, you are still aware of the pain sensation in your stomach, as long as the sensation is conscious. Although you are not aware of it focally or attentively, you are nonetheless aware of it, so to speak, *peripherally*.

The psychological reality of inner awareness has been a central element of philosophers' reflection about mind and consciousness for centuries, from Aristotle, to Descartes, Brentano, the phenomenological tradition, and early analytic philosophy. In contemporary philosophy of mind, it plays an essential role in some of the most prominent theories of phenomenal consciousness—especially higher-order representationalism and self-representationalism, on which inner awareness

is what *makes* a mental state conscious.<sup>1</sup> Recently, it has been called into question, especially by proponents of the so-called *transparency of experience* thesis (Harman 1990) and first-order representationalists about consciousness (Dretske 1995; Tye 1995), who argue that for a mental state to be conscious it suffices that it represents something (some worldly object or property) in the right way—no meta-level state of (inner) awareness is needed (indeed, they argue, none is present).

Advocates of inner awareness often consider it *phenomenologically* evident: that we are aware of our own current conscious experiences is just a plain phenomenological fact:

[Consider] the case of thinking about  $x$  or attending to  $x$ . In the process of thinking about  $x$  there is already an implicit awareness that one is thinking about  $x$ . There is no need for reflection here, for taking a step back from thinking about  $x$  in order to examine it [...] When we are thinking about  $x$ , the mind is focused on  $x$ , not on our *thinking* of  $x$ . Nevertheless, the process of thinking about  $x$  carries with it a non-reflective self-awareness. (Goldman 1970: 96)

Many consider it also *conceptually* evident: a conscious experience of which one is unaware is simply inconceivable.

Detractors too, however, claim phenomenology on their side. They appeal to the “transparency” of experience:

When you see a tree, you do not experience any features as intrinsic features of your experience. Look at a tree and try to turn your attention to intrinsic features of your visual experience. I predict you will find that the only features there to turn your attention to will be features of the presented tree [...]. (Harman 1990: 39)

According to transparency theorists, what is phenomenologically evident is that we are aware of things in our surroundings (and their properties): no inner awareness is phenomenologically manifest. Indeed, they argue, none is there to be found, even upon careful introspection. Detractors, of course, also disagree on the conceptual evidence of inner awareness and argue that conscious experience of which the subject is unaware is not only perfectly conceivable, but also actual (Dretske 1993). Given such introspective and conceptual disagreement, the quarrel needs to be settled on theoretical, and not just purely phenomenological or intuitive, grounds.

The main goal of this paper is to put forward a theoretical defense of inner awareness.

However, before developing or evaluating any argument concerning inner awareness, we need to get clear exactly on *what thesis* the argument is supposed to support or disprove. This is what I am going to do in the remainder of this section. At the end of the section, I will make explicit the defense of which thesis constitutes *my* primary goal here. At least the following theses can be formulated about inner awareness:

1. *Existence*: we are innerly aware of at least *some* of our conscious experiences at the time of their occurrence. More formally:

---

<sup>1</sup> Both higher-order representationalism (HOR) and self-representationalism (SR) explain inner awareness in terms of meta-representation: a subject  $S$  is aware of a mental state  $M$  in virtue of harboring a mental state  $M^*$  that represents  $M$ . On HOR,  $M$  and  $M^*$  are distinct states; on SR, they are the same state. As for HOR, see Armstrong (1968), Lycan (1996), and (Rosenthal 1997). As for SR, see especially Kriegel (2009); see also Gennaro (1996), Carruthers (2000), Van Gulick (2004), and Frank (2015). For a collection of papers on SR, see Kriegel and Williford (2006).

1. EXISTENCE: There is some conscious state M of a subject S at  $t$ , such that S is innerly aware of M at  $t$ .
2. *Ubiquity* (or *Universality*): we are innerly aware of *each* of our conscious experiences at the time of its occurrence. This thesis may come into two varieties: an *unrestricted* ubiquity thesis quantifies over all subjects of conscious experience; a *restricted* ubiquity thesis quantifies only over a subset of those subjects. The most natural restricted ubiquity thesis quantifies over *neurotypical human adult subjects*. The idea is that not all conscious subjects may feature inner awareness: animals, infants, and some human adults whose cognitive capacities are impaired, may have conscious experiences that are not accompanied by inner awareness. Inner awareness, on this view, is ubiquitous only in the conscious life of all humans that have reached full development of their cognitive capacities (*adults*, rather than infants whose psychology is still in the process of development) and whose cognitive capacities are not impaired (*neurotypical*, rather than affected by mentally impairing conditions):

2a. UBIQUITY<sub>R</sub>: For any conscious state M of a neurotypical human adult S at  $t$ , S is innerly aware of M at  $t$ .

Many inner awareness theorists, however, defend a stronger, unrestricted thesis, one that quantifies over *all* subjects of conscious experience.

2b. UBIQUITY<sub>U</sub>: For any conscious state M of a subject S at  $t$ , S is innerly aware of M at  $t$ .

3. *Necessity*: *Necessarily*, we are innerly aware of each of our conscious experiences at the time of its occurrence. Depending on the kind of modality at play, different versions of this thesis may be formulated. The most relevant involve nomological modality and metaphysical modality respectively. A *nomological necessity* thesis claims that all conscious experiences that obey the actual world's laws of nature are such that their subject is aware of them at the time of their occurrence (“necessarily<sub>n</sub>” stands for “nomologically necessarily”):

3a. NECESSITY<sub>N</sub>: Necessarily<sub>n</sub>, for any mental state M of a subject S at  $t$ , if M is conscious, S is innerly aware of M at  $t$ .

A *metaphysical necessity* thesis claims that all conscious experiences, in all possible worlds, are such that their subject is aware of them at the time of their occurrence (“necessarily<sub>m</sub>” stands for “metaphysically necessarily”):

3b. NECESSITY<sub>M</sub>: Necessarily<sub>m</sub>, for any mental state M of a subject S at  $t$ , if M is conscious, S is innerly aware of M at  $t$ .

4. *Grounding*: Our being innerly aware of our own experiences is what *makes* them conscious. I assume that grounding implies metaphysical necessity: if  $x$  grounds  $y$ , the occurrence of  $x$  metaphysically implies the occurrence of  $y$ . Not everybody agrees on this, but most philosophers do. Accordingly, we have:

4. GROUNDING: Necessarily<sub>m</sub>, for any mental state M of a subject S at  $t$ , if M is conscious, M is conscious in virtue of S's being innerly aware of M at  $t$ .

The four theses are presented in order of strength (from the weakest to the strongest) and in reverse order of entailment (GROUNDING entails NECESSITY which entails UBIQUITY which entails EXISTENCE). EXISTENCE says that inner awareness *exists*. UBIQUITY says that a subject (either *any* subject, or any *neurotypical human adult*, depending on the version) is aware of *each* of their conscious

experiences at the time of its occurrence. It is a descriptive thesis: it says that inner awareness always accompanies consciousness, but this may just be a *contingent* fact. NECESSITY, by contrast, introduces the normative claim that the one described by UBIQUITY is a *necessary* fact about consciousness (either nomologically necessary or metaphysically necessary, depending on the version). Note, however, that inner awareness is here still possibly just an *accompaniment* of consciousness (albeit a necessary one), rather than a constitutive element of it: the idea is that in every possible world conscious experience is accompanied by inner awareness. I think that many advocates of inner awareness defend either some version of UBIQUITY or some version of NECESSITY. Finally, GROUNDING is a thesis about the metaphysical *constitution* of consciousness: inner awareness is not just a necessary accompaniment of conscious experience, but it is what *makes* it a conscious experience at all. I take higher-order representationalists and self-representationalists to defend GROUNDING.

If one accepts EXISTENCE, however, there is a way to reason abductively from it to stronger theses, all the way to GROUNDING. First, one may reason from EXISTENCE to UBIQUITY on the assumption that (i) there does not seem to be any fundamental phenomenological difference between two kinds of conscious experiences—those that are accompanied by inner awareness and those that are not—and (ii) a view on which some experiences are accompanied by inner awareness and some are not might be unnecessarily uneconomical. The idea is that, since there is no apparent principled reason to divide experiences into two groups—those of which the subject is aware and those of which the subject is unaware—we should not so divide them. Accordingly, if we accept EXISTENCE, we should also accept UBIQUITY. As noted, UBIQUITY is compatible with inner awareness being just contingent to consciousness. However, one may think that the best explanation of inner awareness’s ubiquity is its being *necessarily* attached to consciousness: all conscious experiences *are* accompanied by inner awareness because all conscious experiences *must* be so accompanied. So, it seems at least *prima facie* plausible to move from UBIQUITY to NECESSITY via inference to the best explanation. Finally, it seems natural to reason from NECESSITY to GROUNDING too via inference to the best explanation, the idea being that inner awareness is a *necessary* feature of consciousness because it is an *essential* feature of consciousness—it is (part of) what makes something a conscious state at all: inner awareness is necessary to consciousness because it is *constitutive* of consciousness. Of course, each step of this “abductive climbing” may be contested. It is not my purpose to *defend* the abductive climbing, only show that this could be a natural way for an advocate of inner awareness to go. (Though I will use and motivate a similar way of reasoning while arguing for some stronger theses about inner awareness in §4.)

My primary aim is to defend UBIQUITY<sub>n</sub>, that is, the version of UBIQUITY that quantifies over neurotypical human adults:

IA<sub>n</sub>: For any conscious state M of a neurotypical human adult S at *t*, S is innerly aware of M at *t*.

My main (minimal) thesis, then, is that inner awareness is ubiquitous in the conscious life of all humans that have reached full development of their cognitive capacities and whose cognitive capacities are not impaired.

My secondary (maximal) thesis is very ambitious, for it is the strongest thesis that can be formulated about inner awareness, that is, GROUNDING:

IA<sub>M</sub>: Necessarily<sub>m</sub>, for any mental state M of a subject S at *t*, if M is conscious, M is conscious in virtue of S’s being innerly aware of M at *t*.

On  $IA_M$ , inner awareness is a necessary and constitutive element of consciousness as such: any conscious mental state, in every possible world, is conscious (partly) in virtue of its subject's being aware of it.

Here is how I will proceed. In sections 2 and 3, I pursue my primary aim. In §2 I develop an argument for  $IA_m$ , the Memory argument. The argument is not new, but I propose some new ways to reconstruct it. I call it the “Modest Memory argument” because it supports the most modest thesis about inner awareness I intend to defend. In §3 I consider and rebut an objection to the Memory argument recently raised by Daniel Stoljar (forthcoming). Finally, in §4, I propose a way to develop the argument so that it supports not only the modest thesis, but also the more ambitious thesis that constitutes the secondary aim of this paper,  $IA_M$ ; accordingly, I call it the “Ambitious Memory argument.”

A note before continuing. The theses explored in this section are about the *existence/universality/necessity/grounding role* of inner awareness. None of them is about the *nature* of inner awareness. A thesis about the nature of inner awareness makes a claim about *in virtue of what* a subject is innerly aware of their experience. Some take inner awareness to be a *primitive* relation between the subject and their experiences (some, but not all, *acquaintance* theorists maintain something along these lines). Others argue that a subject is aware of their experience in virtue of harboring a mental state directed at the experience. Higher-order representationalism and self-representationalism adopt the latter strategy and explain inner awareness in terms of meta-representation. In what follows, issues around the *nature* of inner awareness are set aside. My concern, here, is to establish a thesis about the *existence, necessity, grounding role, and (level of) universality* of inner awareness, independently of how inner awareness is constituted or explained.

## 2. The Modest Memory argument

Theoretical arguments for inner awareness are not easy to find. Those who accept the psychological reality of inner awareness do so primarily based on *phenomenological*, rather than *theoretical*, considerations. As noted by Kriegel (2019: 144), many of them take inner awareness to be “so self-evident as to *allow* no cogent argument that would derive it from truths even more obvious than it.” The Memory argument is probably the best extant argument that aims to do just that. At its core is the observation that we can recall our past experiences even if we do not attend to them when they occur. But we can only recall something if we were aware of it when it occurred. So, for us to recall an experience, we must be somehow aware—innerly aware—of it at the time of its occurrence.<sup>2</sup>

The argument was originally formulated by the 5<sup>th</sup>-6<sup>th</sup> century Buddhist philosopher Dignāga. It has received extensive attention by Buddhist scholars (see, e.g., Ganeri 1999; Perrett 2003; Kellner 2010; Thompson 2011), but perhaps not as much as it would deserve by philosophy of mind theorists of inner awareness. A notable exception is Uriah Kriegel (2019), who provides a rigorous reconstruction and a thorough defense of the argument.

In this section, I expand on Kriegel's treatment and propose different ways in which the Memory argument may be formulated, by specifying the exact thesis each formulation supports and how it supports it, and by considering some *prima facie* objections.

---

<sup>2</sup> Here and throughout, “remembering” and “recalling” are treated as *factive*: remembering (or recalling) E implies that E occurred. “Seeming to remember” and “seeming to recall” are the non-factive counterparts of the verbs.

A first formulation of the argument runs roughly as follows:

- P1. We can only remember something if we were aware of it at the time of its occurrence.
- P2. We remember some of our own conscious experiences.
- C. Therefore, there are some conscious experiences of ours of which we are aware at the time of their occurrence.

The target of this argument is, visibly, a version of EXISTENCE. The idea is that, since we remember some of the experiences we had, there are at least some experiences that are such that we were aware of them at the time of their occurrence. A more precise formulation of the argument is:

ARGUMENT 1

- P1. For any subject S, event E occurring at time  $t$ , and time  $t_1$  such that  $t < t_1$ , S can remember E at  $t_1$  only if S is aware of E at  $t$ .
- P2. There is some conscious state M of S at  $t$ , such that S remembers M at  $t_1$  (where  $t < t_1$ ).
- C. There is some conscious state M of S at  $t$ , such that S is aware of M at  $t$ .

For C to follow from P1 and P2, the argument needs a suppressed premise, namely that conscious states are events. Here I assume that a phenomenally conscious state is a Kimean event, i.e., the instantiation of a phenomenal property by a subject at a time (Kim 1993).

P1 seems straightforwardly true. Consider: the police will ask for your help and rely on your memory to reconstruct the dynamics of a car accident only if you *witnessed* it—i.e., only if you were somehow *aware of* the accident. Obviously, they will make no use of your memory if you did not see, hear, or experience the accident in any way. It may be objected that, in fact, we can and do remember events we have never been aware of. I can remember that French Revolution occurred in 1789, even though I could not be aware of *anything* in 1789. Note, however, that what I remember here is the *fact* (the true proposition) that French Revolution occurred in 1789: *French Revolution*—the *event* itself—is something I *cannot* recall. At play here is the distinction between so-called *episodic* and *semantic* memory (Tulving 1972). The former consists in first-person recollection of events of one’s personal past; the latter is third-person memory of facts of the world. The former takes an *event* as its object; the latter takes a *proposition*. Reports of the former deploy a noun or noun phrase (“I recall *the accident*,” “I recall *eating sushi* in Houston”); reports of the latter deploy a that-clause (“I remember *that* the accident occurred in Rue Sainte-Anne,” “I remember *that* sushi is one of your favorite foods”). One may have semantic memory of an event of one’s personal past without having episodic memory of it: I may remember *that my grandmother used to take me* to Villa Giulia when I was a child without recalling *being taken there* by my grandmother. There is no agreement on the analysis and definition of episodic memory. However, there is general consensus that having *first-personal experience* of an event is a necessary condition for episodic memory (though not a sufficient one: arguably, I was aware of my grandmother’s taking me to the park; nevertheless, I do not have episodic memory of that event—I forgot it). Naturally, one can have also semantic memory of what one has episodic memory of. If you recall (episodically) the accident, you may also remember (semantically) *that* the accident occurred in Rue Sainte-Anne. *Experiencing* the accident is necessary for you to recall it episodically, but is neither necessary nor sufficient for you to remember (semantically) the fact that it occurred in Rue Sainte-Anne.

The very notion of episodic memory thus seems to imply that one can only recall (episodically) events one is aware of when they occur. If so, a more precise formulation of P1 is true by definition; P2 is updated accordingly:

## ARGUMENT 2

- P1. For any subject S, event E occurring at time  $t$ , and time  $t_1$  such that  $t < t_1$ , S can *episodically* remember E at  $t_1$  only if S is aware of E at  $t$ .
- P2. There is some conscious state M of S at  $t$ , such that S *episodically* remembers M at  $t_1$  (where  $t < t_1$ ).
- C. There is some conscious state M of S at  $t$ , such that S is aware of M at  $t$ .

P2 seems to be just a plain fact about what we can and do remember. You can recall not only *the accident*—the *car*, the *bang*, the *smoke*, etc.—but also your experience(s) of the accident—the way the car *looked* to you, the way the bang *sounded* to you, the way the smoke *smelled* to you. The latter are conscious experiences: your visual experience of the car, your auditory experience of the bang, your olfactory experience of the smoke. Your memory of them is *episodic*: you *first-personally* recall *seeing* the car, *hearing* the bang, and *smelling* the smoke. You may, *additionally*, remember semantically *that* you saw a car, *that* you heard a bang, *that* you smelled smoke. But those semantic memories do not exhaust your memories of your experiences of the accident. For you can remember not only facts or propositions *about* your experiences: you can recall the experiences themselves.

From P1 and P2 follows that we must be aware of at least *some* of our experiences—those we episodically recall (C).

It may be objected that the kind of awareness that appears in C need not be inner awareness (the non-attentive and non-conceptual kind of awareness introduced in §1). The underlying thought would be that the only experiences we remember are those that we introspected (those we attended to when they occurred). If so, the argument does not show that some experiences are object of inner awareness—only that some experiences are introspected.

That we can only remember the experiences we introspect, however, is implausible. We might have more *vivid* or more *detailed* episodic memories of introspected experiences than of non-introspected ones. However, although perhaps less vividly and in less detail, it seems undeniable that we can and do remember some experiences we did not introspect. Take the car accident again. Arguably, when you witnessed the accident, your attention was (almost) fully captured by the event taking place before your eyes—your attention was focused on the accident, rather than on your experience of it. Even so, when you recall the accident later on, you recall not only the accident (and things and people involved in it), *tout court*, but you recall the accident *from a certain point of view*: you recall how the things and people involved in it *looked* to you, *sounded* to you, *smelled* to you, etc.—you recall *the way you experienced* those things and people.<sup>3</sup> Therefore, not only introspected experiences can be recalled, but also non-introspected ones. Accordingly, the argument goes:

## ARGUMENT 3

- P1. For any subject S, event E occurring at time  $t$ , and time  $t_1$  such that  $t < t_1$ , S can episodically remember E at  $t_1$  only if S is aware of E at  $t$ .
- P2. There is some *non-introspected* conscious state M of S at  $t$ , such that S episodically remembers M at  $t_1$  (where  $t < t_1$ ).
- C. There is some *non-introspected* conscious state M of S at  $t$ , such that S is aware of M at  $t$ .

---

<sup>3</sup> The reasoning is partly inspired by Siewert (2004).

It is now explicit that the kind of awareness involved in C is inner awareness in the strict sense: a non-attentive and non-conceptual awareness of experience.

The argument just presented supports only EXISTENCE. There are two ways to produce an argument for the stronger and more interesting thesis of UBIQUITY<sub>R</sub>—one abductive and one deductive.

The abductive reasoning is that the best explanation of C in ARGUMENT 3 is that, for *any* conscious experience of ours, we are aware of it at the time of its occurrence. Who is “us”? At a minimum, *neurotypical human adults*. Accordingly, we may also reason as follows:

ARGUMENT 4

- P1. For any neurotypical human adult S, event E occurring at time  $t$ , and time  $t_1$  such that  $t < t_1$ , S can episodically remember E at  $t_1$  only if S is aware of E at  $t$ .
- P2. There is some non-introspected conscious state M of a neurotypical human adult S at  $t$  such that S episodically remembers M at  $t_1$  (where  $t < t_1$ ).
- C1. There is some non-introspected conscious state M of a neurotypical human adult S at  $t$ , such that S is aware of M at  $t$ .
- P3. The best explanation of C1 is that for *any* conscious state M of a neurotypical human adult S at  $t$ , S is aware of M at  $t$ .
- C2. For any conscious state M of a neurotypical human adult S at  $t$ , S is aware of M at  $t$ .

The second part of this Memory argument is abductive. If some of our experiences are such that we are aware of them at the time of their occurrence, it is plausible that we are aware of *all* our conscious experiences the same way (P3). For there seems to be nothing special to the experiences we actually remember—nothing that makes them fundamentally different from the experiences we do not remember.

The key idea is this. The difference between the experiences we remember and those we do not is *not* a *fundamental* difference between (a) experiences it is *nomologically possible* and (b) experiences it is *nomologically impossible* for us to recall. Rather, it is a difference between (in principle equally rememberable) experiences we *happen to remember* and experiences we *happen to forget* (due to contingent factors). It is a difference not in the experiences themselves, but in the subsequent cognitive processes through which information acquired through experience is managed and stored. Although we actually recall only *some* of our (non-introspected) experiences, we *could* recall any of them. So the reason there are experiences we do not recall is *not* that we *could not* recall them, but that we *forgot* them. For any conscious experience we have forgotten, there is the nomological possibility for us to recall it (had it entered the right web of associations that traps it in long-term memory). And there can only be the nomological possibility for us to recall an experience (or indeed any event) if we were aware of it at the time of its occurrence.

Note that the same reasoning applies to our memories of external events. Some external events we episodically remember, some we do not. But this does not imply that all the events we do not remember are events we did not experience. Obviously, some events we cannot remember because we did not experience them. But others we did experience, and could remember, but happened to forget. The same reasoning applies, then, to both experiences and external events. For any external event we can recall, we can recall both the event itself and our experience of it. Therefore, at the time of its occurrence, we are aware of both the event and our experience of it. Naturally, we can also recall experiences that are not *of external events*, as when one recalls one’s euphoric mood, one’s

strong desire, or one's thought. And we can recall those experiences regardless of whether we introspected them at the time of their occurrence.

These considerations suggest an alternative formulation of the argument. In ARGUMENT 4, the first two premises are relatively undemanding—they simply register facts about human memory—and most of the work is done by P3, which underwrites the inference to the best explanation. However, a *deductive* reasoning to UBIQUITY<sub>R</sub> can be developed, via a thicker premise:

#### ARGUMENT 5

- P1. For any neurotypical human adult S, event E occurring at time  $t$ , and time  $t_1$  such that  $t < t_1$ , S can episodically remember E at  $t_1$  only if S is aware of E at  $t$ .
- P2. For any conscious state M of a neurotypical human adult S at  $t$ , there is a time  $t_1$  (where  $t < t_1$ ) such that S can episodically remember M at  $t_1$ .
- C. For any conscious state of a neurotypical human adult S at  $t$ , S is aware of M at  $t$ .<sup>4</sup>

Given that here P2 is stronger, no extra (abductive) premise is needed. In the abductive argument, P2 merely registers the fact that some experiences are actually remembered. In the deductive argument, P2 makes the slightly stronger claim that all experiences are *memorable*, “in the weak sense that it is possible for the subject to remember its occurrence at some later time” (Kriegel 2019: 146). The stronger claim, however, is quite plausible, as per the considerations above. Although many conscious experiences are not recalled, this is not because their subject was not innerly aware of them when they occurred, but because their subject forgot them. Even if some are forgotten, all conscious experiences are such that it is nomologically possible for them to be later recalled by their subject.<sup>5</sup>

Both ARGUMENT 4 (abductive) and ARGUMENT 5 (deductive) support the main thesis I aim to defend, that is:

- IA<sub>m</sub>: For any conscious state M of a neurotypical human adult S at  $t$ , S is innerly aware of M at  $t$ .

### 3. A new objection

In a recent paper, Daniel Stoljar (forthcoming) advances a systematic critique of arguments for inner awareness, including the Memory argument. He proposes a reconstruction of the argument that is importantly different from those offered above, and develops an objection based on that reconstruction. His objection targets a version of UBIQUITY. I will treat it as an objection to the weakest version of UBIQUITY, UBIQUITY<sub>R</sub>, i.e., IA<sub>m</sub>.

---

<sup>4</sup> This is very close to the way Kriegel (2019: 146) reconstructs the argument.

<sup>5</sup> Kriegel also considers an objection to P1 hinging on cases where S can correctly be said to remember E, but E's occurrence was prior to S's awareness of E. For example, I can correctly say that I remember the sun shining yesterday at 17:13, although what I was experiencing at 17:13 was really the way the sun was at 17:05 (given that it takes eight minutes for the light to get from the sun to my retina). He proposes two ways to deal with this: one is “to loosen sufficiently the relevant notion of simultaneity to allow for whatever additional time is needed for the causal process whereby E causes perception-of-E;” the other is “to restrict the scope of the first premise to ‘nearby events,’ understood as events for which there is no pertinent time lag between their occurrence and perceptual awareness of their occurrence.” (Kriegel 2019: 149). He opts for the latter, but I think that either option is satisfactory. I myself adopt the former and accordingly make no change to P1.

In what follows, I present Stoljar's reconstruction and his objection; I show that his reconstruction does not do justice to the real commitments of the original Memory argument and highlight the weaknesses of his objection; finally, I distill a deeper worry that seems to underlie his objection and try to dissolve it.

Stoljar's reconstruction starts with the following considerations:

Suppose I (currently) remember that I saw the blue sky yesterday. [...] It [...] follows that I currently know that I saw the blue sky yesterday; if you (genuinely) remember something, you know it is the case. But it does *not* follow that yesterday I knew that I was seeing blue sky. After all, it may be that yesterday my mind was so crowded out with other things [...] that I took no notice of the fact, and so did not know, that I was seeing the sky, even though I was. (Stoljar, forth.: 14)

The idea is that there must be *something* that grounds my current state of knowing that I saw the blue sky yesterday. However, that cannot be a *previous* state of knowing, given that yesterday I did *not* know that I was seeing the sky—arguably because, since I did not attend to or notice that I was seeing it, I could not form the belief that I was seeing the blue sky. The inner awareness theorist, argues Stoljar, proposes that what grounds my current state of knowledge is my *previous state of inner awareness*—my being aware of seeing the blue sky yesterday. Here is how I understand his reconstruction of the argument:

- P1. I remember that I saw the blue sky yesterday.
- P2. If I remember that I saw the blue sky yesterday, I know that I saw the blue sky yesterday.
- P3. I know that I saw the blue sky yesterday only if either (i) yesterday I knew that I was seeing the blue sky, or (ii) yesterday I was innerly aware of seeing the blue sky.
- P4. Not (i).
- C. (ii)

The underlying idea is that there are only two possible explanations of the fact that I currently know that I saw the blue sky yesterday. The first is what Stoljar calls the “knowledge model” (corresponding to option (i) in the argument):

On this model, I remember that I saw the blue sky yesterday just in case (a) I currently know that I saw the blue sky yesterday; (b) yesterday I knew that I saw the blue sky; and (c) my current state of knowledge depends in the right way on my previous state of knowledge. (*ibid.*)

However, this model is not applicable to cases in which I remember an experience of mine even though I did not know that I was having it when it occurred (because I was not attending to or introspecting it). Therefore, Stoljar's reconstruction goes, the only option left is what he calls the “inner awareness model” (ii):

I remember that I saw the blue sky yesterday just in case (a) I currently know that I saw it yesterday; (b) yesterday I was (phenomenally) aware of seeing it [...]; and (c) my current state of knowledge depends in the right way on my previous state of inner awareness. How does my current state of knowledge depend on my previous state of awareness? Here I think the suggestion is that my current state of knowledge depends on the fact that I am currently (phenomenally) aware of *having seen* the blue sky yesterday, and this state of awareness is connected in the right way to the fact that yesterday I was (phenomenally) aware of my seeing the sky. (*ibid.*)

As Stoljar reconstructs it, the argument only supports an *existential* thesis about inner awareness. However, it may be reformulated to support a *universal* thesis (I assume that the premises quantify over neurotypical human adults):

- P1. At  $t_1$  S remembers that S was in M at  $t$  (where  $t_1 > t$ ).
- P2. If at  $t_1$  S remembers that S was in M at  $t$ , then at  $t_1$  S knows that S was in M at  $t$ .
- P3. At  $t_1$  S knows that S was in M at  $t$  only if either (i) at  $t$  S knows that S is in M, or (ii) for any subject S and conscious state M of S at  $t$ , at  $t$  S is aware of being in M.
- P4. Not (i).
- C. (ii).

The idea is that there are only two available explanations of the fact that, at  $t_1$ , S knows that S was in M at  $t$ : either at  $t$  S knew that S was in M, or a universal thesis about inner awareness is true. Since a subject can be in M without knowing that s/he is, the best explanation of the fact that one can know that one *was* M is that one is innerly aware of all of one's conscious states. As the reader can see, this reconstruction is very different from any of the reconstructions offered in §2. I will come back to this later.

Stoljar's objection targets P3. He suggests that the "knowledge model" and the "inner awareness model" do not exhaust the relevant options and proposes a third model—what he calls the "perceptual model:"

I remember that I saw the blue sky yesterday just in case (a) I know that I saw the sky yesterday; (b) yesterday I saw the sky, and so was perceptually aware of the sky's being blue; and (c) my current state of knowledge depends on my previous perceptual state. How does my current state of knowledge depend on my previous perceptual state? The answer is that my current state of knowledge depends on a current state of perceptual awareness, according to which the sky *was* blue yesterday, and this state in turn bears the right relation to the state of perceptual awareness I was in yesterday when I saw the sky. (Stoljar, forth.: 15)

So, according to Stoljar, there is a third possible explanation of the fact that, at  $t_1$ , S knows that S was in M at  $t$ : at  $t$ , S is aware of what M is *about*—the state of affairs that constitutes the content of M. So, I now know that I had the experience of seeing the blue sky yesterday in virtue of my now being *perceptually* aware of the sky *having been* blue yesterday, and my current state of perceptual awareness depends on the fact that yesterday I was perceptually aware of the blue sky. On this model, having an experience (of a certain external event) is sufficient for one to subsequently remember that one had that experience—inner awareness of it is not required.

Stoljar argues that his "perceptual model" is superior to the "inner awareness model" based on theoretical-economy considerations: it has the same amount of explanatory power but does not entail the existence or ubiquity of inner awareness. Moreover, he argues, the "perceptual model" provides a unified explanation of both the fact that I now know that I saw the blue sky yesterday and the fact that I now know that the sky was blue yesterday (even if yesterday I knew neither fact, because I formed neither the belief that I was seeing the blue sky nor that the sky was blue): both are explained by my having been perceptually aware of the blue sky yesterday (Stoljar, forth.: 15-16).

There are several problems with Stoljar's argument. But first let me point out how his reconstruction of the Memory argument differs from the reconstructions discussed in §2. On his reconstruction, the argument is much weaker and seems to drift considerably apart from the core ideas underlying the original Memory argument. As noted, the original argument relies on what seems to be a simple fact about episodic memory, i.e., that we can only episodically recall events of which we were aware at the time of their occurrence, and on the observation that we can episodically recall our own experiences. As noted in §2, a *deductive* argument draws from these two observations the conclusion that we are aware of our own experiences:

## ARGUMENT 5

- P1. For any neurotypical human adult S, event E occurring at time  $t$ , and time  $t_1$  such that  $t < t_1$ , S can episodically remember E at  $t_1$  only if S is aware of E at  $t$ .
- P2. For any conscious state M of a neurotypical human adult S at  $t$ , there is a time  $t_1$  (where  $t < t_1$ ) such that S can episodically remember M at  $t_1$ .
- C. For any conscious state M of a neurotypical human adult S at  $t$ , S is aware of M at  $t$ .

This is a *deductive* argument whose premises are *facts* about episodic memory (i.e., memory of first-personally experienced *events* of one's personal past). Stoljar's reconstruction, by contrast, does not rely on facts about episodic memory, but rather revolves around *how* and *on what grounds* one *knows* that one had a certain experience. Its premises 1 and 2 concern *semantic*, rather than *episodic* memory: they concern (propositional) memory of *facts* about one's past experience ("I remember *that I saw the sky* yesterday"), rather than experiential (objectual) memory *of the experience* itself. But we saw that the Memory argument is *not* about semantic memory. Arguably, part of the reason why Stoljar's reconstruction revolves around *knowledge of facts* about past experiences is that he takes the kind of memory involved in the argument to be semantic, and semantic memory (if it is *genuine*, rather than *apparent* memory) grounds knowledge of facts about past events. So, Stoljar attributes to the inner awareness theorist the claim that a universal thesis about inner awareness is the best explanation of how such knowledge is grounded—via what he calls the "inner awareness model" of past-experience knowledge acquisition. However, this is *not* the claim that underlies the original Memory argument. Indeed, the original argument is much stronger, partly because it relies on claims (about facts about episodic memory) that are much more plausible than the claim underlying Stoljar's reconstruction (a claim about what grounds one's knowledge about past experiences).

So, even if Stoljar's objection to the argument as he reconstructs it was convincing, nothing in what he says speaks to the original Memory argument's underlying idea. Even if his objections were convincing, they would only show that one possible Memory argument (i.e., Stoljar's own reconstruction) is weak—they would not threaten the argument to which proponents of inner awareness actually appeal to.

However, Stoljar's objection is *not* convincing. First, it is unclear what "being *perceptually* aware of the sky's *having been* blue" amounts to. Arguably, we can only perceive objects, properties, and events that are *present*: perception involves a causal process connecting the event perceived and the subject's perceptual representation of it via stimulation of the subject's perceptual system, where the perceived event and the subject's representation of it are (roughly) simultaneous. An event that is no longer occurring cannot be, strictly speaking, *perceived*. Indeed, arguably, this is part of what fundamentally distinguishes perception from other mental states such as mere thinking, imagining, or remembering. Unlike those mental states, it is a necessary condition for perceptual experience that what is perceived is *represented as present* and, if the experience is veridical, what is represented *is present*. It may be objected that we can perceive events of the past, as when we see the shining of a star that is so far away from Earth that when its light hits our retina, the star has already stopped shining for years. However, for one thing, what seems crucial to perception is that the *physical stimulation* of the perceptual system be (roughly) simultaneous with the perceptual experience it causes (I say "roughly" to allow for the time lapse due to the causal process from the physical stimulation to the conscious perceptual representation); this accommodates cases such as seeing stars. For another, there is a crucial *phenomenological* feature that distinguishes perceptual experience from mere thought, imagination, and recollection: what is perceived is perceived *as present, as here and now*. Stars are perhaps perceived as being far away, but are still perceived as

present. Experiencing something *as having occurred*, by contrast, simply does not have the phenomenology of perception. Therefore, one cannot literally *perceive* the sky as *having been* blue.

Perhaps what Stoljar has in mind is not literal perceiving, but rather a looser notion corresponding roughly to “outer awareness,” so that by “perceiving the sky’s having been blue” he would mean “being *outwardly aware* of the sky’s *having been* blue.” Arguably, this amounts to a representational state whose content is a past event—“representing the sky as having been blue.” But, if so, it is unclear what “perceiving the sky’s having been blue” consists in if not just “recalling the blue sky,” plus a commitment to a model of memory such that memory is a mental representation whose content is constituted by an event and its pastness—or by the fact that a certain event occurred in the past. And if “perceiving the sky’s having been blue” is just remembering the blue sky, it is unclear what explanatory role it is supposed to play, given that the subject’s memory of the sky is just a datum in the argument.<sup>6</sup>

Stoljar anticipates the objection that “one cannot be perceptually aware of the sky’s *having been* blue yesterday” (forth.: 15) and offers two replies. One is that “we are plausibly committed to the perceptual model anyway” (*ibid.*). He argues that the perceptual model is still needed to account for cases in which one remembers an external event (rather than an experience): when I remember that the sky was blue yesterday, I know that the sky was blue yesterday and what explains my current state of knowledge is my current state of *perceiving* the sky’s *having been* blue yesterday, which is connected in the right way to my yesterday’s state of perceiving the sky’s being blue. However, as noted, it is unclear what “*perceiving* the sky’s *having been* blue” amounts to if not just “remembering the blue sky,” so more should be said about why and in what sense Stoljar’s “perceptual model” is indispensable for the explanation of memory of external events. Besides, it seems that a simple “awareness model” could have at least the same explanatory power: I can now recall the blue sky (partly) in virtue of the fact that yesterday I was aware of the blue sky (and partly in virtue of the fact that information acquired via yesterday’s awareness was correctly stored in long-term memory which I can now consciously access).

Stoljar’s other reply is that “the proponent of the inner awareness model requires the analogous thing to be true of inner awareness, since it requires that one is phenomenally aware of *having seen* the sky yesterday” (*ibid.*). Now, if by “being phenomenally aware of *having seen* the sky yesterday” he simply means “recalling seeing the blue sky,” then of course the inner awareness theorist is committed to that, but there is nothing so compromising about this commitment: for most would agree that we can remember our experiences. If, instead, Stoljar means anything thicker than just “recalling seeing the blue sky,” then the inner awareness theorist is *not* committed to that. For one thing, no thesis about inner awareness implies that our awareness of our experiences is *analogous to perception*. For another, differently from what Stoljar suggests, the Memory argument does not imply commitment to what he calls the “inner awareness model.” In particular, it does not imply commitment to the existence of a mental state of “being phenomenally aware of having seen the blue sky” if this is understood as distinct from or more theoretically loaded than simply “recalling seeing the blue sky.” The Memory argument does not assume any “inner awareness of the past,” simply because, unlike Stoljar’s “perceptual model” (that, as noted, on the most charitable interpretation, presupposes an account of memory in terms of outer awareness of past events), it

---

<sup>6</sup> Perhaps Stoljar’s idea is that construing “remembering the blue sky” as “representing the sky as having been blue” makes a dialectical difference, based on some transparency assumptions about memory. I am going to consider this below, and show that it fails to make the perceptual model more plausible.

does *not* presuppose any specific account of episodic memory. Therefore, the inner awareness theorist is *not* anything analogous to Stoljar’s notion of outer awareness of past events.

This becomes even more evident when we consider how significantly Stoljar’s reconstruction of the argument drifts apart from the original argument’s core ideas. As noted, the reconstructions offered in §2 do not involve any claim about *how* or *on what grounds* a subject currently *knows* that s/he had a certain experience, which is, instead, what constitutes the heart of Stoljar’s reconstruction. The question whether an “inner awareness model” is needed arises only if the argument is reconstructed in such a way that it requires a *model* for the acquisition of propositional knowledge about one’s past experiences. But the Memory argument does not need to be so reconstructed. Quite the contrary, for it seems that Stoljar’s reconstruction does not do justice to the original argument’s real commitments.

There is, however, a way to distill, by a charitable reading of Stoljar’s discussion, a more pointed objection targeting the original argument. The idea is that, when it comes to experiences, what is required for one to be able to episodically remember them is not that one is aware of the *experience*, but rather that one is aware of *what the experience is about* at the time of the experience’s occurrence. So, for me to recall my visual experience of the blue sky I do not need to be aware of the visual experience itself when it occurs, but only of what the experience is about, i.e., the blue sky. So construed, the objection targets P1 of the original argument (i.e., the premise that one can only episodically remember an event if one is aware of it when it occurs): to episodically remember an event one must be aware of it at the time of its occurrence *except when the event is an experience*; to episodically remember an experience it is sufficient that one is aware of what the experience is about, at the time the experience occurs. Accordingly, the objector would propose a *disjunctive* formulation of P1:

P1\*. For any neurotypical human adult S, event E occurring at time  $t$ , and time  $t_1$  such that  $t < t_1$ , if E is not an experience, S can episodically remember E at  $t_1$  only if S is aware of E at  $t$ ; if E is an experience, S can episodically remember E at  $t_1$  if at  $t$  S is aware of what E is about.

So, the idea is that there are some events—experiences—such that, for S to remember them, it is sufficient that S was aware of something *other* than the event itself at the time of its occurrence. From P2 and *this* formulation of P1 the conclusion (UBIQUITY<sub>R</sub>) does *not* follow.

However, there are reasons not to consider even this more pointed objection as compelling. First, the principle underlying P1 in the original argument reflects (i) a *unified* account of episodic memory and (ii) what seems to be a *conceptual truth* about episodic memory. As noted in §2, it seems that the very concept of episodic memory implies that one cannot episodically recall an event of which one was unaware at the time of its occurrence. Episodic memory is memory of one’s personal past *from the first-person perspective*: it is memory of *first-personally experienced* personal past. This implies that one can only remember events that one first-personally experienced. It thus seems to be a conceptual truth about episodic memory that only events of which one was aware of at the time of their occurrence can be episodically remembered. (i) and (ii) are very good reasons to stick to the principle. Giving up that principle (as P1\*’s disjunctive claim implies) would be reasonable only in exchange for an exceptional theoretical or explanatory advantage.

However, the disjunctive claim does *not* come with a better explanation of facts about episodic memory of experiences. For, as we saw, Stoljar’s proposal involves the notion of *perception of a past event*—such as “perceiving the sky’s having been blue”—which, at the very least, would require

much theoretical backup. Moreover, even if the notion of perception of a past event could be made sense of, the explanation of episodic memory of experiences based on it would be cumbersome—certainly *less economical* than the explanation available to inner awareness theorists. On the inner awareness explanation, what explains episodic memory of experiences is the same process that explains episodic memory of *any* event: the current memory is based on information acquired via the previous awareness of the event—simple and elegant as that. On the disjunctive-based explanation, instead, episodic memory of the experience is grounded in awareness of *something other* than the experience itself. Visibly, much theoretical work is needed to get a plausible explanation of the fact that episodic memory of the experience is a mental state *about the experience*, rather than (just) about what the experience is about. For, arguably, *mere awareness of a past event* is not sufficient to explain one’s memory of one’s *experience* of that event: mere awareness of the sky having been blue does not suffice to explain my memory of my *experience* of the blue sky. Being merely aware of the sky having been blue does not yet give me information about *the way* I was aware of the sky—information about how the sky *looked to me*.

That we can and do remember our own experiences, rather than just the events they are about, is something that Stoljar seems to grant. However, a harsher opponent may simply reject the existence of memory of experiences (thus reject ARGUMENT 5’s P2). The underlying idea is a sort of “transparency thesis” about episodic memory: we can only recall the external events we experienced—we cannot recall our experiences of them.

However, a transparency view of this sort would return wrong results, both phenomenologically and epistemically (Fernández 2006). When a subject has an episodic memory (or an apparent episodic memory as) of an event E, it is part of the memory experience’s phenomenology that what is remembered was *experienced* by the subject in the past. As noted by Fernández, “a subject’s episodic memories appear to her as coming from her own past perceptions, as opposed to testimony or reasoning. Once I appear to be remembering a certain event episodically, the question whether or not I seem to have perceived that event is no longer open.” (2006: 43). Being memory *of one’s first-personal experience* of E, rather than just memory of E, is what enables one to distinguish, introspectively, episodic from semantic memories: if in episodic memory one could only recall a past event and not also one’s experience of that event, “I simply would not be able to tell, in the absence of collateral information derived from testimony or reasoning, those memories of mine that appear to be episodic from those memories that appear to be semantic.” (*ibid.*). So, the transparency thesis returns the wrong *phenomenological* results.

It also returns the wrong *epistemic* results. Imagine that you witness a car accident between a blue car and a red car, but you undergo a visual illusion such that the red car looks *green* to you. So, although the actual *event* that is the object of your experience is an accident between a blue car and a *red* car, what you visually *experience* is an accident between a blue car and a *green* car. Based on this experience, you form the episodic memory of an accident between a blue car and a *green* car. This is what you report to the police. For years you do not think about the accident, but one day you seem to remember witnessing an accident between a blue car and a *red* car. You are misremembering, since what you visually *experienced* was an accident between a blue car and a *green* car. On the transparency view, however, this turns out to be a genuine memory of yours, since the actual event remembered was an accident between a blue car and a red car. But this is a wrong result. To get the right epistemic result, we need to assume that episodic memory of E is not only memory of E, but also of S’s experience of E (Fernández 2006; see also Kriegel 2019).

Finally, the disjunctive claim underlying P1\* seems to be *ad hoc*, and the objection based on it to beg the question against the inner awareness theorist: they seem to be motivated by a previous commitment to a transparency thesis and *against* a ubiquity thesis about inner awareness. Why does the objector assume a disjunctive principle about the memorability of events, according to which all events but experiences are memorable only if the subject is aware of them at the time of their occurrence? Arguably, they assume the disjunctive principle because they antecedently assume that experiences are *not* objects of awareness—they antecedently assume that we are aware of external objects, properties, or events, but not of our own experiences.

This puts the objector in an embarrassing position and forces them, at the very least, to provide compelling motivation for the transparency assumption. Note that the premises of the original Memory argument for inner awareness, by contrast, do *not* assume any thesis about the existence or ubiquity of inner awareness: those premises only concern *general* (not *ad-hoc*) facts about what can and what cannot be remembered.

I conclude that Stoljar’s objection, even on a most charitable reading, does not shake the Memory argument, which remains one of the strongest and most convincing arguments for the existence and ubiquity of inner awareness.

#### 4. The Ambitious Memory Argument

In §§2-3 I defended IA<sub>m</sub>:

IA<sub>m</sub>: For any conscious state M of a neurotypical human adult S at *t*, S is innerly aware of M at *t*.

If the reader has been on board with me so far, I achieved the primary aim of this paper!

Now, as announced in §1, I would like to make some further steps toward the following stronger thesis:

IA<sub>M</sub>: Necessarily<sub>m</sub>, for any mental state M of a subject S at *t*, if M is conscious, M is conscious in virtue of S’s being innerly aware of M at *t*.

IA<sub>M</sub> is the strongest thesis about inner awareness, i.e., GROUNDING. My defense of it proceeds in two steps. First, we reason from IA<sub>m</sub> (that is, UBIQUITY<sub>R</sub>) to NECESSITY<sub>M</sub>, that is, the thesis that all conscious experiences, in all possible worlds, are such that their subject is aware of them at the time of their occurrence:

NECESSITY<sub>M</sub>: Necessarily<sub>m</sub>, for any mental state M of a subject S at *t*, if M is conscious, S is innerly aware of M at *t*.

Then, we take a further step from NECESSITY<sub>M</sub> to GROUNDING.

There are two ways to move from UBIQUITY<sub>R</sub> to NECESSITY<sub>M</sub>, one abductive and one deductive. I will go through the abductive way first, and then move to the deductive way.

A first step in the abductive reasoning consists in moving from UBIQUITY<sub>R</sub> to UBIQUITY<sub>U</sub> (i.e., from restricted to unrestricted ubiquity). Part of what makes the Memory argument for UBIQUITY<sub>R</sub> plausible is that neurotypical human adults are capable of episodic memory. However, as noted by Kriegel, “the memorability claim cannot be true of *every* conscious state, since plausibly, some conscious creatures *have* no episodic memory. Some insects, for instance, may have rudimentary

experiences of their environment without having any capacity for recollecting them.” (2019: 151). If you think that insects do *not* have experiences, you can still conceive of some subject who does have experiences but no ability to recall them—e.g., some other animals, or someone like the protagonist of *Memento*, who is a human adult but has a condition such that, although he has conscious experiences, his brain cannot store information about them in long-term memory.

Although not all conscious creatures can episodically remember their experiences, we can still reason abductively from  $UBIQUITY_R$  to  $UBIQUITY_U$ . The idea is that, plausibly, if all the experiences of memory-capable subjects are such that their subject is innerly aware of them, so should be the experiences of non-memory-capable subjects: “After all, the existence of episodic memory is not claimed here to generate the subject’s awareness of her conscious state, but merely to proffer an instructive symptom of it.” (Kriegel 2019: 152). Accordingly, we have:

ARGUMENT 6

- P1. For any neurotypical human adult S, event E occurring at time  $t$ , and time  $t_1$  such that  $t < t_1$ , S can episodically remember E at  $t_1$  only if S is aware of E at  $t$ .
- P2. For any neurotypical human adult S, and conscious state M of S at  $t$ , there is a time  $t_1$  (where  $t < t_1$ ) such that S can episodically remember M at  $t_1$ .
- C1. For any neurotypical human adult S and conscious state M of S at  $t$ , S is aware of M at  $t$ .
- P3. The best explanation of C1 is that, for any mental state M of a subject S at  $t$ , if M is conscious, S is innerly aware of M at  $t$ .
- C2. For any mental state M of a subject S at  $t$ , if M is conscious, S is innerly aware of M at  $t$ .

The first part of the argument is deductive (i.e., the deductive argument for  $UBIQUITY_R$ ); the third premise makes the overall argument abductive. An extra abductive step takes us from  $UBIQUITY_U$  to  $NECESSITY_M$ : the best explanation of the fact that all conscious experiences are such that their subject is innerly aware of them is that inner awareness is metaphysically necessary to consciousness. However, there is also a different way to move from  $UBIQUITY_R$  to  $NECESSITY_M$ .

Similarly to before, a deductive argument can be developed by strengthening one of the premises. P1 of the abductive argument is that neurotypical human adults can episodically remember only what they experience. Arguably, this is a (rather weak) claim about what is *nomologically* impossible for neurotypical human adults: the way human memory works is such that it is nomologically impossible for us to episodically remember something of which we were not aware. However, as noted in §2, it seems that the very *concept* of episodic memory implies that one cannot episodically recall an event one was unaware of at the time of its occurrence. For episodic memory is not just memory of one’s personal past—as noted, we can have also semantic memory of our personal past. To be *episodic*, memory of one’s personal past must be *experiential*, i.e., *from the first-person perspective*. Its being experiential in this sense implies that episodic memory must be memory of *first-personally experienced* personal past. For it is impossible for one to have a genuine first-person perspective memory of an event without having experienced it first-personally at the time of its occurrence. It thus seems that the impossibility to recall events of which one was unaware is not just nomological, but also conceptual.

The first premise of the deductive argument for  $NECESSITY_M$  is thus that it is *conceptually impossible* for one to episodically recall something of which one was unaware at the time of its occurrence. This is how Kriegel (2019: 152) proposes to reconstruct the argument accordingly:

- (1) For any subject S and nearby event E, if S is not aware of E when E occurs, then it is conceptually impossible for S to remember E later;

- (2) Every conscious state is such that it is conceptually possible for there to be some later time at which its subject remembers its occurrence; therefore,
- (3) Every conscious state is such that its subject is aware of it at the time of its occurrence.

Note that a further modification is introduced by Kriegel concerning the modality scope in P2, to deal with the objection that there could be creatures who are capable of conscious experience but are so short-lived (their life has the same duration as their instantaneous experience), that it is conceptually impossible for them to recall the experience at a later time, just because they are such that there *is no* later time in which they could recall it. To overcome this objection, and to make the argument more precise, the scope of modality in P2 is made explicit. What is claimed by P2 is *not* that for any subject S there is a time  $t_1$  such that it is conceptually possible for S to recall M at  $t_1$ . Rather, that for any subject S it is conceptually possible for there to be a time  $t_1$  such that S recalls M at  $t_1$ . This is plausible and it accommodates the short-lived creature objection.

By reframing the premises consistently with the rest of this paper, we have:

ARGUMENT 7

- P1. For any subject S, event E occurring at time  $t$ , and time  $t_1$  such that  $t < t_1$ , it is conceptually impossible for S to episodically remember E at  $t_1$  if S is not aware of E at  $t$ .
- P2. For any conscious state M of S at  $t$ , it is conceptually possible for there to be a time  $t_1$  (where  $t < t_1$ ) such that S episodically remembers M at  $t_1$ .
- C. Necessarily<sub>m</sub>, for any conscious state M of S at  $t$ , S is aware of M at  $t$ .

The premises of this deductive argument are fully based on what is *conceivable* about memory and subjects of consciousness—they do not contain any empirical claim, but only a priori considerations. They apply not only to subjects in the *actual* world, but also to subjects in *any* other possible world. Therefore, they apply to all conceivable (or metaphysically possible) subjects of experience. The conclusion is that in every possible world, for any subject S and conscious state M of S at  $t$ , S is aware of M at  $t$ . Of course, *unconscious* mental states can exist without their subject's being aware of them. So, inner awareness is necessary not for a state to be *mental*, but for a mental state to be *conscious*: necessarily, for any subject S and mental state M of S at  $t$ , if M is conscious, S is innerly aware of M at  $t$ . Since, as noted, the argument's premises are fully a priori, the argument supports NECESSITY<sub>m</sub>. Accordingly, the conclusion is a metaphysical necessity claim about all conscious experiences.

A final step takes us from NECESSITY<sub>m</sub> to GROUNDING. The underlying consideration was already mentioned in §1: plausibly, if inner awareness is a *necessary* feature of consciousness this is because it is an *essential* feature of consciousness. But being an essential feature of consciousness amounts to being *constitutive* of it—i.e., being (part of) what makes something a conscious state at all. If so, we get to the strongest thesis about inner awareness:

- IA<sub>m</sub>: Necessarily<sub>m</sub>, for any mental state M of a subject S at  $t$ , if M is conscious, M is conscious in virtue of S's being innerly aware of M at  $t$ .

I do not expect the reader to be on board with all the further argumentative steps toward IA<sub>m</sub>. However, I hope to have at least provided sufficient motivation to support the minimal and weaker thesis I have aimed to defend, that is:

IA<sub>m</sub>: For any conscious state M of a neurotypical human adult S at  $t$ , S is innerly aware of M at  $t$ .<sup>7</sup>

---

<sup>7</sup>This paper benefitted from very helpful discussions with Davide Bordini, Arnaud Dewalque, and Uriah Kriegel. I am particularly grateful to Uriah Kriegel for generous and extensive comments on a previous draft.

## References

- Armstrong, David M. 1968. *A Materialist Theory of the Mind*. New York: Humanities Press.
- Carruthers, Peter. 2000. *Phenomenal Consciousness: A Naturalistic Theory*. Cambridge: Cambridge University Press.
- Dretske, Fred. 1993. "Conscious Experience." *Mind* 102 (406): 263–83.
- . 1995. *Naturalizing the Mind*. Cambridge, MA: MIT Press.
- Fernández, Jordi. 2006. "The Intentionality of Memory." *Australasian Journal of Philosophy* 84 (1): 39–57.
- Frank, Manfred. 2015. "Why Should We Think That Self-Consciousness Is Non-Reflective?" In *Pre-Reflective Consciousness*, edited by Sofia Miguens, Gerhard Preyer, and Clara Morando Bravo. London: Routledge.
- Ganeri, Jonardon. 1999. "Self-Intimation, Memory and Personal Identity." *Journal of Indian Philosophy* 27 (5): 469–83.
- Gennaro, Rocco J. 1996. *Consciousness and Self-Consciousness*. Amsterdam: John Benjamins Publishing.
- Goldman, Alvin I. 1970. *A Theory of Action*. Princeton: Princeton University Press.
- Harman, Gilbert. 1990. "The Intrinsic Quality of Experience." *Philosophical Perspectives* 4: 31–52.
- Kellner, Birgit. 2010. "Self-Awareness (*Svasamvedana*) in Dignāga's *Pramāṇasamuccaya* and *-Vṛtti*: A Close Reading." *Journal of Indian Philosophy* 38: 203–31.
- Kim, Jaegwon. 1993. *Supervenience and Mind: Selected Philosophical Essays*. Cambridge Studies in Philosophy. Cambridge: Cambridge University Press.
- Kriegel, Uriah. 2009. *Subjective Consciousness: A Self-Representational Theory*. Oxford: Oxford University Press.
- . 2019. "Dignāga's Argument for the Awareness Principle: An Analytic Refinement." *Philosophy East and West* 69: 143–55.
- Kriegel, Uriah, and Kenneth Williford. 2006. *Self-Representational Approaches to Consciousness*. Cambridge, Mass.: MIT Press.
- Lycan, William G. 1996. *Consciousness and Experience*. Cambridge, MA: MIT Press.
- Perrett, Roy W. 2003. "Intentionality and Self-Awareness." *Ratio* 16 (3): 222–35.
- Rosenthal, David M. 1997. "A Theory of Consciousness." In *The Nature of Consciousness: Philosophical Debates*, edited by Ned Block, Owen J. Flanagan, and Güven Güzeldere, 729–53. Cambridge, MA: MIT Press.
- Siewert, Charles. 2004. "Is Experience Transparent?" *Philosophical Studies* 117 (1–2): 15–41.
- Stoljar, Daniel. forthcoming. "Is There a Persuasive Argument for an Inner Awareness Theory of Consciousness?" *Erkenntnis*.
- Thompson, Evan. 2011. "Self-No-Self? Memory and Reflexive Awareness." In *Self, No Self?: Perspectives from Analytical, Phenomenological, and Indian Traditions*, edited by Mark Siderits, Evan Thompson, and Dan Zahavi. Oxford: Oxford University Press.
- Tulving, Endel. 1972. "Episodic and Semantic Memory." In *Organization of Memory*, edited by Endel Tulving and Wayne Donaldson, 382–404. New York: Academic Press.
- Tye, Michael. 1995. *Ten Problems of Consciousness: A Representational Theory of the Phenomenal Mind*. Cambridge, MA: MIT Press.
- Van Gulick, Robert. 2004. "Higher-Order Global States (Hogs): An Alternative Higher-Order Model of Consciousness." In *Higher-Order Theories of Consciousness: An Anthology*, edited by Rocco J. Gennaro. Amsterdam: John Benjamins Publishing.