

The Lamp that Illuminates Itself: Consciousness and Inner Awareness¹

Nico Silins

In Loving Memory of Yale-NUS College

Shortened version for MEW, including material from “The Conscious Theory of Higher-Orderness”
NS

Introduction

When you sit down on the park bench and realize to your dismay that the bench was wet, your feeling of wetness is a phenomenally conscious mental state. What makes it a conscious mental state? According to higher-order theories of consciousness, we can explain why your mental state is conscious in terms of your awareness of the state. This approach has a long history, with important and visible proponents both in philosophy and psychology.²

I agree with the high-order theorist that, whenever we are in a conscious mental state, we are aware of it, or so to speak have *inner awareness* of it (Necessary Awareness). But Necessary Awareness is only a correlational claim, holding that consciousness is sufficient for inner awareness, but leaving the explanation of this correlation open. My proposal is that, once we

¹ Thanks to Jack Beaulieu, Amit Chaturvedi, Monima Chadha, David Chalmers, Nilanjan Das, Jay Garfield, Anna Giustina, Shao-Pu Kang, Malcolm Keating, Uriah Kriegel, Edvard Meza, Matthias Michel, Carlotta Pavese, Adriana Renero, David Rosenthal, Mark Siderits, Jonathan Simon, Declan Smithies, Daniel Stoljar, Lu Teng, and Alex Watson. Thanks also to the participants in my spring 2024 philosophy of mind seminar and audiences at Kobe University, Vin University, and the Université de Montréal.

² See e.g. Rosenthal 2005, Kriegel 2009, or Lau 2022, as well as Locke's *Essay Concerning Human Understanding*.

accept Necessary Awareness, we should explain inner awareness in terms of consciousness.

When you are in a conscious mental state M, you are aware of M, but you are aware of M because M is conscious, not vice versa.

My view converges with an ancient (and contested metaphor) accepted by many Buddhist philosophers, on which consciousness is a lamp that illuminates itself. My main project here is to unpack this metaphor with its most philosophically rich and defensible reading. In doing this work, I will offer exegetical options to interpreters of classical Indian and Tibetan philosophy, as well as theoretical options to investigators of consciousness in any tradition.

As the 7th century CE Indian Buddhist philosopher Dharmakīrti says,

As an illuminating lamp is considered to be the illuminator of itself, because of its nature, just so is awareness aware of itself (translated in Ganeri 2021: 31)

With the specification that consciousness is self-illuminating "because of its own nature", Dharmakīrti suggests that there is an explanatory connection between consciousness and inner awareness, with consciousness coming first in the explanatory order. Here we have a dramatic departure from theorists such as Rosenthal and Kriegel, since those theorists defend the reverse explanatory claim that inner awareness accounts for consciousness.

We can find similar claims to Dharmakīrti's in other sources. For example, the Tibetan Buddhist philosopher Mipham writes in his extensive commentary on Śāntarakṣita that,

it [consciousness] is, by its own nature, autocognizant, or self-aware, and self-illuminating (2005: 201)

To put the point in contemporary terms, consciousness is 🔥. Since Rosenthal and others in effect instead hold that consciousness is illuminated by inner awareness, we could illustrate their view as follows: 😊. Consciousness is moon.

My first step will be to clarify my reading of the metaphor of consciousness as a lamp that illuminates itself. I will go on to contrast my reading with rival takes on the explanatory relation between consciousness and inner awareness, and to defend it against those rivals. I close by responding to several important objections to my own position.

Throughout the paper there will be no strict division between doing philosophy and doing history of philosophy---each one opens important possibilities for the other. We will learn about logical space by thinking about positions from the past, and we will open interpretative options for the past by thinking about logical space. For example, we will see that inner awareness need not be achieved by looking inside, and that reflexive and non-reflexive views of consciousness need not conflict.

Section 1. Set Up

I'll begin by working through the following claim about a relation between consciousness and inner awareness:

(Necessary Awareness): For any conscious state M and subject S, if S is in M, then S is aware of M.³

³ While our formulation does not make this explicit, it does require that when you are in a conscious mental state at a certain time, you are aware at that very time of the mental state.

If consciousness is a lamp that illuminates itself, at least this much must be true. Clarifying Necessary Awareness will be a useful way to set up the key terms of our debate. We can then turn to examine what else must be true if consciousness is a lamp that illuminates itself.

Necessary Awareness tells us that there is a necessary correlation between consciousness and inner awareness, and by itself not much more than that. As a stipulation, I will speak very broadly of any awareness or representation of one's own conscious states by oneself as "inner awareness." Much remains to be clarified about what inner awareness is and what inner awareness does. In particular, we should say more about the object of inner awareness, the character of inner awareness, and the options for more specific claims about the relation between consciousness and inner awareness. I'll now take up these questions in turn.⁴

As far as the **object** of awareness is concerned, I am simply taking it to be the mental state M itself. Alternative targets of awareness include for example oneself, or the fact that one is in a given mental state (further candidates are laid out in Nida-Rümelin 2017 and Stoljar 2021).

I'm less clear on the status of a further option widely discussed in the literature. As Zahavi puts it when interpreting Husserl:

prior to reflection one perceives the intentional object, but one experiences [erlebt] the intentional act. Although I am not intentionally directed at the act (this only happens in the subsequent reflection, where the act is thematized), it is not unconscious but conscious, that is self-given (Zahavi 2003: 159).

So if a gardener is seeing a tree, the tree is an intentional object of his perception, and he is in some way aware of his seeing of the tree, yet his seeing of the tree is not any way an intentional

⁴ For useful further set up, see Farrell and McClelland 2017, Stoljar 2021, or Giustina 2022.

object of his awareness. To assess such claims, we need to be clear about what it takes for something to be the object of a mental state. When I say "object" here, I simply have in mind whatever your awareness is awareness of, not in any way assuming something stronger, as would for example a sense of "object" implying non-sentience. In my fairly undemanding use of "object", there is a straightforward contradiction in saying "that awareness of a mental state does not have any object". In other words, a so-called "nonthetic" state of awareness of something does have an object if it involves awareness of something.

Now, one might use "object" in a more demanding way, for example assuming that it is impossible for a mental state to be its own object. Zahavi clearly is making such an assumption here:

one should avoid theories that describe self-awareness as a kind of relation---be it a relation between different experiences, or between the experience and itself---since every relation, especially the subject-object relation, presupposes a *distinction* between two (or more) relata (2003: 158)

Zahavi should be applauded for putting his cards on the table, but I don't think these cards are very good. Here I would deny that we have a good theoretical use for such a demanding sense of "object", especially since it is simply false that every relation holds only between distinct relata. The relation of being exactly as tall as holds between two "identical" twins, and between each of themselves as well.

A further question concerns the **character** of inner awareness itself. We can ask whether inner awareness is a form of perception, a form of thought such as belief or knowledge, a form of acquaintance, or some unique relation of its own not present elsewhere. Here I will leave our options open to be able to engage with as many thinkers as possible across the debate.

There is an important cautionary note to make about the term "inner awareness". The awareness is of a mental state that is in some sense an "inner" state, but do not assume that such awareness must be achieved by a literal or figurative form of looking inside. Whether inner awareness is perceptual or not, it might be obtained via outer perceptual or non-perceptual investigation. Views with such a structure are developed by Evans (1982) and others, often going under the heading of "transparency". According to Evans, for instance, we may gain knowledge of our own perceptual states by bracketing our background "extraneous" information about how the world is, coming to an assessment via perception of how the world is, and then prefix that assessment with "it seems to me as though..." For example, bracket your background information that the object bobbing in the lake is a decoy duck, and consider that in the absence of having such information you would judge there to be a duck over there, so conclude that it (perceptually) seems to you that there is a duck over there.⁵ Now Evans did not himself consider this method as any way of coming to awareness of anything inner:

[W]hat we are aware of, when we know that we see a tree, is *nothing but a tree*. In fact, we only have to be aware of some state of the world in order to be in a position to make an assertion about ourselves (1982: 231).

Here Evans rejects the claim that we are aware of seeing a tree at all. But we can read Evans---arguably more charitably---only as excluding all forms of perceptual awareness, leaving it open for us to say that his method is a route to cognitive inner awareness. (I would say there is quite a tension in "I know that I'm seeing a tree, but I'm not aware of seeing a tree").

For perhaps the most striking illustration of "inner awareness" achieved by looking outside in the context of Indian philosophy, consider the radically idealist Yogācāra. This

⁵ How to develop the view further is open to discussion, with efforts to cash out "extraneous" in Peacocke (2008), and a rival account in Byrne (2018).

approach can agree with the letter of what Evans said: what you are aware of, when you see a tree, is indeed nothing but a tree. Yes, but what is a tree?! Here the tree is in some way an aspect of your mental state, and you are aware of your mental state by being aware of the tree (which is not yet to say that you are aware of it as being mental).⁶ Again, the term of "inner" awareness should therefore be taken with much caution, since your awareness of inner mental states might come through at least putatively outward looking means.

The final and crucial question I will address here concerns the **vehicle** of inner awareness. Many will say that, when you are in a conscious state M, and have inner awareness of M, M is itself the vehicle of your inner awareness of M. Sample proponents of this view include Dignāga, Dharmakīrti, and Kriegel. Here I will say that each conscious mental state M is "reflexive" or that inner awareness is "reflexive". (Bearing in mind we are not saying in any strong sense that a mental state can be an "object" for itself). Such philosophers endorse at least the following sort of claim, if not more as well:

(Necessary Reflexive Awareness): For any conscious state M and subject S, if S is in M, then S is aware of M, and M is a vehicle of S's awareness of M.

In framing the claim in terms of vehicles, I am avoiding building into the thesis that M is in some sense the subject of awareness of itself. The view could be developed this way, but does not have to be developed this way. Here I will continue to speak as if S is around to be a subject of awareness of its own conscious mental states, even if a more complete Buddhist analysis would purge such talk. We could even avoid implying any subject of awareness at all, for example if

⁶ For further discussion of how to spell out Buddhist idealist views here, see Williams 1998: 33-4, Finnigan 2018 and Watson 2022, esp. 147-8.

we simply say that M represents itself. Compare how a billboard could say “this space available” without being a subject.

A classic way to argue for Necessary Reflexive Awareness is via some form of regress argument. As Williams summarizes:

If it is necessary to know a consciousness by a further consciousness, then there would be an infinite regress. . . It [consciousness] has the nature of luminosity . . . [it has] self-luminosity to itself of also its own nature without needing to depend on another . . . (1998: 93)

In other words, unless Necessary Reflexive Awareness is true, we might have to say that each conscious state is the object of a distinct conscious state. Here we would have an endlessly bubbling chat of conscious states and their successors, the texts would never stop.

We should not conflate Necessary Reflexive Awareness with the rejection of any form of hierarchy of inner awareness. The presence of reflexive inner awareness does not entail the absence of higher-order inner awareness. There is room for views on which all conscious states come along with reflexive awareness, where we also have higher-order awareness of some or many of our conscious mental states. It could even be that every conscious mental state is the vehicle of awareness of itself, and at the same time every conscious mental state M comes along with a distinct vehicle of awareness of M. Call this view Necessary Double Awareness. To illustrate:



fig.2. Necessary Double Awareness

To make the same point in a different way, Necessary Reflexive Awareness tells us that each conscious mental state is **a** vehicle of awareness of itself, the claim does not yet say that each conscious mental state is **the only** vehicle of awareness of it. Necessary Reflexive Awareness leaves each conscious mental state comes along with at least one buddy aware of it too.

My guess is that we missed this possibility because of an overestimation of the strength of regress arguments in this context. I suspect we have tended to conflate the acceptance of reflexive requirements for consciousness with endorsement of regress objections to second-order awareness requirements on consciousness.

But regress objections and reflexive views need to be much more sharply distinguished. A major issue here is that regress arguments for reflexive views are far from decisive.⁷ Even if a regress argument did refute the claim that every conscious state comes along with inner awareness by some distinct state, it is a leap to conclude that every conscious state comes along with reflexive inner awareness (see e.g. Ganeri 2012: 177). The most we could so far conclude is that some conscious mental states come along with reflexive inner awareness.

To see the problem, compare the structure of classical regress arguments for foundationalism in epistemology. If every justified belief is justified by a distinct justified belief, we threaten to have a vicious regress of justification. To block the regress you might affirm that there are some foundationally justified beliefs not justified by distinct beliefs. But it would be a leap to conclude that all justified beliefs are foundationally justified (it would be counterproductive too since we would end up only with foundations and no superstructure).

⁷ For further evaluation of regress arguments as used in the context of debates about consciousness and inner awareness, see e.g. Kellner 2011, Stoljar 2018, Das 2022: sec. 4.1., or Garfield 2022: 70-9.

Using a regress argument to conclude that all conscious states are reflexively self-aware does not yet seem to fare any better.⁸

There is a way to try to respond. As Kellner puts it,

... Dharmakīrti makes clear that the regress cannot be stopped by assuming that one arbitrarily chosen final cognition in the chain is self-establishing, and makes use of this possible choice of his opponent to strengthen his own claim: since there is no cause for cognitions to be different from one another in this respect, if one cognition is self-establishing, all cognitions have to be (2011: 423).

A shifting stopping point across cognitions might indeed be arbitrary, resulting in a kind of unplanned urban skyline of ascents of awareness, some stopping at level 2 and others going to 90 or higher. But the critic of regress arguments need not endorse a shifting standard to block infinite ascent, and it's not yet clear what would be arbitrary about excluding reflexivity from our ground floor conscious states, and allowing it only at some fixed higher-order level above them.

In any case, further major hurdles for regress arguments remain. First, as Rosenthal 2005 in effect points out, perhaps consciousness does always come along with inner awareness, but where that inner awareness need not itself be conscious. So Necessary Double Awareness does not entail any infinite ascent. Second, perhaps we do have an infinite ascent of 🔥, but the regress is not vicious (see Kriegel 2012 for discussion of this option). As perhaps the most extreme option, Necessary Double Awareness is compatible with a structure in which each

⁸ Considering regress arguments for foundationalism reveals further options for the debate about consciousness and inner awareness. What about circles of inner awareness? In its tightest form, mental state 2 would be a case of inner awareness of mental state 1, and mental state 1 a case of inner awareness of mental state 2 (This possibility is mentioned in Stoljar 2018 but not pursued further there). Or what about analogues to coherentism, where a conscious mental state is not the object of inner awareness of 1 distinct state, but instead somehow of a field of distinct states?

conscious mental state comes along with an infinite ascent of conscious mental states, each of which is also reflexively self-aware!

Regress arguments do not by themselves establish Necessary Reflexive Awareness, and might fail to establish anything about consciousness and inner awareness. The case for Necessary Reflexive Awareness will need to come from somewhere else, and might leave room for Necessary Double Awareness.

Is there any positive support for Necessary Double Awareness? Here it might help to remember that regress arguments are not the only potential route to reflexive forms of inner awareness. Different strategies could proceed through phenomenological reflection, conceptual analysis, or other forms of a priori reflection on what it takes for a state to be conscious. Also, once we get more fine-grained about potential targets of inner awareness, it becomes more plausible that there is a need for us to have both reflexive and non-reflexive forms of inner awareness. For example, perhaps reflexive inner awareness is best suited to grasp a conscious mental states' being mine, or alternatively having any conscious character at all, where here reflexive inner awareness would be invariant across my conscious mental states (see Kriegel 2009: 181-190). Assuming that we invariably have inner awareness of our conscious states' being mine, or being conscious period, it would follow that they all come along with reflexive inner awareness. On the other hand, as our conscious mental states vary in terms of what they represent or are of, perhaps varying states of higher-order inner awareness are better suited to grasp variations of what our conscious mental states represent or are of.⁹ Assuming that we invariably have inner awareness of what our conscious states represent or are of, we would then have a case for them all coming along with higher-order awareness. Putting the two threads

⁹ For sample surveys of the rich range of candidate targets of inner awareness here, see Ganeri 2012 or Howell 2023.

together, perhaps a case for Necessary Double Awareness could be built from the need for different forms of inner awareness to capture different aspects of conscious states.

Section 2. The Lamp

We've now worked to clarify the object, character, and vehicle of inner awareness. More could be done, but we should have enough to enable us to understand Necessary Awareness and stronger claims such as Necessary Reflexive Awareness. Necessary Awareness and Necessary Reflexive Awareness tell us that there is a correlation between consciousness and inner awareness, but they do not tell us anything about what might explain that correlation, or whether anything does at all. Let's get back to the metaphor of consciousness as a lamp that illuminates itself. I will now try to provide the most philosophically defensible interpretation of the metaphor for consciousness as a lamp that illuminates itself.

To my ear, the metaphor holds that consciousness explains its correlation with inner awareness. In particular, I propose to read the metaphor as follows:

(Self-Illuminating Consciousness): For any conscious state M and subject S, if S is in M, then S is aware of M in virtue of the fact that M is conscious, where a vehicle of that awareness of M is M itself.

Here I am building reflexivity into our reading of the metaphor. The addition of reflexivity is not redundant. In principle, it could be that you only have higher-order inner awareness of M in virtue of M's being conscious. Reflexivity will be largely on the backburner, so to speak, in what follows, but we will return to it when we address objections. Note also that we have not ruled

out claims along the lines of Necessary Double Awareness, although claims with that structure won't be in the spotlight in what follows.

Let's now explore Self-Illuminating Consciousness and cognate claims in greater depth. As far as the exegesis of classical Buddhist philosophical texts is concerned, I will focus on the following key passages:¹⁰

Dharmakīrti:

As an illuminating lamp is considered to be the illuminator of itself, because of its nature, just so is awareness aware of itself¹¹

Mipham:

it [consciousness] is, by its own nature, autocognizant, or self-aware, and self-illuminating (2005: 201)

Mipham:

The nature of consciousness, on the other hand, is unlike matter. For it to be known, it depends on no condition other than itself. Therefore, it is perfectly acceptable to say that it is self-cognizant. In the very instant that consciousness arises, the factors of clarity and knowing are present to it. And although other things are known by it, it is not itself known by something else and is never without self-awareness (it is never “self-unaware”). It is like a boatman taking himself over the river at the same time as his boat, or the lamp that is self-revealing as it shines. Although it renders a pot visible in a dark room, it has no need of another source of light for itself to be seen. This is what is meant by the expression “self-illuminating lamp.”... Broadly speaking, and by its nature, every consciousness entails self-cognition (2005: 202-3).

¹⁰ I'm not sure whether there are contemporary Western versions of the claim beyond the one I defend in my forthcoming. For the closest parallels I have been able to find, see section 4.2 of Chalmers 2003, Chalmers 2013: 345, section 6 of Nida-Rümelin 2017, and Stoljar forthcoming. I don't count figures such as Husserl and Sartre as parallels given that they formulate their claims in terms of “nonthetic” awareness that allegedly has no object.

¹¹ As translated in Ganeri 2021: 31.

If their claim were merely that a state's being conscious entails awareness of that state, there would be no need for any language about what consciousness does "because of its nature" or "by its nature". It would instead suffice to say here for Dharmakīrti to say that "awareness is aware of itself" and for Mipham to say that "it [consciousness] is self-aware." Done. But these quotes do much more. The language of the "nature" of consciousness introduces questions about the respective explanatory priority of consciousness and inner awareness. Now once we factor in the metaphor of the lamp illuminating itself, where it is the lamp that is explicitly the source of the light that illuminates itself, it also seems clear that consciousness has explanatory priority over inner awareness of conscious states. Here your conscious state makes you have inner awareness of your conscious state.¹² As Dignāga puts it at one point, "self-cognition must be admitted. It itself is a result [of the act of cognizing] (Hattori 1968: 31)".

I should flag that my reading of our key passages is hardly beyond doubt. For one thing, I have so far omitted an awkward bit of the Mipham. The fuller version of the third passage I quoted reads as follows:

Although it renders a pot visible in a dark room, it has no need of another source of light for itself to be seen. This is what is meant by the expression "self-illuminating lamp." **It does not mean that the lamp makes itself visible as its own object...** Broadly speaking, and by its nature, every consciousness entails self-cognition (2005: 202-3, emphasis mine)

I think we have returned to the wrinkle about uses of "object" we already encountered with Zahavi 2003, and I take Mipham to use the term in a demanding way that speaks past my own undemanding use of "object".¹³ Mipham does overstep when he says that "although other things

¹² Here it may be that I part exegetical ways with Ram Prasad 2007 and MacKenzie 2017, who seem to frame their discussions as being about the nature of inner awareness rather than the nature of consciousness. It remains possible that we differ merely in emphasis rather than substantively.

¹³ For further relevant discussion, see also Watson 2014, esp. 412-415. He surveys a view which

are known by it [consciousness], it is not itself known by something else". Again, we need not reject higher-order awareness when we affirm reflexive inner awareness.

To get a feel for what Mipham really cares about, consider the following sort of passage:

when it is said that consciousness is self-knowing, this is not meant in the sense of an ax chopping wood. It does not mean that consciousness apprehends itself as something really other than itself, or that consciousness as the subject and consciousness as the object of the act of cognition are being considered as real and separate entities. To know is simply the nature of consciousness, and for this reason it is acceptable and correct to consider that consciousness is autocognizing (Mipham 2005: 202).

Since my version of reflexivity holds that a conscious mental state is a vehicle of inner awareness of itself, I am hardly holding that consciousness is somehow something other than itself here.¹⁴

"concluded from 2, i.e. the fact that light is not illuminated by another light, not . . . that it is illuminated by itself, but rather that

(4) Light is not illuminated at all; it shines forth by itself...For them light appears (*prakāśate*) by itself, but never as the object of illumination (Watson 2014: 412-413).

I'm not clear on whether there's a substantive disagreement between this view and my own, there may again simply be an equivocation over terms such as "illuminate" and "object". Consider how Watson elaborates the claim that light appears by itself: "i.e. without requiring any means other than itself." (Watson 2014: note 31). When the claim is put that way, I do not have any disagreement with it. All that said, it may be safer to formulate the claim in terms of light sources rather than light. One might deny that light is visible at all, but it is much harder to deny that a light source can illuminate itself without requiring any means other than itself.

¹⁴ For potential rejection of my reading, see Williams: "Self-consciousness here ... essentially does *not* involve in itself consciousness *of* anything ... (1998: 31-2)". At the same time, Williams seems to contradict himself by also describing the relevant form of inner awareness as follows: "what is by nature not insentient is thereby said to be conscious of its very own self (1998: 26)." Williams seems to assume that it is impossible for a mental state to be conscious of itself without the mediation of a distinct mental state. As he puts it, "the character of self-awareness here has nothing to do with taking itself as an *object* in a way which might lead to an infinite regress (1998: 28)." But a mental state can have itself as an object without generating any infinite regress. (Williams has a further argument at pp. 30-1 appealing to the possibility of pure consciousness that is not conscious of anything, but I take the existence of such a form of consciousness to be too controversial to give enough traction in the debate.)

I'll now move on from exegesis. To bring out the philosophical plausibility of Self-Illuminating Consciousness, I will defend the claim that it gives the best explanation of the truth of Necessary Awareness.

The standard case for Necessary Awareness appeals to the sheer absurdity of being in a conscious mental state without being aware of it. The key idea, often expressed as the "Transitivity Principle", is put forth as being platitudinous in many places by David Rosenthal among others:¹⁵

There is a natural way of understanding how conscious states differ from mental states that are not conscious. No mental state is conscious if the individual that is in that state is in no way aware of it. If somebody thinks, desires or feels something but is wholly unaware of doing so, then that thought, desire or feeling is not a conscious state (2012: 1425).¹⁶

Here I agree with Rosenthal and any others making this claim.¹⁷ But note that this only gets us to Necessary Awareness. The "Transitivity Principle" does not entail any theory or combination of theories that would explain a state's being conscious in terms of our awareness of it. The common move to higher-order theories from the putative platitude is an entirely optional leap.

Now I will take it for granted that consciousness is necessarily accompanied by inner awareness, the question is why.

For further criticisms of non-intentional readings of inner awareness, see Garfield 2015 or 2022: 91-7.

¹⁵ For a nice range of references to Brentano, Sartre and others, as well as further arguments of his own for Necessary Awareness, see Janzen 2008 ch. 4.

An important further line of argument for Necessary Awareness comes out of the Buddhist philosopher Dignaga's memory argument. See e.g. Ganeri 1999, Garfield 2006, Kellner 2010, Thompson 2010, Chadha 2017, Kriegel 2019, Giustina 2022a (who also offers interesting discussion of how to potentially bridge the gap between Necessary Awareness and higher-order theories)

¹⁶ For further examples see Rosenthal 2005: 103, 109, 145.

¹⁷ For dispute, see e.g. Dretske 1993, Thomassen 2006, Gertler 2012, Siewert 2013 or Stoljar 2021. For a line of response to some of their objections, see Giustina 2022b.

I'll start by opposing higher-order theories of consciousness which hold that when a mental state is conscious, it is conscious because we have inner awareness of it. These theories face major objections entirely ineffective against Self-Illuminating Consciousness¹⁸. As a result, Self-Illuminating Consciousness provides a much more economical explanation of Necessary Awareness.

First, according to the “rock” objection, our awareness of a rock is not able to make the rock conscious, and by the same token our awareness of our own mental states should not be able to make them conscious either. Even as we consider various more specific forms of awareness, e.g. cognitive or perceptual, or direct or indirect, each form of awareness seems extendable to objects that are not conscious, thereby raising the question of how any form of awareness could endow a mental state with the status of being conscious. The point seems especially clear when we remember that we are focused here on phenomenal consciousness, a matter of what it's like to be in our mental states. If our focus was instead only on access consciousness, it would instead be trivial that being aware of a mental state can make it access conscious. Access consciousness is after all defined in terms of the availability of a mental state to awareness or report of the mental state.

Now, one might respond to the rock objection with the fine print that that awareness can make something conscious only when that something is a mental state. I take such restrictions to be ad hoc, we need a case for why the exception is being made here. There is also the possibility that many forms of awareness could be extended to a mental state without that mental state yet becoming conscious (more on that soon).

¹⁸ For useful overview of these classic objections with further references, see Carruthers and Gennaro 2020: sect. 7. I develop my argument against higher-order theories in more detail in my forthcoming.

Now, the "rock" objection is a non-starter against Self-Illuminating Consciousness. The view does not assign any capacity of creating consciousness to inner awareness. After all, on this view consciousness explains inner awareness rather than vice versa. Endorsing the view thus permits us to smoothly avoid the rock problem for higher-order theories.

Similar considerations unfold for the important case of the Misrepresented State Objection. Here the objection is that, since higher-order representations can misrepresent in various ways, we should not expect them to suffice for the presence of consciousness. For example, presumably it is possible for us have higher-order representations when no conscious state putatively targeted is present at all, a sort of illusion of a conscious mental state.

In response, one might insist that "awareness" is a success term, so that it is impossible to have awareness of a mental state M unless it is present. We can take this point on board. The objection is in effect that it is possible to be aware of a mental state as conscious when the mental state is not conscious. The objection does not need the possibility of a pure hallucination of inner awareness where no mental state targeted is present at all. Now, one might try to further fine-tune the higher-order thesis to the claim that, if you are successfully aware of a mental state M as conscious, then M is conscious. But this claim is trivial.

Perhaps surprisingly, the Misrepresented State objection is irrelevant to Self-Illuminating Consciousness. The direction of our claim is from conscious mental states to awareness or representation of conscious mental states, not from higher-order awareness or representation to conscious states. So Self-Illuminating Consciousness only makes predictions about cases in which a conscious mental state is already present. The view is therefore entirely compatible with the possibility of illusions of conscious states. To illustrate the point with an analogy, consider how to test the claim that Anna always announces herself as present when she is present. You

only need to check for cases in which Anna is present and yet does not announce Anna as present. It is irrelevant whether Billy ever plays a prank and announces Anna as present when she is not present. Going back to our central metaphor, even if consciousness is a lamp that illuminates itself, a matching form of light could sometimes come from elsewhere.

In sum, Self-Illuminating Consciousness explains Necessary Awareness while avoiding the excesses of higher-order theories. That said, even if we set higher-order theories aside, there are further rival options to consider.

One approach would be to accept Necessary Awareness, but to deny that there is any explanatory priority between consciousness or inner awareness. Perhaps we could find some third factor that explains the correlation between consciousness and inner awareness, in principle they could be the joint outputs of some explainer they have in common. While I find this option intriguing, I do not know of any particularly promising candidates to consider as ways of fleshing it out. You might suggest that some physical substrate could do the job. In response I observe we could instead trace an explanatory chain in which some physical substrate explains consciousness, and consciousness in turn explains inner awareness. We would need some reason to prefer the structure on which consciousness and inner awareness are joint independent outputs of a common source, but I don't see how to make any such case effectively.¹⁹

A more promising way to deny explanatory priority between consciousness and inner awareness is to identify them.²⁰

¹⁹ Thanks to Daniel Stoljar for an interesting further suggestion. Perhaps consciousness and inner awareness are interdefined, just as belief and desire are interdefined on some functionalist approaches. Neither consciousness nor inner awareness has priority over the other, but they are not identical either. My tentative response is that such a view offers too loose a link to explain the tight connection involved in each conscious state necessitating inner awareness of it.

²⁰ Here I'm considering forms of explanation on which, if a explains e , and $a=b$, then b explains e . I'm also remaining neutral about potential differences between type-type and token-token identity theories here.

As an exegetical matter, I would again emphasize that for Mipham, consciousness is "by its own nature" self-aware. If there were simply an identity here, there would be no need for such a qualification, we should instead just say that consciousness is self-awareness. For example, given that Clark Kent is identical with Superman, we say simply "Clark Kent = Superman", there is no need to say, "Clark Kent is by his own nature Superman."

As a further philosophical challenge to the identity view, we can return to the possibility of illusory inner awareness. Consider Anton syndrome, where subjects who are in fact blind insist they are not, offering detail about what they are putatively (consciously) seeing. Perhaps this is specifically a case in which the subject is aware of a mental state as conscious even when the mental state is not. Alternatively, recall that for our purposes, unconscious perception is perception that is phenomenally unconscious, never mind whether it is unreportable. There seems then to be logical space for subjects who can report unconscious perception, as with Block's hypothetical cases of 'super-duper-blindsight' (1995: n. 16, 246). Such subjects have the ability to report their states of (phenomenally) unconscious perception, and even to do so without the mediation of inference. They arguably then have inner awareness of their perceptual state, even while it remains unconscious.

Now, these cases are by no means conclusive, we can debate what is actually happening in them. My point is that these cases do pose a threat to the identity view of consciousness and

One might try to read Dharmakīrti as endorsing an identity view, but the move is tricky considering Dharmakīrti's emphasis on "non-difference" rather than sameness. Arguably his view is only that consciousness and inner awareness are necessarily interdependent in some way that falls short of being one and the same. For more detail on this point, see e.g. Taber 2010: 292-3.

inner awareness, whereas they are irrelevant to Self-Illuminating Consciousness.²¹²² Again, Self-Illuminating Consciousness is not committed to any non-trivial entailment of consciousness by inner awareness. So the view can comfortably brush aside potential cases of inner awareness in the absence of consciousness.

I'll now move on from the view that consciousness and inner awareness are identical. A final option is that nothing explains the correlation between consciousness and inner awareness, perhaps it is simply a brute fact that the former is always accompanied by the latter. Given that we have seen so many rivals to this option with at least some promise, I don't think we need to worry about the brute connection view at this stage.

We've now judged a parade of rival views about the respective explanatory priority of consciousness and inner awareness. I hope to have established that Self-Illuminating Consciousness gets best in show. It does not say too much, and thereby avoids major challenges to higher-order theories and identity theories that stick their necks out too far. Self-Illuminating Consciousness also does not say too little, and thus remains rich enough to offer explanatory insight, unlike the brute connection view. All that said, there is a major caveat about how much our argument so far can establish. Given that the reflexivity component of Self-Illuminating

²¹ There is a tricky further possibility about the scope of identity claims once we bring in the distinction drawn by many Buddhist philosophers between conventional truths and ultimate truths (for a book length discussion of how this distinction bears on the debate, see Williams 1998, see also Garfield 2006). Depending on how the distinction is cashed out, I might leave open the possibility that consciousness and inner awareness are identical at the ultimate level, my primary claim would instead be that consciousness explains inner awareness at the conventional level.

²² For discussion of introspective error in the context of Indian philosophy, specifically with respect to past perceptions of absence, see Beaulieu 2021 (as well as Garfield 2015 for a more general discussion). For discussion in the context of Indian philosophy of the much more radical possibility that we seem to have conscious mental states without ever having any conscious states at all, see Garfield 2016 and Chaturvedi forthcoming.

Consciousness has not played a critical role in this section, that component of the view remains undefended. Either regress arguments need to be rescued, or we have to find some other way.

Section 3. Objections and Replies

I'll now respond to some important objections.

One line of objection is to my case in favor of Self-Illuminating Consciousness, in particular to the way that I challenged higher-order and identity theories. My complaint was that these theories make claims that are too strong, overpredicting the presence of conscious states in cases where conscious states are in fact absent. One might respond that it is no surprise that these theories should be more vulnerable than mine, given that they promise us an account of what consciousness is. The theories might have more costs but arguably they offer much more as well. In contrast, I made much of the way that Self-Illuminating Consciousness avoids making claims about how inner awareness generates or otherwise accounts for consciousness. Perhaps this silence of the view makes it too weak.

In response, I want to emphasize that Self-Illuminating Consciousness can be smoothly supplemented to give us a view that does tell us what consciousness is. The claim here would be that conscious states are uniquely characterized by the way consciousness explains inner awareness. In other words, if we ask for an account of what a conscious mental state is, we can answer that by pointing to those states whose conscious character explains inner awareness. Dharmakīrti and Mipham were presumably telling us as much when they spoke of what consciousness does “because of its nature” or “by its nature”.

To formulate the view properly requires some complexity:

(Definitive Self-Illuminating Consciousness): For any mental state M and subject S, M is conscious if and only if M is such that, if S is in M, then S is aware of M in virtue of the fact that M is conscious, where a vehicle of that awareness of M is M itself.

The Self-Illuminating Consciousness thesis is weaker since it only says that all conscious mental states make us have inner awareness of them. It didn't give any necessary condition for being a conscious mental state, and so it didn't rule out the possibility that non-conscious mental states also make us have inner awareness of them. Definitive Self-Illuminating Consciousness does rule out this possibility, by adding the claim that only conscious mental states are such that being in them suffices to make us have inner awareness of them.²³²⁴

We are not still signing up for the claim that any form of awareness non-trivially suffices for consciousness. The defining feature of a conscious mental state is actually not our being aware of the state, as is held by higher-order approaches. The key is instead much more complex. The defining feature of being a conscious mental state is being such that, if you are in the mental state, then you are aware of it in virtue of its being conscious. Here we have a further stark contrast with higher-order theories. On higher-order approaches, being conscious is a contingent feature of a mental state, and the status of one and the same mental state as conscious can wax and wane depending on whether the state is illuminated by inner awareness.²⁵

Consciousness is moon indeed: 🌙. On the present picture, any mental state that is conscious

²³ See also MacKenzie 2017: 345-7.

For a rival reading of Śāntarakṣita here, see Williams 1998: 18 or Arnold 2005 and 2010. Here is a quote that gives the flavor of Arnold's view of inner awareness: “not, that is, as itself simply another kind of intentional cognition—but as intentionality itself—that is, as what it is in virtue of which any cognition could count as a token of that type. (2005: 97).”

In contrast, I in effect am reading my key passages as describing another kind of intentional cognition.

²⁵ Compare Rosenthal 2005: 27-9.

could not have failed to be conscious. I leave open how further to adjudicate the question of whether being conscious is a contingent feature of a mental state (see Kripke 1972: 152-3 for a classic discussion on which it is not).

Now, one might demand a more robust and reductive account of what consciousness is. In particular, one might seek an explanatory factor that determines the presence of consciousness without itself specified using the notion of consciousness. Given our focus on something above consciousness determined by consciousness, we have not yet met such a demand.

If you do have that itch, it actually is possible to hold our definitional claim and scratch the itch. It is entirely coherent to combine the view with some non-higher-order account of how consciousness is determined, and then trace a chain of explanation/determination from the first-order resource, to a conscious mental state, then to inner awareness of the mental state. The chain would look like this:

Factor X $\longrightarrow$ M is conscious $\longrightarrow$ subject is aware of M²⁶

All that said, it remains up for grabs whether we should seek a robustly reductive account of consciousness. Another option would be to mirror the knowledge-first program of Williamson 2000 (see also Unger 1974). The project here is to use knowledge as an explanatory tool, rather than to try to explain knowledge in any robustly reductive way. One might likewise

²⁶ As Jack Beaulieu pointed out to me, we can find an instance of such a view in Mokṣākaragupta: “Just as a lamp illuminates itself, so, too, are states of awareness, which are distinguished from insentient things insofar as they are produced just by their causes to have the nature of illumination, established to be reflexively aware.” Kajiya (p. 48): https://archive.org/details/tarkabhashamokshakaraguptaanannotedtranslationwuichikajiyama_202003_76_a/page/n55/mode/2up

set out to explain inner awareness and other conditions in terms of consciousness, as opposed to trying to explain consciousness in terms of inner awareness or other conditions.

Either way, Self-Illuminating Consciousness can smoothly be supplemented to provide more or less reductive accounts of what consciousness is.

A further challenge here would hold there isn't even a correlation between consciousness and awareness to explain. This sort of traditional objection to higher-order theories is not targeted specifically at the explanatory ambition of higher-order theories (even though in practice it seems to be pushed without distinguishing between Necessary Awareness and the further specifics of higher-order theories).

One variant of the challenge proceeds in terms of limits on our introspective self-knowledge. Given plausible limits on our self-knowledge, perhaps there are many cases when we are in a conscious mental state without being aware of it. However, since inner awareness need not be a form of knowledge, efforts to reveal dramatic limits to our self-knowledge will not suffice to establish dramatic limits to our inner awareness. There is room for the possibility that we have inner awareness of conscious mental states even when we do not know that we are in them. Moreover, even if inner awareness does turn out to be some form of knowledge, it need not yet be a form of propositional self-knowledge, the usual target of contemporary critiques.

A more promising challenge is the over-intellectualization objection pushed by Dretske 1995 and many others. This objection uses putative cases of animals or human infants who arguably have conscious mental states without (yet) the equipment to have awareness of any of those mental states. (Here we are concerned with blindness to conscious mental states, not with hallucinations or illusions about them).

In response, we can remind the objector that, in order to over-intellectualize, you need to intellectualize in the first place. Since the relevant form of awareness of conscious mental states can come in perceptual, attentional, or sui generis forms, it's not clear why a capacity for inner awareness shouldn't be widely available across the animal kingdom, at least for any creature complicated enough to have consciousness in the first place. (A separate line of response could instead work with a more cognitively demanding form of awareness, just in a dispositional rather than non-dispositional form. I will continue to set that option aside).

Conclusion

"I'm not a historian." This phrase is sometimes used by contemporary philosophers to close off engagement with the past rather than to engage the past with epistemic humility. I'm not a historian. I lack the linguistic skills and deep contextual knowledge of the experts in Indian and Tibetan philosophy I have cited. But if we restrict discussion to the experts, contemporary analytic philosophy will never open up more thoroughly to other voices and times.

With my interpretation and defense of the Buddhist metaphor of consciousness as a lamp that illuminates itself, I hope to have helped to illustrate how contemporary analytic philosophy can benefit from opening itself up.

References

- Arnold Dan (2005) Is *Svasaamvitti* Transcendental? A Tentative Reconstruction Following Śāntarakṣita, *Asian Philosophy*, 15:1, 77-111, DOI: [10.1080/0955236052000341050](https://doi.org/10.1080/0955236052000341050)
- Beaulieu, Jack (2021). Gaṅgeśa on Absence in Retrospect. *Journal of Indian Philosophy* 49 (4):603-639.
- Byrne, A. (2018). *Transparency and self-knowledge*. Oxford: Oxford University Press.
- Chaturvedi,
- Carruthers, Peter and Rocco Gennaro (2020), 'Higher-Order Theories of Consciousness', in Edward N Zalta, ed. *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab. Stanford University. <https://plato.stanford.edu/archives/fall2020/entries/consciousness-higher/>.
- Chadha, Monima (2017) ' is Essential to Consciousness: A Buddhist-Abhidharma Perspective', *Review of Philosophy and Psychology* 8: 83-101. doi: 10.1007/s13164-015-0295-7.
- Chalmers, David. 2003. "The content and epistemology of phenomenal belief." In *Consciousness: New Philosophical Perspectives*, edited by Q. Smith and A. Jokic. Oxford: Oxford University Press.
- Chalmers, David. 2013. "The contents of consciousness: Reply to Hellie, Peacocke and Siegel." *Analysis Reviews* 73 (2): pp. 345–368.
- Chaturvedi, Amit (forthcoming). Is the Mind a Magic Trick? Illusionism about Consciousness in the "Consciousness-Only" Theory of Vasubandhu and Sthiramati. *Ergo: An Open Access Journal of Philosophy*.
- Cohen, Michael A, and Daniel C Dennett (2011) 'Consciousness Cannot Be Separated From Function', *Trends in Cognitive Sciences* 15: 358-64. doi: 10.1016/j.tics.2011.06.008.
- Coseru, Christian (2012) *Perceiving Reality: Consciousness, Intentionality, and Cognition in Buddhist Philosophy*. Oxford University Press.
- Coseru, Christian (2015) 'Taking the Intentionality of Perception Seriously: Why Phenomenology is Inescapable', *Philosophy East and West* 65: 227-48. doi: 10.1353/pew.2015.0004.
- Coseru, Christian (2020) 'Whose Consciousness? Reflexivity and the Problem of Self-Knowledge', in Mark Siderits, Ching Keng & John Spackman, eds., *Buddhist Philosophy of Consciousness Tradition and Dialogue*: 121-53. Leiden.
- Das, Nilanjan, "Śrīharsa", *The Stanford Encyclopedia of Philosophy* (Winter 2021 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2021/entries/sriharsa/>.
- Evans, G. (1982). *The varieties of reference*. Oxford: Oxford University Press.
- Farrell, Jonathan, and Tom McClelland (2017) 'Consciousness and Inner Awareness', *Review of Philosophy and Psychology* 8: 1-22. doi: 10.1007/s13164-017-0331-x.
- Finnigan, Bronwyn (2018) 'Is Consciousness Reflexively Self-Aware? A Buddhist analysis', *Ratio* 31: 389-401. doi: 10.1111/rati.12200.
- Funes Maderey, Ana Laura (2020) 'The Spaciousness of Self-Awareness: A Phenomenological Account of Self-Reflexivity in Patañjali's Yoga philosophy', *Asian Philosophy* 30: 295-306. doi: 10.1080/09552367.2020.1846845.
- Ganeri, Jonardon (1999) 'Self-Intimation, Memory and Personal Identity', *Journal of Indian Philosophy* 27: 469-83. doi: 10.1023/a:1004490802605.
- Ganeri, Jonardon (2012) *The Self: Naturalism, Consciousness, and the First-Person Stance*. Oxford University Press.

- Ganeri, Jonardon (2021). *Inwardness: an outsider's guide*. New York: Columbia University Press.
- Garfield, Jay L (2006) 'The Conventional Status of Reflexive Awareness: What's at Stake in a Tibetan Debate?', *Philosophy East and West*, **56**: 201-28. doi: 10.1353/pew.2006.0020.
- Garfield, Jay L (2015) *Engaging Buddhism: Why it Matters to Philosophy*. Oxford University Press.
- Garfield, Jay L. (2016). Illusionism and Givenness. *Journal of Consciousness Studies*, 23(11–12), 73– 82.
- Garfield, Jay L. (2022). *Losing Ourselves*. Oxford University Press.
- Giustina, Anna (2022) 'A defense of inner awareness: The memory argument revisited', *Review of Philosophy and Psychology*, **13**: 341-63. doi: 10.1007/s13164-021-00602-0.
- Harman, Gilbert (1990) 'The intrinsic quality of experience,' *Philosophical Perspectives* **4**: 31–52. doi: 10.2307/2214186.
- Howell, Robert (2023) *Self-Awareness and the Elusive Subject*. Oxford University Press.
- Jinpa, Thupten (2021) *Illuminating the Intent: An Exposition of Candrakīrti's Entering the Middle Way*. Simon and Schuster.
- Kellner, Birgit (2010) 'Self-awareness (svasaṃvedana) in Dignāga's Pramāṇasamuccaya and-vṛtti: A Close Reading', *Journal of Indian Philosophy* **38**: 203-31. doi: 10.1007/s10781-010-9091-y.
- Kellner, Birgit (2011) 'Self-awareness (Svasaṃvedana) and Infinite Regresses: A Comparison of Arguments by Dignāga and Dharmakīrti', *Journal of Indian Philosophy* **39**: 411-26. doi: 10.1007/s10781-011-9139-7.
- Kriegel, Uriah (2009a) *Subjective Consciousness*. Oxford: Oxford University Press.
- Kriegel Uriah (2009b) 'Self-Representationalism and Phenomenology', *Philosophical Studies* **143**: 357–81. doi: 10.1007/s11098-008-9204-6.
- Kriegel Uriah (2012) 'In Defense of Self-Representationalism: Reply to Critics', *Philosophical Studies* **159**: 475–84. doi: 10.1007/s11098-011-9764-8.
- Lau, Hakwan (2022) *In Consciousness We Trust: The Cognitive Neuroscience of Subjective Experience*. Oxford University Press.
- Mipham, Jamgon (2005) *Introduction to the Middle Way: Candrakīrti's "Madhyamakāvatāra" with Commentary*, translated by the Padmakar Translation Group. Shambala.
- Mohanty, Jitendranath (1971). *The Concept of Intentionality*. St. Louis: Warren H. Green.
- Nida-Rümelin, Martine. 2017. "Self-Awareness." *Review of Philosophy and Psychology* 8 (1): pp. 55-82.
- Peacocke, C. (2008). *Truly understood*. Oxford: Oxford University Press.
- Ram-Prasad, C (2007) *Indian Philosophy and the Consequences of Knowledge. Themes In Ethics, Metaphysics, and Soteriology*. Ashgate.
- Rosenthal, David (2005) *Consciousness and Mind*. Oxford University Press.
- Siderits, Mark (2011). Buddhas as Zombies: A Buddhist Reduction of Subjectivity. In Mark Siderits, Evan Thompson & Dan Zahavi (eds.), *Self, No Self?: Perspectives From Analytical, Phenomenological, and Indian Traditions*. Oxford University Press.
- Silins, Nicholas (forthcoming) 'The Conscious Theory of Higher-Orderness', *Oxford Studies in Philosophy of Mind*.
- Stoljar, Daniel (2018). The Regress Objection to Reflexive Theories of Consciousness. *Analytic Philosophy* 59 (3):293-308.

- Stoljar, Daniel. 2021. "Is there a persuasive argument for an inner awareness theory of consciousness?." *Erkenntnis*: pp. 1-21.
- Stoljar, Daniel (2023) 'Is There a Persuasive Argument for an Inner Awareness Theory of Consciousness?', *Erkenntnis* **88**: 1555-75. doi: 10.1007/s10670-021-00415-8.
- Stoljar, Daniel. forthcoming. "A Euthyphro Dilemma for Higher-order Theories of Consciousness." In *Grounding and Consciousness*, edited by G. Rabin. Oxford: Oxford University Press.
- Thompson, Evan (2011) 'Self-No-Self? Memory and Reflexive Awareness', in Mark Siderits et al, eds., *Self, No Self?: Perspectives from Analytical, Phenomenological, and Indian traditions*: 157-75. Oxford University Press.
- Timalsina, Sthaneshwar (2009). Consciousness in Indian philosophy: the advaita doctrine of 'Awareness-Only' **53**: 1–25. doi: [10.1146/annurev.psych.53.100901.135114](https://doi.org/10.1146/annurev.psych.53.100901.135114).
- Watson, Alex (2014) 'Light as an Analogy for Cognition in Buddhist Idealism (*Vijñānavāda*)', *Journal Of Indian Philosophy* **42**: 401–21. doi: 10.1007/s10781-013-9192-5.
- Williams, Paul (1998) *The Reflexive Nature of Awareness: A Tibetan Madhyamaka Defence*. Curzon Press.
- Yao, Zhihua (2005) *The Buddhist Theory of Self-Cognition*. Routledge.
- Zahavi, Dan (2003). Inner time-consciousness and pre-reflective self-awareness. In Donn Welton (ed.), *The New Husserl: A Critical Reader*. Bloomington: Indiana University Press. pp. 157-180.