

Theoretic Minimalism Among (Advaita Vedānta)
Defenders of the “Self-Illuminating Consciousness” Thesis

The recent literature on the mind/body problem has identified several distinct types of consciousness. Phenomenal consciousness, access-consciousness, monitoring-consciousness, and self-consciousness, for example, have been distinguished. Some argue that some of these are dependent on one or another of these or that one or the other does not in reality exist, but much labor has gone into distinguishing these conceptually. In contrast, the type of consciousness that is the topic of this paper has not, so far as I am aware, had much if any attention drawn to it, that is, not by philosophers belonging to the APA. Now this type of consciousness might be called self-consciousness, as it is indeed in Sanskrit by its champions in classical Indian philosophy—and sometimes in English, too, by those who, like myself, have the Sanskrit tradition in mind. But in the recent professional literature on the mind-body problem, self-consciousness is usually talked about as the possession of a concept of self or as the ability to use such a concept in thinking about oneself, whereas the self-consciousness that is my topic here—Advaita self-consciousness—is considered non-conceptual and independent of all thought even that about it. My point in this paper is to put forth faithfully the notion of this consciousness dear to the classical Indian school known as Advaita Vedānta, to contrast this notion with ideas of consciousness more familiar to Western philosophers, as well as, finally and evaluatively, not so much to wonder whether there is any such thing as to show and review an interesting insularity of the Advaita view to developments in science.

First, some terminological clarification. Were this an audience of India-specialists, “self-consciousness,” as I have said, might work as a label for what the Advaitins are about. Given that many in this group probably already have definite views about self-consciousness, I need another term. “Self-illuminating consciousness” is my choice, a term drawn from the earliest strand of the Vedānta tradition, the *Bṛhadāraṇyaka Upaniṣad* (c. 800 BC). The usage occurs in a passage about transformations of consciousness in dream and mystic trance. A person is said first to dream by his “own light” and then to become “self-illuminated,” *svayaṃ jyotiḥ* (4.3.9). Light is chosen as an analogue apparently because light illumines itself. A lamp illumines objects other

than itself but does not require another lamp to be itself seen. Furthermore, in the *locus classicus* for Advaita views, in the works of Śaṅkara (c. 700 AD), namely, we find the insistence that this consciousness is “non-dual,” *a-dvaita*, that it knows itself by being itself, that is, that it knows itself non-reflectively in a non-intellectual and indeed non-observational manner. Thus “self-illuminating consciousness” seems appropriate, capturing what seems to be kind of a phenomenal self-content.

In other words, whether there is anything like this or not, what the Advaitins appear to have in mind is a phenomenal consciousness whose content is itself. This is nevertheless supposed to be a *state consciousness*, not a *consciousness-of*, not a transitive consciousness but a, so to say, intransitive one. Alternatively, we could say that this is a *consciousness-of* in a sense; it is a consciousness of itself. So here the *consciousness-of* relationship would have to be understood not as the asymmetrical relation it is normally taken to be but rather as something like identity. Note that in the Advaita understanding of this as self-consciousness, “self” is not taken to refer to the body or even the person but rather only the consciousness that is self-aware.

Discussions of types of consciousness typically proceed by presenting examples that are analyzed as exhibiting the one type in contradistinction with others with which it might be confused. But here, by the admission of distinguished members of Śaṅkara’s very own school, the best that can be done is a dubious so-called “indicatory” or ostensive definition. Directions are given where self-illuminating consciousness may be found (e.g., the injunction, “Meditate”), which are said to be like a phenomenal definition of ‘red’ that describes conditions under which one would normally experience the color. (By the way, I say this indicatory definition is dubious because many—especially those in Advaita’s classical adversary school of Nyāya or “Logic”—have not been able to find this kind of consciousness, that is, have been unable to follow the directions. More about this below.) But Advaitins claim that this is available to anyone capable of understanding the directions (“Meditate”) such that consciousness attends to itself. There is no third-person access to this consciousness, so it is said. It is denied that it shows itself in action. The gap between this consciousness and everything else makes discourse that seems to be about it problematic—perhaps the telltale

failure or number-one inadequacy of the Advaita view. But for the moment let us ignore the difficulties about language.

For, Advaitins do contrast self-illuminating consciousness with other types of consciousness or cognition, *jñāna*, presupposing, it seems, an audience of compatriots who know directly, or who could know, what this consciousness is. That there is an epistemic specialness to self-illuminating consciousness is sometimes said to be proved by a sublatability argument. Now this argument is not put forth to show that the ability of other types of consciousness to guide action should be questioned; Advaitins are mislabelled illusionists and skeptics. They do not deny the epistemic value of perception, for example, but use epistemic terms to distinguish self-illuminating consciousness. The contrasting argument, then, goes like this. Everything dualistically experienced is capable of being shown to have been misrepresented through experiential sublation. Thus all cognitions could be sublated and shown to be non-veridical except self-illuminating consciousness. For, since among cognitions only self-illuminating consciousness is not an appearance of one thing in or qualifying another, only in its case is a precondition for sublation not met. That is to say, a perceptual illusion can be sublated because a perception presents an object as qualified by or bearing a property. A piece of silver can be misrepresented as mother-of-pearl because invariably the perceived is taken to be perceived as being some way or other, that is, as possessing a property. And a property presented may not qualify the object perceived in fact, as in the case of silverhood presented as qualifying what is in fact a piece of mother-of-pearl. Normally, cognition is qualificative; something *a* is cognized as F. Sublation is an ensuing experience that shows the *a* not to be F. But self-illuminating consciousness is not in this way “qualificative,” to render the Sanskrit expression, *vaiśiṣṭya*. Its non-dualistic mode of presentation precludes sublation, in sum. It cannot be shown non-veridical—unlike all perception and indeed all thinking (remembering, inferring, understanding what someone has said, et cetera), all normal cognition which is invariably consciousness-of of the transitive type. Self-illuminating consciousness is self-authenticating, at least so Advaitins say to themselves and, it seems, to classical philosophers who were, by the way, pretty much obsessed with an epistemic point of view no matter what the school or banner under which they philosophized. Interestingly, Advaitins also say that there is no real point here to the

question about authentication. To itself, self-illuminating consciousness stands self-revealed.

It is not my intention to pursue today the sublatability argument. I mention it mainly in the contrasting spirit already discussed. But we do need to know that Advaitins also extol the value of self-illuminating consciousness along such lines. All contrasts drawn serve the agenda of somehow encouraging people to attend to this. Indeed, the Advaita school is defined by its commitment to the reality and value of self-illuminating consciousness. What I want to show is how this connects with a theoretical minimalism and Advaita's attitude towards science.

Advaitins—from the eighth century and Śāṅkara to the recent end or tapering off of the Indian classical era and writing in Sanskrit but also right now in our own century—have been able to align themselves with several distinct theories about the operation of the sense-organs and the generation of cognitions through physical processes. I say this is understandable because the Advaita commitment to self-illuminating consciousness appears compatible with all science and externalist theory, except, I suppose one has to add, that which proposes to explain self-illuminating consciousness itself. More about this exception in a moment. First it is important to see that there is no call from the Advaita side to try to explain the world in relation to self-illuminating consciousness. For, self-illuminating consciousness is self-contained and there is nothing to be explained. Furthermore, any such explanation would have to employ terms that are learned through ostensive training involving teachers and pupils and the customs of everyday life, whereas self-illuminating consciousness, we have already learned, is accessible only, so to say, from the inside. This radical internalism is said to have the consequence that speech can only indicate it in the sense explained, not refer to it directly. We'll come back to this claim about language. The main point now is that self-illuminating consciousness is non-relational whereas an explanation would purport to find a tie between explanandum and explanans. Self-illuminating consciousness does not explain anything.

Nor can it be explained. That there is mystery in consciousness revealing diversity, and in the transition from the one to the other, is readily admitted by Advaitins. That is to say, why there is both self-illuminating consciousness and the worldly display is said to be non-explicable, *a-nirvacanīya*. This thesis flows from the non-relationality of

self-illuminating consciousness and the nature of explanation as supposing a relation between explanandum and explanans.

The list of great Advaita philosophers includes several who are expert at dialectics, expert at concocting counterexamples and in general disputing philosophic theories. Often one can find no motive behind the Advaita refutations other than sheer sport. And, to come to the point, Advaitins typically embrace the science of their day (although there are exceptions). Nevertheless, there are, from the Advaita point view, theories—let us call them metaphysical theories—that overstep the bounds of science and purport to explain self-illuminating consciousness itself. Invariably, they try to do so in relational terms, try, that is, to integrate an understanding of “self-consciousness” into a holistic theory. And such theorists, say the Advaitins, necessarily make a mistake, because they treat self-illuminating consciousness as related to other things.

In this context, it is important to evoke the epistemic perspective, because epistemology provides the canons of winning and losing a debate or dispute and we are now engaged in a dispute, so say champions of the Advaita cause. Self-illuminating consciousness is self-authenticating and, unlike other conscious states and material phenomena, has an exclusive access to itself. Thus only it has the right to pronounce on itself, so to say. Those who would explain self-illuminating consciousness are trespassers who invariably get their putative explanandum wrong.

Personally, however, I find the non-relational thesis that underlies this line of argument unsatisfactory. Just how is it that Advaitins know that self-illuminating consciousness cannot be related to other things? The idea seems to be that this consciousness’s being absorbed with itself exclusively translates into its being explanatorily unavailable. An explanation would be like an unwanted disturbance violating self-illuminating consciousness’s self-absorbed trance. But how is it that Advaitins know so much about this inexplicable as to be able to explain why it is not explicable? This brings us back to the problem of the language Advaitins use.

The insularity from other areas of theory is purchased by Advaita too cheaply. Advaitins draw a distinction between descriptive use of words and the indicatory, as I mentioned earlier. There is supposed to be a difference between (1) describing Devadatta’s house, explaining what a house is—say, to a child—and (2) indicating Devadatta’s house

conversationally by saying that it is the one where some crows pointed to are currently hovering. The *viśeṣaṇa/upalakṣaṇa* distinction—a distinction between true and pseudo qualifiers in connection with use of words to pick out something (*a*) by way of a property it really has (*Fa*) as opposed to words used to direct a hearer's attention to something (*a*) by means of a thoroughly contingent and accidental relation it has to something else (*b*) which is what the words literally refer to—is, for Advaita talk of self-illuminating consciousness, unhelpful. For, in their arguments Advaitins consider themselves to know some true properties of self-illuminating consciousness. Why else would they say that it is self-illuminating, or non-relational, and then reason according to these attributions?

The distinction between indication and description that we find in classical Indian arguments and philosophy of language is too big to dig into much further today. I'll end with a mere footnote about the distinction in another context. Advaita's superrealist adversaries in Nyāya also cite the distinction when confronted with their own peculiar difficulties. To take up again the example of Devadatta's house, some of the house's properties, according to Nyāya, are causally relevant to any perception of the thing as a house. But indicators such as the hovering crows have the faintest relevance to our house perception and, indeed, to what the thing is. Furthermore, indicators and mention of indicators seem not to be explained by the thing but by our purposes. The Nyāya program is to account for what a person says out of perceptual evidence by a realism of objects possessing real properties, those mentioned in the verbalizations of perceptions. But how then to account for the hovering crows mentioned on the basis of a perception of Devadatta's house? What are the criteria whereby true properties are to be distinguished from mere indicators? Here the controversy centers on the realist assumption that a perception's content is a guide to real features of real things in a world external to and causally determinant of the content. Some content turns out to be determined solely by extraneous cognitive and communicational circumstances, content, namely, that indicates something that it does not qualify. The realist should be able to provide, therefore, a clear and easy way to distinguish indicators from true qualifiers, proper properties. But the history of the classical debate shows that this is something that he may not be able to do. Personally, I find it indicative of Advaita obscurantism

about the language they use concerning self-illuminating consciousness that they resort to the very distinction employed by the realists to try to escape from their own difficulties. In the one case, the indicatory pointing can only work if, despite the Advaitin contention, there is a relation between that which is indicated—self-illuminating consciousness, namely—and the meanings of the terms used, and in the other case, in the case of Nyāya, indication (*upalakṣaṇa*) seems to show that perception is not a sure-fire guide to something's properties.

Stephen Phillips
University of Texas at Austin