

## How a materialist can deny that the United States is probably conscious – response to Schwitzgebel

### Abstract:

In a recent paper, Eric Schwitzgebel argues that if materialism about consciousness is true, then the United States is likely to have its own stream of phenomenal consciousness, distinct from the streams of conscious experience of the people who compose it. Indeed, most plausible forms of materialism have to grant that a certain degree of functional and behavioral complexity constitutes a sufficient condition for the ascription of phenomenal consciousness – and Schwitzgebel makes a case to show that the United States as a whole fulfills this condition.

One way to avoid this counter-intuitive consequence of materialism about consciousness is to adopt what Schwitzgebel calls an “anti-nesting principle”: a principle that states that there can be no nested forms of phenomenal consciousness and that therefore a conscious whole cannot have parts that are themselves conscious. However, Schwitzgebel then proceeds in his paper to draw up various objections, notably based on thought experiments, in order to dismiss these kinds of “anti-nesting” principles.

My aim in this paper is to present a version of a sophisticated anti-nesting principle that avoids Schwitzgebel’s objections. This principle is reasonable, intuitive, and as non-arbitrary as possible. Moreover, it can resist the objections mounted by Schwitzgebel against simple anti-nesting principles. This principle helps materialists avoid the implication that the United States has its own stream of consciousness, while granting consciousness to some entities which, in many cases, are intuitive instantiators of phenomenal consciousness (among which are cases of authentic group consciousness). This principle therefore constitutes a way out for a materialist who wants to deny that the United States is conscious.

**Keywords:** consciousness, experience, materialism, group consciousness

### 1. Introduction:

In a recent paper (Schwitzgebel, forthcoming), Eric Schwitzgebel argues that “if Materialism is true, the United States is probably conscious”; that is to say: if materialism is true concerning phenomenal consciousness then the US, given its functional and behavioral complexity, is likely to have its own independent stream of phenomenal consciousness. As Schwitzgebel himself puts it, this is a counter-intuitive consequence of materialism, and most contemporary materialists would be happy to avoid it. In his paper, Schwitzgebel discusses and rejects what he calls “anti-nesting principles” (Schwitzgebel, forthcoming, p. 6-7): principles that entail that no conscious whole can have conscious subparts. If we assume that individual humans are conscious, such a principle excludes the possibility of larger conscious wholes composed of human individuals, such as the problematic conscious whole of the United States. Schwitzgebel rejects anti-nesting principles on the grounds that currently available articulations of such principles seem arbitrary and are vulnerable to a variety of objections based on thought experiments.

My goal here is to present and defend a particular version of an anti-nesting principle that I find reasonable, intuitive, and non-arbitrary, the acceptance of which would allow materialists to deny that the USA is probably conscious. Accepting my anti-nesting principle allows both the denial of the claim that entities like the United States are singly conscious, and the ascription of consciousness in a number of real and imaginary cases to entities that intuitively instantiate it – among which are cases of authentic group consciousness.

In order to present and defend my view, I will proceed as follows: In section 2 I will briefly present Schwitzgebel's argument, as well as his objections to anti-nesting principles. In section 3, I will present the more sophisticated anti-nesting principle I intend to defend, and in sections 4 and 5 I will explain how this principle can resist Schwitzgebel's objections.

## **2. Schwitzgebel's argument and the problem with anti-nesting principles**

Schwitzgebel's argument rests on the untendentious claim that if we accept materialism we are likely to admit that the phenomenally conscious states of a given entity supervene on its functional organization.<sup>1</sup> If we do so, Schwitzgebel argues, then we have to accept that the US (for example) is probably conscious, given that the kind of functional organization instantiated by the US is not fundamentally different when it comes to organizational and behavioral complexity from the kind of functional organization instantiated by entities that we tend to think of as phenomenally conscious. The US, understood as the vague-boundaried group of American citizens who sometimes act in a coordinated manner, does seem to be able to massively gather and process information. It can represent and self-represent, and act in a coherent, semi-intelligent fashion attuned to environmental inputs. In his paper, Schwitzgebel argues that it is difficult to find a capacity (relevant for consciousness) the USA could lack that would be possessed, for example, by a simple creature to which we consensually ascribe phenomenal consciousness (take, for example, animals such as rabbits).

Schwitzgebel addresses many objections to his provocative thesis. He notably mentions one possible way for the materialist (who accepts the supervenience of phenomenally conscious states on functional organization, or something approaching) to deny that the US is phenomenally conscious, on which I want to focus now. This way endorses what he calls an "anti-nesting principle" (Schwitzgebel, forthcoming, p. 6). Anti-nesting principles are principles which assert that it is impossible for a conscious entity to be constituted by other conscious entities. One typical example of such a principle is Putnam's principle (Putnam, 1967), according to which no organism capable of feeling pain possesses a decomposition into parts separately capable of feeling pain. A generalization of this principle could be expressed as follows: it is impossible for a phenomenally conscious whole to have subparts that are themselves phenomenally conscious.<sup>2</sup>

If materialists accepted a principle of that kind, they would be able to deny that the US is phenomenally conscious even given its functional organizational properties, since the US has human beings as subparts and nobody in this debate wants to deny that human beings are phenomenally conscious. However, according to Schwitzgebel, there are two reasons which preclude us from accepting anti-nesting principles.

---

<sup>1</sup> This is not exactly true, given that Schwitzgebel's characterization of materialism is compatible with some forms of type-materialism (Schwitzgebel, forthcoming, p. 8) which identify human consciousness with specific biological processes (McLaughlin, 2007; Polger, 2004). However, even if such theories are not committed to the supervenience thesis just stated, their most plausible versions have to accept that certain levels of complexity regarding the functional organization of a system constitute a condition for the presence of a stream of consciousness attached to this system. For reasons of simplicity however, I will focus on a functionalist understanding of materialism which implies that the phenomenally conscious states of a system supervene on its functional organization.

<sup>2</sup> Another anti-nesting principle discussed by Schwitzgebel is Tononi's (Oizumi, Albantakis, & Tononi, 2014; Tononi, 2012).

The first is that, according to Schwitzgebel, anti-nesting principles are arbitrary. That is to say, they seem to be merely designed to avoid counterintuitive implications of whatever materialist theory of consciousness one is willing to adopt, and have no independent support in their favor. The second reason (which I believe plays the more important dialectical role) is that such principles seem to be at odds with the intuitive conclusions of various thought experiments. Here are two examples of such thought experiments.

The first example is a thought experiment put forth by Schwitzgebel himself, which features “Antarean Antheads”. These creatures “look like woolly mammoths but [...] act much like human beings”. However, “they are quirky in a few ways [...] For example, their cognitive activity takes them on average ten times longer to execute”; moreover, “their heads and humps contain not neurons but rather ten million squirming insects, each a fraction of a millimeter across. Each insect has a complete set of minute sensory organs and a nervous system of its own” (Schwitzgebel, forthcoming, p. 5-6). And it is the interactions between these ants, which we can suppose are (at least possibly) phenomenally conscious, that instantiate the functional organization on the basis of which Antheads act like humans.

Are Antheads phenomenally conscious? Schwitzgebel is inclined to say yes. But if they are conscious and if we grant that the insects inhabiting their brains – which are responsible for the complex behavior of Antheads – are conscious too, then we face an intuitive counter-example to the kind of anti-nesting principle suggested by Putnam.

The second example is a thought experiment presented by Schwitzgebel in his paper (Schwitzgebel, forthcoming, p. 7) and first suggested by Ned Block (Block, 1978): let’s imagine that tiny conscious organisms are incorporated into our brains, and that these organisms decide to play the role of some of our neurons. If Putnam’s anti-nesting principle is true, it means that the incorporation of *even* one of these tiny conscious organisms into our brains would render us unconscious despite the fact that the functional organization of our brains and minds – and our behavior – would remain the same. This seems to be an extremely counter-intuitive consequence, which makes such a strong anti-nesting principle unlikely to be true.

For these two reasons, Schwitzgebel seems to think that we should refuse to endorse any version of an anti-nesting principle. However, I think that such a conclusion is misguided. I think that there is a correct intuition behind the idea of anti-nesting principles, even if Putnam’s version of such a principle is subject to counterexample. My aim is to put forth a more sophisticated version of an anti-nesting principle, which is able to resist the objections raised by Schwitzgebel. I will presuppose the truth of a thesis implied by most versions of materialism<sup>3</sup>, but which doesn’t itself imply materialism, given that it is compatible not only with materialism about consciousness but also with the kind of naturalistic dualism defended by David Chalmers (Chalmers, 1996), according to which the phenomenally conscious states of a given entity supervene on its functional organization. I will therefore presuppose that a given entity is phenomenally conscious and has phenomenally conscious states (phenomenally conscious perceptions, emotions, beliefs, etc.) if it has the right kind of functional organization. I will not start out by coming down on the side of any particular functional materialist theory<sup>4</sup>, but my anti-nesting principle may in the end imply that certain kinds of functional organization *cannot* be the basis of a genuine stream of consciousness.

### 3. The Sophisticated Anti-Nesting Principle

I now want to present an anti-nesting principle which seems to me to be able to fulfill the conditions just mentioned. I call it the “Sophisticated Anti-Nesting Principle” (SAP), to distinguish it from other anti-nesting principles already discussed in the literature, and it is formulated such that it can be added to any theory of consciousness (e.g., the theories of consciousness considered by

---

<sup>3</sup> But not all of them, strictly speaking: see the first footnote of this paper.

<sup>4</sup> For the main theories discussed by Schwitzgebel, see (Chalmers, 1996; Dennett, 1991; Dretske, 1988, 1995; Humphrey, 1992, 2011; Tononi, 2004).

Schwitzgebel). Here is how things should work: any theory of consciousness which asserts that the conscious states of an entity supervene on its functional states (whether this is a materialist theory or not) will have to determine, given a conscious mental state *S*, a functionally individuated property *P* such that an entity instantiating *P* is a sufficient condition for the entity instantiating *S*.<sup>5 6</sup> Given any such theory, one can add to this theory an anti-nesting principle, which supplements the theory by determining a set of cases that constitute exceptions to the theory such that *P* is instantiated but *S* is not. Once supplemented by this principle, each theory of consciousness would contain statements such as: if an entity has *P*, then it has *S*, *except if it falls under the cases described by the sophisticated anti-nesting principle*.

Let us now expose the precise formulation of the SAP:

**Sophisticated Anti-Nesting Principle (SAP):** Given a whole *W* that instantiates the functional property *P*, such that *W*'s instantiation of *P* is normally sufficient for *W* to instantiate the conscious mental state *S*, ***W* does not instantiate *S*** if *W* has at least one subpart that plays a role in its functional organization which fulfills at the same time the two following conditions:

- (A) The performing of this role by the subpart *requires* (given the nature of this functional role and our theory of consciousness) that this subpart has conscious mental states (beliefs, emotions, hopes, experiences, desires, etc.) that represent *W* (what it is, what it does, what it should do). That is to say, this subpart has a functional property *Q*, *Q* being a sufficient condition for the subpart having the conscious mental state *R* (where *R* is a mental state representing *W*).
- (B) If such a functional role (i.e., a functional role of such a kind that it requires that the subpart performing it has conscious mental states representing *W*) was not performed by at least one of the subparts of *W*, *W* would no longer have the property *P* (or *any other* functional property sufficient for the having of *S*<sup>7</sup>). In other words: if no subpart of *W* had *R*, then *W* would no longer have *S*.

The SAP therefore asserts that a whole *W* does not have *S* if the having of *P* by *W* (having of *P* which, in “normal” cases, is sufficient for the having of *S*) features the having, by a subpart of *W*, of a functional property *Q* which is a sufficient condition for this subpart having a conscious mental state *R* (where *R* is a mental state representing *W* itself), in such a way that, had the subpart not had *Q* (or any other functional property sufficient for the having of *R*), *W* would not have had *P* (or any other functional property sufficient for the having of *S*). In other words: according to the SAP, if the functional complexity of a whole, which would in “normal” cases (that is to say, cases without nested consciousness) be a sufficient condition for the whole having a given conscious state, only exists in virtue of one of the subparts consciously representing the whole itself, then we face a case of genuine nested consciousness, such that the whole does not have this given conscious state.

#### 4. Defending the SAP against Schwitzgebel's thought experiments

<sup>5</sup> Of course, it may be that a given theory will have to determine many properties of that sort. Ideally, it will have to give an exhaustive list of those, which will allow to refer to the property *Z*, described by the disjunctive statement of *all* the properties of the kind *P*. *Z* will be such that an entity having *Z* will constitute a *necessary and sufficient condition* for an entity having *S*.

<sup>6</sup> I speak of *P* as a sufficient condition for *S*, but not as *constituting* *S*, because amongst the theories I am considering here are some anti-materialist theories of consciousness, such as Chalmers' naturalistic dualism. And according to this kind of theory, the having of a functional property *P* by an entity can be considered as a sufficient condition for the having of a phenomenal property *S* by the same entity (the laws of nature being fixed) without the having of *P* *constituting* the having of *S*.

<sup>7</sup> That is to say, it would no longer have the property *Z*, where *Z* is the (possibly disjunctive) property, the having of which by an entity is a *necessary and sufficient* condition for this entity having *S*.

My aim is now to defend the SAP, by showing that it can resist Schwitzgebel's objections against anti-nesting principles.

First, it should be noted that if we adopt the SAP we can grant that Antheds are phenomenally conscious. Indeed, the insects located in their heads have no conscious states representing the whole creature in itself (so condition A is not fulfilled), and in any case their functional role could be played by "simple" and obviously non-conscious subparts such as neurons (so condition B is not fulfilled). The SAP also allows that the replacement of my neurons by tiny conscious organisms willing to play the neurons' role, as described in Block's thought experiment, wouldn't preclude me from being conscious, given that this role can obviously be played by a simple non-conscious subpart such as a neuron. So in this case, condition B is not fulfilled in even if condition A is fulfilled.

But the SAP nevertheless allows us to deny that the US is phenomenally conscious. Indeed, some subparts of the US (American citizens, for example) have conscious mental states representing the US: what the US is, what the US does and what the US should do. Moreover, if American citizens were replaced by entities without any conscious mental states representing the US, the US would no longer display the kind of complex functional organization which is necessary for a system to have conscious mental states. A "country" where neurons, or even simple computers, replaced the inhabitants, or a country in which the inhabitants had no idea of their existence as a group, wouldn't be able to display the kind of behavioral and cognitive complexity displayed by the US, which formed the original basis of Schwitzgebel's challenge. The case of a potentially conscious US fulfills conditions A and B of the SAP; therefore, the SAP correctly predicts that the US cannot instantiate conscious mental states

The SAP also has another advantage: it allows, in principle, for some kinds of group consciousness. Conscious beings could actually be among the subparts of a conscious whole entity, as long as the functional organization of the whole does not depend on the fact that the members of the group have conscious attitudes representing the whole.

For example, it *could* be the case that we humans, without knowing it, are currently engaged in a kind of collective functional organization which is responsible for a form of group consciousness, of which we are unaware. It could also be the case that some members of humankind could come to *discover* the existence of the collective organization or of the group consciousness without making it disappear. But if this discovery was a necessary condition of a new behavior of these humans, new behavior which in turn would be a necessary condition of new aspects of the functional organization of the whole, these new aspects could not be the basis of *genuine new conscious states of the whole* (even if it doesn't mean that the whole wouldn't preserve some of its "old" phenomenal states, linked to a functional organization which doesn't require the discovery of the whole as a whole by the members of the group).

This allows us to point out another advantage of the SAP: given that it doesn't bear on the general ascription of "phenomenal consciousness" to an entity, but rather on the ascription of *phenomenal mental states* to such an entity, it allows us to make *subtle discriminations* among the conscious states that we can ascribe to a whole constituted by other conscious beings. For example, consider Block's example: let's imagine that some of my neurons are replaced by very clever tiny conscious organisms. If these organisms decide to play the role that used to be played by my neurons, I would, according to the SAP, remain conscious all the same. However, if these tiny organisms have the conscious intention, for example, of changing my linguistic behavior in order to make me talk about an extremely complex mathematical theorem M (a theorem that no normal human would be able to understand, for example), then according to the SAP I would not have conscious thoughts about M (though one could be tempted to ascribe me those thoughts on the sole basis of my external behavior), even if I still retained my good old "normal" conscious states<sup>8</sup>, and

---

<sup>8</sup> An anonymous reviewer suggested a variation on that case which could constitute an objection to the SAP: let us suppose that the tiny clever organisms decide to make me talk about M by changing my brain *directly* in

even if I would probably gain some new sensory phenomenology (for example, auditory phenomenology of myself talking about M). I think that these consequences of the SAP fit nicely with our intuitions regarding those cases.

The SAP shares the spirit of Chalmers' objection to the existence of a stream of consciousness attached to the United States, which Schwitzgebel discusses in his paper (Schwitzgebel, forthcoming, p. 17-18) - even though Chalmers' objection doesn't rely on the explicit formulation of an anti-nesting principle. Chalmers' objection goes like this: perhaps the United States lacks consciousness because its complex cognitive capacities arise largely in virtue of the capacities of the individual people composing it, and only to a very small extent in virtue of the functional relationships between those people. Schwitzgebel's answer to this objection relies on the fact that such a description could very well apply to the human mind itself. For instance, theories based on the "global workspace" model of the mind (Baars, 1988; Dehaene & Naccache, 2001) have it that most of our cognitive capacities arise in virtue of the cognitive work done within specialized subsystems, which then feed information into a global workspace, which is where consciousness is actually realized; however, nobody seem to think that this would imply that human beings are not themselves conscious.

The SAP has the advantage of retaining the intuition behind Chalmers' objection while escaping Schwitzgebel's reply. Indeed, according to the SAP, the US cannot be conscious because its complex cognitive capacities arise largely in virtue of the capacities of the people composing it, *and because those capacities are such that those people have conscious mental states representing the US itself without which the complex cognitive capacities of the US could not arise*. This is clearly not unlike the case of the hypothetical modules proposed by the global workspace model; no theorist

---

order to "insert" a conscious thought about M in me. It may seem that the SAP would imply that such thoughts cannot be conscious. This would be an utterly counter-intuitive consequence of the SAP : indeed, in that case, it is stipulated that the actions of the clever organisms only "cause" my thoughts about M, without "constituting" them, and it is hard to see why the fact that my thoughts have such causes rather than others, would make them less conscious – even though it could make them be out of my control, in such a way that I would not be their author (Shepherd, forthcoming; Vosgerau & Voss, 2014). However, I don't think this is a problem for the SAP in the end. Indeed, when we detail this case, it can be described in two versions. In the first, we can imagine that the tiny clever organisms change my brain in a temporally localized manner, so that after the change I remain able to talk about the complex theorem by myself and without their intervention. In that version, I grant that I would have conscious thoughts about M, and denying that I am conscious of M would be very counter-intuitive. But this is perfectly compatible with the SAP, because the conscious intervention of the clever organisms changed (once and for all) my functional organization, but cannot be considered as a part of my own functional organization (while the SAP only applies when a whole has a conscious part playing a certain role in one's functional organization). Indeed, for X to be a part of the functional organization of Y requires a certain *lasting* (even if merely dispositional) contribution of X to the causal properties of Y, and this is not the case when we only consider a temporally localized modification (in the same way as an act of brain surgery can change my functional organization, without this very act being in itself a part of this organization). In the second version of the case, we can suppose that a permanent intervention of the tiny organisms upon my brain is necessary to maintain the appropriate behavior and organization. In that case, the SAP indeed states that my "thoughts" about M won't be conscious thoughts, as me having them requires that the conscious organisms play a certain role in my functional organization, a role which itself requires that these organisms have thoughts about me as a whole, and which could not be played by non-conscious, non-intelligent entities. In that case, the functional complexity that makes my "thoughts" be indeed about M really comes from the clever tiny organisms and requires that these organisms have mental states representing me; in a way, the actions of the organisms don't only "cause" my thoughts, but also "constitute" them, as only they have the functional features which are required to make my thoughts be really about M. In this last version of the case, I would simply bite the bullet and claim that I don't have conscious thoughts about M, but I don't think that, so described, this would be such a counter-intuitive claim after all. Of course, the SAP does not lead to denying that, in this last version of the case, the intervention of the clever organisms can cause me to have a slightly different stream of consciousness compared to what I would have had without this intervention – however it would not give me conscious thoughts about M.

would say that those modules have conscious mental states regarding the human being as a whole, or the human mind as a whole. In fact, they are not supposed to have any conscious mental states *at all*.

In conclusion: the SAP allows us to deny that the USA is phenomenally conscious; it also grants that in the Anthheads' case and in Block's neuron replacement case the whole is conscious. It allows for some kinds of group consciousness, under certain conditions, and I think that these conditions precisely match the conditions under which we would intuitively think that group consciousness can appear. This principle escapes the counterexamples mounted by Schwitzgebel against other anti-nesting principles, as well as against Chalmers' objection. Finally, the SAP is consistent with, although it does not itself entail, the possibility of spatially discontinuous conscious entities such as Schwitzgebel's "Sirian Supersquids"<sup>9</sup>, as much as the possibility of animal consciousness and alien consciousness. To use Schwitzgebel's terms, this principle doesn't entail any form of "contiguism" or "neurochauvinism"<sup>10</sup>.

## 5. Defending the SAP: this principle is reasonable, intuitive, and not arbitrary

Finally, even if this point seems more difficult to make, I will do my best to show that the SAP is intuitive, reasonable and as minimally arbitrary as possible.

First, I think that anti-nesting principles *are* intuitive: there is an intuitive pull behind Putnam's idea according to which it is impossible for a phenomenally conscious whole to have subparts which are themselves phenomenally conscious. Perhaps Putnam's mistake was to give an interpretation of the "subpart" aspect which seemed to refer to *spatial* or *material* subparts, whereas a better interpretation would understand it as referring to *functional subparts*. This is precisely what the SAP does. While Putnam's style of principle states that no phenomenally conscious whole could have subparts that are themselves conscious, the SAP predicates that the problem only arises when the phenomenal consciousness of the whole exists *in virtue* of certain kinds of phenomenal states instantiated by parts of the whole, and of some particular functional organization of the whole that these phenomenal states make possible in virtue of their phenomenality. This, together with the other specifications given by the SAP, allows us both to retain the intuitive pull behind anti-nesting principles *and* to give more plausible accounts of a wide range of imaginary cases. Therefore, this principle seems to me to be both intuitive and reasonable.

Is the SAP arbitrary? I think that it is formulated in sufficiently *general* terms so that it will not be considered *ad hoc*. Moreover, I think that this principle is not arbitrary, as it can be justified by appealing to a broader understanding of what consciousness is supposed to be. This can be shown by contrasting the SAP with other anti-nesting principles, notably Putnam's. Indeed, one can say that the problem with Putnam's principle was that the principle stated that nested consciousness is impossible, without the statement being clearly supported by its relations to other beliefs concerning consciousness. On the other hand, the SAP states that it is not impossible for a conscious whole to have conscious subparts, but it is impossible for a conscious whole to be conscious in virtue of the

---

<sup>9</sup> "Sirian Supersquids" (Schwitzgebel, forthcoming, p. 3-4) are imaginary intelligent creatures, whose intellectual achievements are comparable to those of humans, but whose nervous system is distributed through their head and a thousand tentacles. They can detach those tentacles, but the nervous signals (based on light signals rather than chemical signals) can still be reliably transmitted from a distance. For this reason, their spatially discontinuous nervous system can remain fully and coherently integrated from a functional and cognitive point of view. Schwitzgebel takes it that denying a genuine stream of consciousness to Supersquids is arbitrary, and requires that we appeal to an unjustifiable form of prejudice against spatially discontinuous cognitive systems ("contiguism") according to which only spatially continuous entities can have their own streams of consciousness.

<sup>10</sup> The meaning of "contiguism" has been explained in the previous note. "Neurochauvinism" refers to the kind of prejudice (unjustified according to Schwitzgebel) that would lead us to deny consciousness to any creature whose cognitive functioning and information-processing is not neuron-based.

fact that its subparts are themselves conscious of the whole (when certain conditions are fulfilled). This thesis can be seen as a consequence of a broader principle that seems to be an important *prima facie* constraint on our theories of consciousness: one should not needlessly multiply ascriptions of consciousness, which is itself a specialized version of Occam's Razor. In particular, one should only ascribe consciousness to an entity when one cannot explain the behavior and the organization that seems to justify this ascription as the consequence of mental states of other, distinct subjects – notably mental states of other subjects which bear on the very behavior and organization of the entity. This general principle, I think, is the reason why we are reluctant to ascribe genuine intentional states to books and computers, or genuine conscious states to a puppet which is acted on by a human being. It is also the reason why we are reluctant to ascribe a genuine stream of consciousness to the USA – the complex behaviors exhibited by each of these entities can be fully explained in terms of the conscious mental states of other, distinct entities. Such a principle can itself be justified (even though a satisfying justification would require further reasoning) by appealing to the idea that the ascription of consciousness is, amongst other things, supposed to play a role in the explanation and justification of the behavior of the entities to which it is ascribed. But, as for all explanations and justifications, one should always choose the simpler explanation/justification when faced with many explanations/justifications. For example, in the case of explanations, the simpler explanation will typically be the one whose *explanans* posits as few tokens of explanatory entities as possible (*ceteris paribus*).

The SAP can be seen as an attempt to give a precise and detailed application of this principle in the case of consciousness, in the particular case of a part/whole situation. For this reason, I think that the SAP does not encounter the same difficulty as Putnam's principle, as it can be supported by appealing to some independent and broader principle stating what role ascriptions of consciousness should play.

Of course, in spite of these considerations, the SAP can still appear to be arbitrary to a certain extent. Indeed, one can always ask meaningfully "Why is it that a whole is conscious in this case, but not in that case?" But we have to keep in mind that this kind of meaningful open question can seemingly be asked about *any* theory of consciousness. This is what supports Chalmers' thesis according to which it seems that there is a "hard problem" of phenomenal consciousness, or Levine's idea (Levine, 1983) that there is an "explanatory gap" between consciousness and the physical and the functional. Given any physical or functional state, it seems that we can intelligibly ask the questions: "Why is this physical state correlated with a conscious state, and precisely this conscious state?"<sup>11</sup> or "Why does this physical state constitute a phenomenal state, and precisely a phenomenal state of this kind, while this other physical state does not constitute a phenomenal state (or a phenomenal state of a different kind)?" For this reason, the fact that the SAP seems residually arbitrary cannot constitute an objection which applies specifically to this principle or even to anti-nesting principles in general, given that this residual arbitrariness concerns, to a certain extent, any theory of consciousness. Indeed, the question of knowing *why* consciousness is correlated with (or constituted by) certain physical processes, rather than with others, seems to stay always at least partly unanswered. This explains why we should not consider this residual "arbitrary" aspect of the SAP as a fatal flaw.

## 6. Conclusion

The Sophisticated Anti-Nesting Principle I just presented seems able to fulfill the two requirements of a successful anti-nesting principle. First, this principle, coupled with many different functionalist materialist theories of consciousness, entails intuitive answers to many thought experiments that posed a challenge to other anti-nesting principles. Second, this principle is intuitive,

---

<sup>11</sup> This arbitrary aspect concerns materialist theories of consciousness, as well as some dualist theories of consciousness (like Chalmers' (Chalmers, 1996), which have to posit *basic laws* linking phenomenal properties to physical properties).



reasonable and as little arbitrary as possible. For these reasons, I think that this principle should be accepted by materialists (and possibly by dualists who endorse the supervenience of conscious states on functional organization). The SAP gives the materialist a natural way to deny, *pace* Schwitzgebel, that the USA is phenomenally conscious, and more generally, a way to understand and assess various possible cases of group consciousness in an intuitive and reasonable.

#### Reference List:

- Baars, B. (1988). *A Cognitive Theory of Consciousness*. Cambridge: Cambridge University Press.
- Block, N. (1978). Troubles with Functionalism. *Minnesota Studies in the Philosophy of Science*, 9(261-235).
- Chalmers, D. (1996). *The Conscious Mind: In Search of a Fundamental Theory*. Oxford University Press.
- Dehaene, S., & Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework. *Cognition*, 79, 1-37.
- Dennett, D. (1991). *Consciousness Explained*. Penguin.
- Dretske, F. (1988). *Explaining Behavior : Reasons in a World of Causes*. MIT Press.
- Dretske, F. (1995). *Naturalizing the Mind*. MIT Press.
- Humphrey, N. (1992). *A History of the Mind: Evolution and the Birth of Consciousness*. Chatto&Windus.
- Humphrey, N. (2011). *Soul Dust: The Magic of Consciousness*. Princeton: Princeton University Press.
- Levine, J. (1983). Materialism and qualia: the explanatory gap. *Pacific Philosophical Quarterly*, 64(October), 354-61.
- McLaughlin, B. (2007). Type Materialism for Phenomenal Consciousness. In M. Velmans & S. Schneider (Ed.), *The Blackwell Companion to Consciousness*. Malden, MA: Blackwell.
- Oizumi, M., Albantakis, L., & Tononi, G. (2014). From the Phenomenology to the Mechanisms of Consciousness: Integrated Information Theory 3.0. *PLOS Computational Biology*, 10(5), e1003588. doi:10.1371/journal.pcbi.1003588.
- Polger, T. (2004). *Natural Minds*. Cambridge, MA: MIT Press.
- Putnam, H. (1967). Psychological Predicates. In W. Capitan & D. Merrill (Ed.), *Art, Mind, and Religion*. Pittsburgh: University of Pittsburgh Press.
- Schwitzgebel, E. (forthcoming). If Materialism is True, the United States Is Probably Conscious. *Philosophical Studies*. DOI: 10.1007/s11098-014-0387-8.
- Shepherd, J. (forthcoming). Conscious Control over Action. *Mind and Language*.
- Tononi, G. (2004). An Information Integration Theory of Consciousness. *BMC Neuroscience*, 5(42).
- Tononi, G. (2012). Integrated Information Theory of Consciousness: an Updated Account. *Archives Italiennes de Biologie*, 150, 290-326.
- Vosgerau, G., & Voss, M. (2014). Authorship and Control over Thoughts. *Mind and Language*, 29(5), 534-565.