

---

# Representationalisms, Subjective Character, and Self-Acquaintance

Kenneth Williford

---

In this study I argue for the following claims: First, it's best to think of subjective character as the self-acquaintance of each instance of consciousness—its acquaintance with itself. Second, this entails that all instances of consciousness have some intrinsic property in virtue of which they, and not other things, bear this acquaintance relation to themselves. And, third, this is still compatible with physicalism as long as we accept something like *in re* structural universals; consciousness is a real, multiply instantiable, natural universal or form, but it likely has a highly complex, articulated structure, and “lives” only in its instances. In order to make these cases, I give a characterization of subjective character that accounts for the intuition that phenomenal consciousness is relational in some sense (or involves a subject-object polarity), as well as the competing and Humean intuition that one of the supposed relata, the subject-relatum, is not phenomenologically accessible. By identifying the subject with the episode or stream of consciousness itself and maintaining that consciousness is immediately self-aware (“reflexively” aware), these competing intuitions can be reconciled. I also argue that it is a serious confusion to identify subjective character with one's individuality or particularity.

I argue that deeper reflection on the fact that consciousness has only incomplete self-knowledge will allow us to see that certain problems afflicting acquaintance theories, like the one I defend, are not the threats to certain forms of physicalism that they might seem to be. In particular, I briefly consider the Grain Problem and the apparent primitive simplicity of the acquaintance relation itself in this light.

## Keywords

Acquaintance | Consciousness | Direct realism | First-order representationalism | For-me-ness | Harder problem | Heidelberg school | Higher-order representationalism | Individuality | Individuation | Intrinsic property | Mineness | Naturalize | Particularity | Phenomenal consciousness | Phenomenal intentionality | Physicalism | Qualitative character | Reflexive awareness | Reflexivity | Relational property | Representation | Representationalism | Same-order representationalism | Self-acquaintance | Self-knowledge | Self-representation | Sense of self | Sense-datum theory | Stream of consciousness | Structural universals | Subject | Subjective-character | The grain problem | Transparency intuition

## 1 Introduction

In this study, I argue for the following claims: First, it's best to think of subjective character as the self-acquaintance of each instance of consciousness—its acquaintance with itself.<sup>1</sup>

<sup>1</sup> As will become clear shortly, contrary to ordinary ways of speaking, I do not hold that persons must be the “subject relata” of acquaintance relations. Rather, I hold that episodes of consciousness are, fundamentally, the subject relata.

## Author

Kenneth Williford  
williford@uta.edu

The University of Texas  
Arlington, TX, U.S.A.

## Commentator

Tobias Schlicht  
tobias.schlicht@rub.de

Ruhr-Universität  
Bochum, Germany

## Editors

Thomas Metzinger  
metzinger@uni-mainz.de

Johannes Gutenberg-Universität  
Mainz, Germany

Jennifer M. Windt  
jennifer.windt@monash.edu

Monash University  
Melbourne, Australia

Second, this does indeed entail that all instances of consciousness have some internal relational property (or intrinsic property) in virtue of which they, and not other things, bear this acquaintance relation to themselves. And, third, this is still compatible with physicalism as long as we accept something like *in re* structural universals. There is always a price, but in this case

it's arguably no more than the price we pay to be scientific realists.<sup>2</sup>

To make these cases, I must consider some important preliminaries. I give a characterization of subjective character that accounts for the intuition that phenomenal consciousness is relational in some sense (or involves a subject-object polarity), as well as the competing Humean intuition that one of the supposed relata, the subject-relatum, is not phenomenologically accessible. If the latter is true, it is hard to explain how we could have immediate evidence (as opposed to some sort of inferential knowledge) of the existence of this relational structure—evidence we do seem to have. If we identify the subject with the episode or stream of consciousness itself (however we individuate or ontologize these)<sup>3</sup> and maintain that consciousness is immediately self-aware (“reflexively” aware<sup>4</sup>), then the intuition of relationality and the Humean intuition of the missing subject can be reconciled.

I also argue that it is a serious confusion to identify subjective character with one's individuality or particularity. This will be considered first from a phenomenological point of view, in relation to our tendency to describe subjective character in terms of ownership or “mineness”, and then from an ontological point of view, in relation to the metaphysical individuation conditions of distinct streams of consciousness.

Further, I argue that deeper reflection on the fact that consciousness has only incomplete self-knowledge will allow us to see that certain problems afflicting acquaintance theories, like the one I defend, are not the threats to certain forms of physicalism that they might seem to

be. In particular, I briefly consider the Grain Problem<sup>5</sup> and the apparent primitive simplicity of the acquaintance relation itself in this light.

Preliminary to all this, we must first briefly consider the inadequacies of representationalism, and at least adumbrate some of the motivations for the recently renewed interest in the idea of acquaintance (see e.g., Chalmers 2003; Tye 2011, pp. 96–102; Gertler 2011, pp. 87–128, 2012; Balog 2012; Howell 2013, chs. 3 & 4; Goff forthcoming). I argue that, indeed, we need to lose our fear of moving beyond reductive naturalistic representationalisms, especially in regard to subjective character. My conclusions, and in many cases arguments, are not entirely new, but I attempt to cast the material in a new light, in a spirit of synthesis.

The dialectical structure of this study is somewhat circuitous. In section 2, I argue that the most plausible representationalist theory of consciousness is a self-representationalist one (or “Same-Order” representationalism) because it captures subjective character, which I view as essential to consciousness, with the smallest theoretical cost. However, I argue, all forms of representationalism about consciousness are ultimately implausible. This leads to a focused discussion of the notion of subjective character in section 3, the notion that motivates higher-order and same-order representationalisms. In that section, I argue that subjective character should be identified with the self-manifestation or self-appearance of consciousness. Consciousness, the claim goes, appears to *itself* no matter what else appears to it. This in turn allows us to make sense of the competing relationality and Humean “no-self” intuitions mentioned above. Combining these elements from sections 2 and 3, I argue in section 4 that we should understand self-manifestation in terms of self-acquaintance rather than self-representation. In section 5, I clear up what I regard to be the not uncommon confusion of subjective character with individuation. And in section 6, I argue

5 The Grain Problem, customarily attributed to Wilfred Sellars, is a problem for any identity theory according to which sensory qualities are really brain properties of some sort. Roughly put, the problem is that brain properties are complex and structured while sensory qualities seem, on the face of it, ultimately simple and unstructured. For good discussions with references to Sellars see Clark (1989) and Lockwood (1993).

2 This is not to imply that scientific realism entails physicalism, of course.

3 This is a difficult issue I will not enter into. See e.g., Dainton (2000, 2008); Strawson (2009).

4 I will occasionally use the terms “reflexivity” and “reflexive awareness” to denote just this characteristic of consciousness (i.e., that of its always being aware of itself). It is not to be confused with “reflection” in the sense of introspection. It is more like the logical usage of “reflexive” (as in “reflexive relation”). The acquaintance relation is reflexive on the domain of conscious states, according to the view accepted here (as well as being anti-symmetric). But not everything that stands in this relation is self-acquainted—episodes of consciousness are, but they are also acquainted with sensory qualities, and these latter are not acquainted with anything.

that though the view espoused here implies that being conscious is a matter of having certain intrinsic properties, this is compatible with a certain type of physicalistically acceptable hylo-morphism—the view that complex kinds of physical objects, properties, or processes involve the concrete instantiation of real structures and cannot be properly understood in abstraction from such a “marriage” of form and matter.

## 2 Representationalisms: From first-order to same-order

In the theory of consciousness, the term “representationalism” has, aptly but somewhat confusingly from a historical point of view, come to designate any view according to which being phenomenally conscious is equivalent to representing the right sort of things in the right sort of way. There is, of course, much internecine disagreement over these things and ways, but the main idea is simple and attractive enough. If we could understand consciousness in terms of representation and representation in terms of some naturalistically acceptable relations, then we could “naturalize” consciousness. I’ll call representationalisms that are coupled with naturalistic theories of content *reductive representationalisms*.

Representationalisms are typically divided up into various “orders.” These orders have, in a way, to do with the kind of content (or object) a conscious representational state supposedly must have. For First-Order (F) representationalisms the relevant states are, fundamentally, just directed at worldly objects and properties (typically the sensible properties of tables, chairs, etc., see, e.g., Tye 1995, 2000; Dretske 1995). For Higher-Order (H) representationalisms, the states must be directed at mental states of “lower-order”—possibly but not necessarily first-order (see e.g., Rosenthal 2005; Lycan 1996). For Same-Order (S) representationalisms, the representational state must be directed at itself (or, perhaps, some part of itself, or a whole of which it is a part, or another part of the whole of which it is a part).<sup>6</sup> I also add Priv-

ileged-Object (P) representationalisms as a distinct category. For these, the state must be directed at some special type of entity—a model of the organism as a representational or embodied homeostatic system, a “proto-self” or, less naturalistically, perhaps an enduring substantial ego entity.<sup>7</sup>

There is, however, no obvious reason why there could not be unconscious representations with any of these contents. And, generally, it seems implausible that something could be conscious in virtue of representing a certain type of object—this is Alvin Goldman’s so-called “Problem of the Rock” (thinking about or seeing rocks does not make them conscious, so why should it make anything else conscious?), which seems to apply to H, S, and P theories—but see below (see e.g., Goldman 1993; Gennaro 2005; Lycan ms).

For F theories, since it is admitted there can be conscious and unconscious states with the same sort of content, another distinguisher between conscious and unconscious mental states will have to be found. For F theorists, this has typically been a functional constraint placed on the representations (e.g., poise, feeding into the mind-reading system, becoming available to the global workspace, see, e.g., Tye 2000 and relatedly Baars 1997; Dehaene & Naccache 2001), sometimes coupled with the necessary condition that the properties represented must be represented in a “non-conceptual” way (whatever that is taken to amount to).<sup>8</sup> For the H theorists, it has been a somewhat different story.

H theorists are generally motivated by a phenomenological inadequacy they see in F the-

theorist. S theory is also often called self-representationalism.

<sup>7</sup> For naturalistic versions, see e.g., Damasio (1999 and 2010), Metzinger (2004), and Sebastian (forthcoming). I am sure that Damasio, Metzinger, and Sebastian would reject this label, but the point of it is that all these theories identify subjective consciousness, in one way or another, with the representation of a “self,” understood in a naturalistically acceptable sense. See e.g., Metzinger (2004), p. 302: “In short, a self-model is a model of the very representational system that is currently activating it within itself” (emphasis original); and Damasio (2010), p. 180: “[T]he brain constructs consciousness by generating a self process within an awake mind. The essence of the self is a focusing of the mind on the material organism that it inhabits.” It should be noted that Metzinger allows that there could be conscious experience that does not involve subjective character (see Metzinger 2004, pp. 559-560). Thus my categorization here applies at most only to his theory of subjective consciousness. Since, for me (as for Damasio), all consciousness necessarily has subjective character, this difference in detail will not loom large in what follows.

<sup>8</sup> See the excellent discussion of the “non-conceptual content” literature in Hopp (2011).

<sup>6</sup> See e.g., Gennaro (2012); Kriegel (2006, 2009); Weisberg (2008, 2014). Williford (2006) can be taken to express a pure S view—the conscious mental state has itself for its own object, not some portion of itself. We can also classify Carruthers as an S theorist; see Carruthers (2000, 2005). Gennaro would not describe himself as an S

ory. F theorists generally stress the so-called “Transparency Intuition”—the idea, roughly put, that first-order consciousness reveals only properties and objects in the world and nothing directly about consciousness itself, the perceiving mind, the subject, or the vehicles of representation (see e.g., Harman 1990; Tye 2000; Byrne 2001). H theorists, on the other hand, are with varying degrees of explicitness motivated by the equally powerful intuition that consciousness involves some sort of “for-me-ness” or “to-me-ness,” often termed “subjective character” (see e.g., Rosenthal 1986, p. 345 and Gennaro 2006. See also Levine 2001, pp. 104–111). This gets encoded in the H mantra that the conscious states are just those that one is “Aware of Being In”, those that one is aware that one is oneself in (see e.g., the “Introduction” to Kriegel & Williford 2006). The thought is that F theory simply does not capture that intuition. F theorists and their fellow travelers would consider such “essentially indexical” contents or the “sense of self” to be more advanced cognitive products or artifacts of social cognition, certainly not in the very ground floor of consciousness (see e.g., Edelman & Tononi 2000, pp. 103–104 and Macphail 1998, pp. 2–5).

There is here an important bifurcation in intuitions about consciousness. Some significant percentage of us thinks that subjective character (however we ultimately understand it) is essential to consciousness, is in the ground floor. And some significant percentage of us thinks that it is not; that somehow qualitative character (perhaps understood as having the right sort of representational content) is essential but that subjective character is derived, secondary, or tertiary. This bifurcation shows up in neuroscientific and psychological thinking on consciousness as well.<sup>9</sup> We will briefly return to the significance of this bifurcation point in the next section.

The H theorist has a few options about the exact content of the H representation, the higher-order thought (or perception [or global state]).<sup>10</sup> There are serious and well-known

problems here. If the represented lower-order state (L state) of, say, visual perceptual awareness were different from the representing H state in terms of relevant content (e.g., if the one represents a phenomenally green ball and the other a phenomenally red one), what would we consciously see? “Red. No, green. No, red...” This is the Problem of the Division of Phenomenal Labor, or mirepresentation problem, as I will sometimes call it, and is related to deep and probably insoluble problems about the epistemology of introspection that are pertinent to such models (of both H and S varieties).<sup>11</sup> If the L state simply did not exist, would your conscious experience in that case be a sort of Meinongian hallucination? This is the Problem of Targetless H States.<sup>12</sup>

To take up the latter problem just a bit, if one takes literally much of the talk one finds in the literature on H theory, the H thought is supposed to *make* the L state conscious. Being conscious is a kind of extrinsic (external relational) property of the L state, a property it has in virtue of its being represented by the H state. Thus, if there is such an H state, it does confer at least a relational property (the property of “being made conscious by the H state”) on the L state. In the cases in which the L state does not exist but the H state directed at it does, some non-existent object, the L state, is made conscious by an H state. Thus the L state would literally have a relational property; it would stand in a relation, even though it does not exist. This literal interpretation of the view entails some form of Meinongianism (at least about non-existent L states) and that you can seem to yourself to be conscious when you are not. Thus, presumably, it should not be taken so literally.<sup>13</sup>

<sup>11</sup> See e.g., Neander (1998); Horgan & Kriegel (2007); Weisberg (2008); Tye (2011, pp. 4–8). See Kidd (ms) for an excellent discussion of these epistemological issues in the (not interestingly different) case of S theory.

<sup>12</sup> See Mandik (2009 and forthcoming) on the “Unicorn problem” and Block (2011). See Rosenthal (2011, 2012); Weisberg (2011a, 2011b); Kiefer (2012); Wilberg (2010), and Berger (2013) for discussions of various strategies for dealing with Higher-Order-Thoughts (HOTs) without Lower-Order-Thoughts (LOTs).

<sup>13</sup> What I am calling the “non-literal” interpretation is, in effect, the position in Berger (2013). And in Rosenthal (2011, p. 436) he in effect claims that the non-literal position (as I am calling it) has always been his view. See Mandik (forthcoming) on this.

<sup>9</sup> For example, Tononi & Koch (2008, pp. 240–241) do not seem to think that the “sense of self” is essential (though Tononi (2014) may have recently changed his view); Damasio (1999, 2010) is in the opposing camp; see also Northoff (2013).

<sup>10</sup> I’ll not go into the Higher-Order Thought vs. Higher-Order Perception debate. See e.g., Gennaro (2012).

The non-literal interpretation, however, is inimical to one of the reductive pretensions of the H strategy. It's not inimical to reductive representationalism as such. But it does draw in to question the idea that a reductive theory must construe the property of being conscious as an external-relational property of otherwise unconscious mental states (see [Rosenthal 1997](#)). Thus, it could only be in virtue of the specific content or structure of the H state itself that there is consciousness. One would then be putting forth the presumably phenomenologically motivated *a posteriori* identity hypothesis that the conscious representational states are just the ones with that content. There may be differences over the specific content (e.g., Is it about some of my other mental states, or is it just about the non-mental objects and properties of the world?) and differences over other criteria (e.g., poise); but otherwise, on the non-literal interpretation, H theory is structurally just like F theory. We can of course wed either of these to a reductive theory of representation, but this will only make "being conscious" into an external-relational property to the extent that the theory of representation adopted makes all representation an external-relational matter.

If one is still conscious when the L state does not exist, then the H state would seem to be doing all the work. And that's what we should focus our explanatory efforts on. What could be special about it? Again, putting aside other types of external relations (e.g., being available to the global workspace), it must have a special sort of content. But it is not in virtue of *being represented* that a state could be conscious. Rather, on this non-literal interpretation, it is in virtue of *being a representation of X* (where *X* is a special object of some sort, e.g., oneself being in a state) or *that p* (where *p* is a proposition with a special content) that the state is conscious; and we can, as with any other sort of contentful state, try to figure out how different naturalistic theories of representation would construe states with that content.

Whatever theory of content we adopt, we'll want to know what salient or interesting properties, from an explanatory point of view, such representations have. What is it about you

that you can represent yourself as being in a state or that a conscious state of yours is occurring now? Find that out, the promise goes, and we will understand consciousness. But, I would argue, none of the theories of representation we have to go on tell us anything very significant about such states. The beaver's tail splash, says Millikan, to take one sort of example, can represent the very time at which it occurs (among other things; [Millikan 1995](#), p. 98). This does not make it conscious. This particular example applies directly to Same-Order theories, but surely the beaver's tail splash *could have* represented a previous tail splash and its content or its simultaneous front paw splash, etc., but that would not in itself make anything conscious either, right?

Naturalistic theories of representation will not themselves tell us anything that interestingly distinguishes H states (or S states) from F states (or P states for that matter). In every case (F, H, S, P), it is just a matter of some physical representational vehicles standing in some set of external (or externally mediated) relations to other physical objects (and sometimes to themselves). From this point of view, we see nothing that interestingly distinguishes the theories.

Moved by these problems, H theorists might try to go the "essential indexical" route (cf. [Weisberg 2012](#)). After all, on Rosenthal's original formulation, the conscious states are those one is aware of *oneself* as being in. But here they are faced with a difficult choice. If they presuppose a teleosemantic theory, then they have to face the fact that on this theory there are no literally *essential* indexicals (see e.g., [Millikan 1990](#)). Change the relevant history and other external relations and you change the content—now an indexical, now a proper name, now a substance term, etc. If they abandon teleosemantics, they could go back down some Fregean rabbit hole.<sup>14</sup> That way lies murk or perhaps triviality (see [Cappelen & Dever 2013](#)). But it seems inadequate just to postulate that

<sup>14</sup> I assume here but will not argue that teleosemantics is the most plausible naturalistic theory of content. There may be other naturalistic options that allow one to make good sense of the notion of essential indexicality in a way that could help H theory here, but I doubt it.

the H state contains a definite description that happens to pick oneself out. Thus the H theorist might be led to consider what is in effect a P theory. One then tries to find a suitable entity to play the role of the privileged object (a privileged signified, if you will): the proto-self, the self-model, or what have you.<sup>15</sup>

It is hard to see how any of these possible objects would somehow help us to make sense of subjective character. And it is hard to see how representing some special object could be that in virtue of which something is conscious. If “essentially indexical” content is either explicable in terms of something more basic (as seems to be the case to me), or impossible (as on teleosemantics), or metaphysically fraught in an ultimately un-illuminating way, then it seems like the best bet is to adopt a version of S theory.

For one thing, we can reduce the metaphysical load that threatens to plague the notion of essential indexicality and solve the non-existence problem at once.<sup>16</sup> All we need are token mental states representing themselves. As a corollary, we can give a deflationary account of “essentially indexical” content in token-reflexive terms<sup>17</sup> that is potentially compatible with teleosemantics (or whatever non-Fregean account one prefers) and find some other way to capture the grain of truth reflected in the opacity arguments presented by Castañeda, Lewis, and Perry.<sup>18</sup> In my view, anyone committed to the intuition motivating H theory should become an S theorist, if for no other reason than because of the non-existent L state problem. The other possible solutions (e.g., Gennaro’s “WIV”) introduce a kind of theoretical inelegance that renders them less plausible.

H theories are better than F theories, given my intuitions anyway, because they en-

code the essentiality of subjective character to consciousness. If that intuition is good then, of the two classes, H theories are the better ones. But H theories face the non-existent L state problem. To solve it, they must either embrace murk or metaphysical baggage (if they go in the direction of some P theories), or embrace the postulation of certain epicycles, or go some order. S, in my view, is evidently the best option for the representationalist.

S theory avoids *ad hoc* moves, better reflects the clarified phenomenological intuitions that are the real motivation, can ground a theory of indexicals, and does not commit one to an enduring self-entity of any sort; nor does it seem to attempt to get subjective character out of something’s representation of something else that is structurally similar to itself, as this last move runs afoul of the Fichte-Shoemaker Regress.<sup>19</sup> S theory evidently does not fall prey to the non-existent L state problem, even if it does not avoid the misrepresentation problem. In the end, however, it is itself nothing more than a type of P theory. The Privileged Object is just the token mental state (or episode) itself. Clearly, there is no self-evident reason why something’s representing itself should make it conscious, even if it is in fact true that all conscious episodes do represent themselves.

We surely cannot seriously imagine that consciousness emanates from a special object it needs to look at, even if that object is just the current experiential time-slice itself. Further, something’s representation of itself, naturalistically understood, is no more theoretically inter-

<sup>19</sup> See Henrich (1982); Frank (2002, 2007); Shoemaker (1968). The issue, which is part of the “essential indexical” problematic, is, when put into a “self-model theory” context (which is not to be identified with Metzinger’s views), just that modeling something structurally isomorphic to oneself is not sufficient for knowledge that one is modeling oneself, as opposed to having behavioral control through such an interface (I could be controlling my doppelgänger unwittingly and just as effectively). One would need to know that the thing modeled is oneself (and not something else that happens to be isomorphic to it, like one’s counterpart in a close possible world). One cannot, on pain of regress, derive such knowledge from a set of descriptions of oneself without already knowing that at least one of the descriptions does indeed apply to oneself. So one must have some direct self-knowledge, such as knowledge by acquaintance that one is the relevant so-and-so. An S theory wedded to a teleosemantic theory of representation and externalist theory of justification has the advantage of being able to accommodate direct reference and non-inferential knowledge of oneself, though one will regard this as a mere simulacrum of the phenomenology.

<sup>15</sup> See Sebastian (forthcoming) for a Damasio-inspired turn toward a P theory (at least, that was my interpretation of it).

<sup>16</sup> We can’t eliminate the misrepresentation problem, however. But we bracket that for now. See Kidd (ms) and Weisberg (2008).

<sup>17</sup> A la Higginbotham (2003 and 2010) and before that (implicitly) Smullyan (1984); see Cappelen & Dever (2013, pp. 160-161). The hyperset model in Williford (2006) is the skeleton of such a theory. See also Kapitan (2006).

<sup>18</sup> See Cappelen & Dever (2013, ch. 10). They attempt to capture this grain by appealing to relatively un-puzzling epistemic limitations. I believe they are on the right track, even if I would characterize the specific limitations in question a bit differently (see the discussion below on our ignorance of what fundamentally individuates us).

esting (or even surprising) than its representation of the world or of one's other thoughts and perceptions. Thus, that does not, *a priori*, appear to be the sort of thing that would be more likely to be equivalent to consciousness than something's representation of something else. Perhaps adding functional constraints would help here but no more than it might help H or F theory.

Even if it is true that all conscious states are self-representational, it is, of course, far from clear how that fact should help us *explain* consciousness. The same can be said for H theory and other P theories. Rather, in all these accounts, we are merely trying to isolate what we take the unique content of consciousness to be and then to apply our theory of representation to states with such content. Absent some strong phenomenological intuitions to the contrary, the conscious mental states, it seems, might well have been all and only those states in which dogs are represented. In the end, though, all "normal" physicalists (i.e., those who reject Russellian Monism, Panpsychism, and Pan-proto-psychism) are reduced to *some* such strategy. All "normal" physicalists, representationalist or not, will identify consciousness with something that is not *a priori* known to be equivalent to it. We return briefly to this familiar problematic at the end.

But, perhaps most alarmingly, reductive S theories (and H theories, and everything in between) are either subject to a version of the old Swampman objection or otherwise untenable.<sup>20</sup> Since the conscious states are, on the theory, just special representational states, they are subject to the constraints of the underlying theory of representation (in this case, teleosemantics). If they don't have the right history, then they don't have the right content. And if they don't have the right content, they are not conscious. Surely there is something simply absurd about the idea that one might or might not be conscious depending on how one's atoms happened to get into the current arrangement.

<sup>20</sup> See e.g., Tye (2000, ch. 6). I will not be able to go into the back and forth over Swampman. Suffice it to say that despite hearing many attempted rejoinders over the years, I still find the objection to be a *reductio* of representationalist theories of consciousness wedded to historico-externalist theories of content.

It is not that one cannot concoct a response to the objection; it is, rather, just the very fact that the view invites such objections in the first place. It demands a rather serious and ugly epicycle; and that counts strongly against it. But if we reject teleosemantics and adopt an internalist theory so as to escape from Swampman, we face equally difficult problems that we cannot, unfortunately, go into here.<sup>21</sup>

The view then is that H theories are better than F theories on phenomenological grounds and that S theories are better than H theories on dialectical *and* phenomenological grounds. But all versions wedded to naturalistic historico-externalist theories of representation are shipwrecked on the Swampman problem, and internalist versions face other equally difficult problems. What then shall we do?

We might consider trying out a non-reductive representationalist version of S theory. This is a possibility we will return to in section 4. But first we need to reflect a bit on what H, S, and P theories are trying to capture in the first place. What is the phenomenological datum designated by this phrase "subjective character," and why is it that F (and related) theorists don't see it as essential to consciousness, while H, S, and some other theorists do?

### 3 Subjective character

Subjective character is often described as a certain "for-me-ness," "mineness," or even "me-ishness" that is phenomenologically manifest and, presumably, always accompanying, even if in a muted or background form, any consciousness whatsoever (see e.g., Zahavi 2005; Levine 2001; Kriegel 2009 and Block 1995). F and related theorists point out that it also seems that one can become so absorbed in one's actions, at one extreme, and perhaps so dulled at the other that one loses all sense of oneself (see e.g., Tononi & Koch 2008, pp. 240–241). Moreover, they might argue that it does not seem reasonable to suppose that worms and bees have a

<sup>21</sup> See e.g., Carruthers (2000, 2005) and Gennaro (2012, pp. 45–49). Briefly, the sort of functional role semantics Carruthers embraces derives actual, occurrent content from dispositions, and it is actually subject to variations on the Swampman theme.

sense of self at all, and yet they may be conscious. A common reply from the defenders of subjective character to the first claim is that we are not talking about focusing on oneself or one's current mental state as an object of attention or concern, and that, if they tried harder, F theorists would realize that even in the most dulled or, at the other pole, absorbed state, they are still aware at some level of themselves (or the very experiential state they are in). To the second objection, the typical reply is that the sort of subjective character we are envisaging does not require the sort of conceptual sophistication or reflective capabilities that would make it impossible for dogs (or even bees and worms) to count as conscious beings (see e.g., [Gennaro 2012](#), chs. 7 & 8). Of course, the replies can be replied to, and so on. And we won't enter into these debates here. Suffice it to say that, unsurprisingly, those who think that subjective character is essential to consciousness have ways of answering objections, just as do those who deny its essentiality. As commonly happens, the answers drive us back to questions that are themselves at least as hard to settle as the ones we began with. Moreover, appeals to the neuroscientific and psychological literature in the attempt to decide these issues sometimes get what plausibility they have from interpretations of the experiments and results that are as questionable as the claims they are supposed to support.

My view here is that one should follow the modeling path inspired by one's "phenomenological muse" and give up fighting phenomenological intuition wars. If you find subjective character to be essential, develop models of consciousness that encode that, and see where they lead. If you don't find it essential but find other things to be more important (multimodal information integration or availability in the global workspace or whatever), model those. And let's not forget that we might all be working on different parts of the same elephant, so perhaps we will be able to combine models fruitfully one day. Eventually we may have ways of more or less decisively testing the different models.<sup>22</sup>

<sup>22</sup> See [Kriegel \(2007\)](#) for an excellent discussion of phenomenological impasses. Thanks to Jennifer Windt for reminding me of this lucid article.

Different intuitions about what is essential to a phenomenon drive different models of the phenomenon. As long as enough people (and don't ask for a number) share one's phenomenological intuitions, one's project won't be, we hope, insane or unmotivated. In regard to the present bifurcation point, many otherwise sane, rigorous, and careful thinkers in many widely distributed traditions and disciplines have had some version of the intuition that consciousness, somehow, involves a sense of self or sense of it-self.<sup>23</sup>

Now, how should we characterize subjective character at the phenomenological level? It does not add much to say that it is a "sense of self." What sort of a sense of self are we talking about? To say that it is "mineness" or "for-meness" makes it seem as though we are talking about the *ownership* of experiences. But this is probably just a certain analogy based on the ownership of property. Yes, for all that matters here, it may well be the case that, always, if I am in a position to know, without having to observe any behavior, that *there is* a pain in the room, then I am in a position to know that it is my *own* pain in the room. But it does not do much good to say that "me-ishness" or "mine-ness" adheres to my experiences like a property or haecceity. It is not as if I just see that my experiences have Willifordhood instead of Zahavihood or Gallagherhood, and thereby know whose are whose—like distinguishing two otherwise qualitatively identical coats by different name tags on the inner pockets.

Note that looking for a special property of the experience is not that different from seeking out its relation to a special object (its owner or The Self) that one may be directly acquainted with. In both cases we are looking for a special something that individuates the experience. There is no interesting difference here between a special unique property that only my experiences have and a special unique self-object to which they all relate.

<sup>23</sup> For just a few examples of the historical pedigree here, see [Caston \(2002\)](#) on Aristotle, [Williams \(2000\)](#) and [Coseru \(2012, ch. 8\)](#) on the Indian and Buddhist debate, [Thiel \(2011\)](#) on the early modern problematic, [Frank \(2004\)](#) on the German Idealist and Romantic discussion, and [Zahavi \(1999 and 2006\)](#) on the Phenomenological movement.



Subjective character should probably not be thought of as a matter of a constant relation to a self-object or as a special property of mine-ness or me-ishness that all experiences come with, all the more is this so if it is possible to misattribute ownership to certain sensations.<sup>24</sup> The first-personal dimension (Zahavi), the sense of self in the act of knowing (Damasio), for-meness, me-ishness (Block), ipseity, *être-pour-soi* (Sartre), *Selbstvertrautheit*, and so on—these are all suggestive names for the phenomenon in question. But we'd like to know if there is not an at least somewhat less ambiguous way of characterizing it.

One name for it that I do rather like depends on a grammatical analogy that can be fleshed out a bit more. Every experience, we may say, involves the appearance *of* something *to* something (or someone). The former can be called the *genitive of manifestation* (appearance-of), the latter the *dative of manifestation* (appearance-to).<sup>25</sup> The genitive of manifestation corresponds to the intentionality of consciousness—its directedness at objects; the dative of manifestation corresponds to subjective character. The identification of subjective character and the dative of manifestation may not at first be so obvious.

The primary intuition here is that there is no such thing as the mere non-relational phenomenal appearance of an object or quality. Objects and qualities don't just phenomenally manifest—full stop. Rather, anything that phenomenally appears, appears to someone or something (cf. Strawson 2011, pp. 41–46). If this were false, phenomenal consciousness would be more like a monadic property of its objects than like a relation between a subject and an object of some sort (see Butchvarov 1979, p. 250). The idea that consciousness could be phenomenally manifest but manifest *to* no one is

either incoherent or, at best, strains credulity. Yet this seems to be exactly what F and related theorists are committed to—aches and pains that can appear (be phenomenally conscious) but appear to no one.

If we accept that there is a dative of manifestation, that objects and qualities appear *to* someone or something, we are closer to but not quite up to subjective character just yet. Subjective character, recall, is supposed to be something phenomenologically detectable. And one might raise the following sort of worry. Suppose phenomenally manifest objects and properties are manifest *to* something or someone. It does not follow from this alone that that *to which* they are manifest is itself manifest or even manifestable. Nor does it follow that the *fact that* they are manifest to something is manifest or even manifestable. In other words, there could indeed be a dative of manifestation and yet no direct phenomenological evidence of this at all. In fact, Hume's famous failure to find his own self and Moore's similar but more tentative musings on this issue can be taken as expressions of the intuition that we do not find a distinct subject relatum in experience.<sup>26</sup> And surely it is true that we do not find a little ubiquitous homunculus—the constant and ever-present thing Hume might have been seeking, like the little face at the bottom of old first-person video games like Quake—to which all our experiences relate—nor do we find a self-haecceity forever re-instantiated by our conscious episodes.

There is, however, this strong intuition that phenomenal consciousness is relational, that it involves a subject-object polarity. And the strong intuition that we do not find any entity or special criterial property that could be a self-entity, me-haecceity, me-ish quale, or subject-relatum is in some apparent tension with this intuition of relationality. Moreover, a *hidden* subject-relatum would not account for the phenomenology of subjective character, evidently. There is a real question here. How is it that consciousness seems to have a subject-object relational structure, and yet we do not

<sup>24</sup> See e.g., Lane & Liang (2011). (Thanks to an anonymous reviewer for pointing this nice article out to me.) If, as I shall argue, subjective character is not fundamentally a representational matter at all, the issue of *representational immunity* to error through misidentification is orthogonal. To the extent that the attribution of ownership is a representational matter, it may or may not be possible to misattribute ownership, as far as the view defended here is concerned.

<sup>25</sup> The terminology apparently derives from Prufer (1975) and is very common in phenomenological quarters. See e.g., Zahavi (1999); Crowell (2011, p. 16).

<sup>26</sup> See Moore (1910); Butchvarov (1979, p. 250, 1998, p. 55), and Williford (2004). On Hume in this regard, see Strawson (2011).

seem to be able to find the subject-relatum, one of the relata of the relation? Isn't it the case that if something non-inferentially seems relational, then we are non-inferentially aware of its (at least) apparent relata? Speaking naïvely and barring certain irrelevant counterexamples, if I see that the cup is on the table, don't I see the cup and see the table too? In the case of the subject-object polarity, do we imagine or project this relation? Is it a product of reflection and memory?

It seems to me that the F theorist should say that it is somehow a product of higher cognition that is projected onto normal adult human conscious experience. But if one is really committed to the intuition that subjective character is an essential and hence ubiquitous feature of conscious experience, then one will simply have to abandon self-relatum and self-haecceity accounts as characterizations of the phenomenology (and as explanatory models, for that matter). What we need is an account of how it is that consciousness manifestly and non-inferentially appears to have a relational structure even though one of the relata is, in a certain sense, invisible.

Here the view that consciousness is self-manifesting can save the day. An episode or perhaps stream of consciousness, on this view, appears to itself at every moment while other things appear to it as well. This will require more unpacking, but at present we just want to clarify the putative phenomenological content of the claim as best we can. We leave the notion of *appearing* or of *phenomenally manifesting* undefined. Or, if you prefer, we define it ostensibly by inner ostension and hope that our interlocutors know what we are talking about and have similar conscious minds (cf. Fales 1996, pp. 147–148).

Let's say that phenomenal manifestation is just the appearance to/in consciousness of something. Let's leave it open what that something is (qualities, facts, objects). We all can know what phenomenal manifestation is, in this purely phenomenological sense, if we are conscious and capable of normal reflection, attention, memory, and conceptual cognition. If we have tasted coffee, then the taste of coffee has

been phenomenally manifest to us. If we haven't, then it has not. And think of this generically—it's what experiencing the taste of coffee has in common with seeing the blue sky and with feeling one's own existence.<sup>27</sup> Now, the claim is that an episode of consciousness is phenomenally manifest to itself whenever anything else is phenomenally manifest to/in that episode. Whenever anything else appears to consciousness, that act or episode or stream of consciousness appears to itself as well. And it is important to remember that this does not mean that one is reflecting on one's experience or that one has any propositional attitude towards that experience or that one is paying any attention to that experience as such.

Now, let us suppose that this is the case. Can we recover a notion of subjective character from this in a way that accounts for both the Humean intuition that the subject-relatum is, in some sense, invisible and that, nevertheless, consciousness has a subject-object relational structure that is phenomenally manifest and non-inferentially knowable? Yes, we can, and at a relatively low price.

The subject-relatum, on the current proposal, is just the episode of consciousness itself. The episode appears to/in the episode. Other things (qualities, objects, etc.) appear in/to the episode as well. The episode is a unified whole, the differentiated qualities and objects appearing in/to it are like its parts (stressing “like”—it's an analogy).<sup>28</sup> We do not find episodes that do not have parts (except perhaps in some very special circumstances), but it is foolhardy to look for some special entity or haecceity that is separable from all the other parts or like a part among the parts. There is no such thing. And that, arguably, is the sort of thing Hume was failing to find. No such subject is given, hence we don't find it. Nonetheless, the true subject-relatum, the episode of consciousness itself, is not invisible. It is manifest.

<sup>27</sup> Cf. Moore (1910, p. 57). (This paper of Moore's is not as well known as his “Refutation of Idealism,” but it deserves to be.)

<sup>28</sup> I will not attempt to offer an account of the (synchronic or diachronic) unity of consciousness in this paper (again, see e.g., Dainton 2000) or of mereological principles governing “parts” of episodes of consciousness and episodes as “wholes.” It is enough for my purposes that one recognize that conscious episodes are internally variegated unities of some sort.

The main price to pay here is that we must try to wrap our heads around the idea that an episode of consciousness could be the phenomenological *subject* of consciousness. I say, and say truly, that such and such appears to *me* or that *I* see, feel, hear, or am conscious of such and such. If I am a subject of consciousness and all subjects of consciousness are just so many episodes, then am I just an episode of consciousness?! I've seen the incredulous stares with my own eyes and have been told that the sentence expressing the view that the subject of consciousness is the episode of consciousness has the same status as sentences like, "Pink dreams sleep furiously."

Indeed, this claim seems wildly counterintuitive at first. But once we realize that there is a certain temporal element connoted in our usage of "I," then this can be ameliorated. "I" normally refers not just to the present experience but to a whole history of connected experiences and much else besides. So it would be a mistake to infer from "I've seen the incredulous stares" the claim that "Incredulous stares were seen in/by this current episode of consciousness." Instead, in the spirit of Four Dimensionalism, one should translate thus: There was a past series of conscious episodes suitably connected to each other and to the present one; incredulous stares were seen by/in them for some time; and the episodes are being recalled in/by the present conscious episode, which bears the same relation (transitively conceived), or some suitable analogue thereof, in the case of broken streams, to that sequence of earlier episodes.

Note, however, that fundamentally the use of "I" is anchored in moment-by-moment, self-manifesting conscious experience. Imagine a person with severe anterograde amnesia and retrograde amnesia as well. Such a person might think, from moment to moment, "I am seeing this," "I am feeling that," but beyond a certain perhaps necessary amount of working memory, they may not carry any of that information into their future. We can imagine truly minimal subjects that have only the minimal amount of working memory required for consciousness, supposing that some amount is required. On the view proposed here, such a conscious being's

consciousness would still have subjective character. It would simply fail to be more or less automatically enriched by memory, projection, familiarity with one's body and dispositions, autobiographical idealizations and distortions, etc., that is, by the autobiographical representational grid through which our experience is normally spontaneously filtered. Perhaps such a person could not think "I" in the sense in which we normally think it. They may lack an "autobiographical self" and even "extended consciousness", as Damasio would put it (see [Damasio 1999](#) and [2010](#)). But their experience would be self-manifest and other things ("parts") would be manifest in/to that experience as well.

Still, isn't it a bit too odd to hold that the whole episode is conscious of its "parts"—however we end up construing these? Or that the "parts" are phenomenally manifest to/in the whole they belong to? Doesn't this still seem like a totally bizarre thing to say? We have to remind ourselves that there is no thing in consciousness, no ego entity, no homunculus that these qualities could be manifest *to*. We don't find any such thing; and no *hidden* thing could allow us to account for the phenomenology. However, we agreed (I hope) that consciousness has a relational, subject-object structure and that this structure is itself phenomenally manifest and not inferred.

Another way to put it is to say that there is a kind of contrast present in our experience all the time. Something is before me, and it is not me. Something is present to consciousness, but it is not that consciousness. Given our mereological analogy, this contrast is a bit *like* that between a whole and its proper parts. The whole is not a proper part. Yet, at a suitably generic level, it bears the same relation to itself that it bears to its constituents (everything is a part of itself too, though an improper part).

Assuming that this relational structure is not projected onto the experience in reflection, assuming that is, that this is a genuine "prepredicative" structure of experience, the contrast between the subject-pole and the object-pole is manifest, even if it normally remains unthematized or attended to as such. On the hypothesis that consciousness is always self-manifesting,

there is no problem here. The relevant contrast is *like* the contrast between the parts and their unified whole. The parts are manifest. The whole is manifest (self-manifest). So all the needed elements are present for their relations (of differentiation, unification, and inclusion) to be manifest.

Moreover, the idea that the difference between the parts and the whole is prepredicatively manifest is no more implausible than the idea that the difference between parts and other parts is prepredicatively manifest, something almost no one would deny. If I see a red patch on a black background, I have a differentiated, contrastive visual experience. The same goes for differences between the sensory modalities: we see and hear simultaneously, etc. If those sorts of contrasts can figure into the ground level of experience, why not the contrast between the unified self-manifesting whole and all its manifest “parts”—the totality of simultaneously manifesting qualities (however we understand them exactly) in all modalities (sensory and possibly cognitive, conative, and affective)?

Subjective character then, on this view, is just the self-manifesting character of an episode of consciousness. This view has the nice feature that it allows us to simultaneously account for the Humean-Buddhist “no-substantial-self” intuition and the intuition of relationality, with its attendant minimalist “sense of self”—as subject-pole.<sup>29</sup> It does this with less metaphysical cost than self-entity and self-haecceity theories, even supposing that those theories are not entirely phenomenologically implausible and explanatorily bankrupt. Let’s remember, however, that this is meant as a phenomenological claim fundamentally: consciousness is self-manifest just as the unified totality of sensory qualities (etc.) is manifest; and their contrast is manifest too, just as the contrast between such qualities (etc.) is manifest. This phenomenological claim has an ontological significance only if we accept that consciousness is indeed how it seems to be upon reflection. A claim that I accept in this case, but one need not accept it to appreciate

the phenomenological point and the virtues of this way of articulating it.

#### 4 From self-representation to self-acquaintance

I gave up on reductive self-representationalism for quite general reasons, reasons affecting all representationalisms. As such, one might be tempted to suggest adopting some non-reductive form of S theory. For example, if one adopts the phenomenal intentionality<sup>30</sup> view, one might hold that whatever phenomenal representation *is*, consciousness represents itself in *that* way. It seems like this view might be just another way of describing the same phenomenological facts belabored in the previous section. If that is so, the phrases “phenomenal intentionality” and “acquaintance” are going to be basically synonymous, and the advocate of the former terminology can just translate. If we build nothing into the notion of representation other than the idea that something (an object, property, episode of consciousness, or whatever) is phenomenally manifest (to someone), then the views are indistinguishable at the phenomenological level and, maybe, the ontological level as well.

If this is not what is intended, however, then it is probably because the phenomenal intentionality theorist wants to mark an important distinction between intentionality (representation) and acquaintance. Perhaps they would prefer not to be committed to acquaintance if possible, and there are several reasons they might want to avoid such a commitment. But I will argue that in a certain sense, to be plausible at all, all forms of representationalism, reductive and non-reductive (including a phenomenal intentionality-based representationalism), ought to embrace a type of acquaintance relation.

Consider, for a moment, fictionalist representationalism about sensory qualities (projectionism about colors, for example). One could embrace a view according to which the sensory qualities are phenomenally manifest, though they in fact are never really instantiated by

<sup>29</sup> I defend this view also in Williford (2011a, 2011b) and in Williford et al. (2012); Dreyfus (2011) is an articulation and defense of a similar view from a Buddhist perspective.

<sup>30</sup> See e.g., Kriegel (2011) and the papers in Kriegel (ed.) (2013), as well as Kriegel’s excellent introduction to that volume.

anything. In such a case, one would not want to think of sensory phenomenal consciousness as a matter of bearing a real acquaintance relation to such qualities or quality instances. Instead one might prefer an adverbial construal of the situation that avoids any commitment to anything literally having (or perhaps even to there being) the properties phenomenally represented. On this view, one denies that there is a relation that supports existential quantification over these immediate objects (whatever they are), and one cannot conclude from the fact that one is phenomenally conscious of a red patch that there exists a red patch of which one is conscious.

Of course, this failure of existential quantification won't apply in the case of one sort of object, namely the conscious episode itself. But it will not be *because* it is an object of phenomenal intentionality that one can validly, existentially generalize from it; generally that fails, just as in other intentional (and intensional) contexts. Rather, it will be because it is the subject or *bearer* of phenomenal intentionality that one can validly generalize from it. In other words, we take episodes of consciousness to be individuals that have this pseudo-relational property. That is why we can quantify over them, and not because of anything that they pseudo-bear that pseudo-relation to. Such “objects,” after all, can be nonexistent. Thomas Reid’s “ambulo ergo sum” would be appropriate here, not the Cartesian *Cogito* conceived in a phenomenologically performative way.

This situation is rather paradoxical. If the only mode of awareness of our own consciousness (even supposing ubiquitous self-manifestation) is via phenomenal intentionality so construed, then our evidence for the very existence of our own consciousness is really no better than our evidence for the existence of phenomenal colors. Just as we might be persuaded that there really are no phenomenal colors, perhaps we could become persuaded that there is no such thing as phenomenal consciousness either. I regard this as absurd. It is like saying that perhaps we only think we think, or that perhaps it only appears to us that things appear to us. Consequently, consciousness must bear some evidentially relevant relation to itself and to its own being, other than the phenomenal

intentionality pseudo-relation it pseudo-bears to phenomenal colors.

Thinking of consciousness as “being-appeared-to-existingly” does not help here, since that applies to phenomenal colors and all other perceived pseudo-objects and pseudo-qualities as well. Any theorist committed to self-manifestation should not try to construe this as just a case of phenomenal intentionality as just described. From our self-consciousness we can conclude that we do exist, and this is not just because we know by inference or in some other way that we are the bearer of a property, as in Reid’s *Ambulo*. We must be acquainted with our own existence—in the sense that every episode of consciousness, however individuated, is acquainted with its own existence. This applies to the subject-pole. What about the object-pole?

In the context of the theory of perceptual consciousness, I think it is a mistake to maintain that any view according to which one can always legitimately quantify over the “immediate objects of conscious awareness” is committed either to some form of direct realism (or perhaps a disjunctivist version thereof) or to old-fashioned sense-datum theory. Any plausible form of representationalism—fictionalist or realist, externalist or internalist, reductive or non-reductive, is, I’ll argue, committed to such quantification, though this must be understood in a particular way. I am not, of course, saying that if we seem to consciously visually perceive a pink rat then we can infer that there exists a pink rat that we see. There is, however, something other than just the conscious state itself (*qua* whole) that we can legitimately, existentially quantify over.

Our conscious perception of differentiation (in unity) entails, even on a representationalist view, that there exists something of which we are aware, namely, at the least, differentiation (or contrast) itself. For example, suppose I hallucinate purple and pinkish smoke clouds arising from stereo speakers as “Fairies Wear Boots” comes on. Evidently I cannot conclude that those purple and pinkish clouds exist. Still, I maintain, we can conclude that there exists some differentiation or contrast of which we are aware. By hypothesis, we cannot say that the difference is that between the pink smoke cloud and the purple one, since they

do not exist. Differences between non-existent objects cannot be appealed to in order to make sense of real differences.<sup>31</sup> But we are aware of some *real* and phenomenally manifest differentiation here. If we say no to that, we'd have to assume that reflection is simply inaccurate when it comes to such hallucinations; that we seem to have a differentiated experience when in fact there is no phenomenal difference at all. But if that itself is a phenomenal state, say a conscious reflection on an ongoing hallucination, we have the same problem all over again.

If the difference we are aware of is not and is not to be accounted for by a difference in the objects (since they do not exist), it must be a matter of the difference in the representations. Hence, albeit in an indirect manner and, as it were, under the guise of a difference in the pink and purple clouds, we must be aware of some differentiation inherent in the representational states themselves.<sup>32</sup> If we reject disjunctivisms, then we ought to maintain that in every case of differentiated phenomenal awareness we are, in fact, acquainted with (and not merely representing) the differences inherent in our episodes of phenomenal consciousness. This is, at any rate, what I think is the most plausible account, even if the considerations just given don't absolutely clinch it. Again, it is not that there cannot be some sort of representationalist response.<sup>33</sup> It is, rather, that I regard the line I take to involve fewer epicycles.

We cannot make good sense of the appearance of a phenomenal difference without direct awareness of differentiation. But, by hypothesis,

<sup>31</sup> We could possibly hold that even if the property instances are not real, the universals represented are, and try to account for the difference in phenomenology in terms of those real differences. But this sort of view does not allow us to make sense of the concrete but hallucinatory representation of different particular instances of the different properties.

<sup>32</sup> I have briefly made similar arguments in Williford (2013).

<sup>33</sup> In particular, a representationalist could say that the represented difference between the pink and purple clouds is just as hallucinatory as the clouds themselves. This is, in a sense, correct. However, representationalists hold (or ought to hold, anyway) that phenomenal differences *always* correlate with differences in the representations themselves (and only normally in the objects of representation). If there are phenomenal differences, there exist some differences inherent in consciousness that are not merely the objects of representation. What I am claiming is that we are acquainted with this differentiation under the guise of differences in objects represented. An adherent of the Transparency Intuition would deny this, of course. And I don't take these considerations to constitute a knock-down argument. (Thanks to an anonymous reviewer for bringing this up.)

in the case of hallucination it cannot be that we are aware of a real difference in the objects of representation. Moreover, it cannot be a difference in something that is *hidden* from conscious awareness—some difference in the externalist conditions determining the content of the representational states, for example—that we are aware of. The most plausible candidate, then, is that we are directly aware of (acquainted with) differentiation or modifications in consciousness itself (and hence the Transparency Intuition (see page 4) is, strictly speaking, false; we are indeed aware of features of consciousness itself even in so-called “first-order” awareness). This applies to both reductive and non-reductive forms of representationalism. If this line of thought is correct, representationalist theories really presuppose some sort of non-representationalist, acquaintance theory.

Implicit in the above discussion is something like this definition of acquaintance:

Acquaintance =<sub>Df</sub> (1) the relation (R) the subject (s) of consciousness (i.e., the episode or stream itself) bears to the differentiated phenomenal manifold ( $D\langle x_1, x_2, \dots, x_n \rangle$ ), such that (2) if  $sR[D\langle x_1, x_2, \dots, x_n \rangle]$ , then we may infer truly that  $(\exists x)(sRx)$ .

Of course, clause (2) can be taken as redundant, given the usual understanding of real relations and that the R of clause (1) is so taken. But in this context it is important to emphasize the point. The first clause is just an inner-ostensive phenomenological characterization that assumes that the relational appearances are indeed the reality; the second is a logico-ontological characterization. Importantly, we can “quantify in” here: If, in any concrete particular case, we stand in that relation to some phenomenally differentiated field, then we can truly infer that *there exists* something differentiated we stand in that relation to. However, it is in general *not* the represented (or intentional) objects that we are thus acquainted with. It is, rather, the common factor of all episodes of phenomenal consciousness, be they hallucinations, dreams, or the “perceptions” of brains in vats. This, again, is often precisely what is denied when one says

that a state is one of representation as opposed to acquaintance. If it is true that *I represent A*, I cannot infer from this that *there is some X such that I represent X*. Adverbialisms and other forms of representationalism were, recall, developed precisely around this insight in order to overcome the problems of sense-datum and other relational theories of perception. Is the theory I am suggesting here a form of old-fashioned sense-datum theory?

Unfortunately I cannot give a short answer to that question and can't give all of the long answer here. This will have to suffice: (1) We can regard sensory qualia (or hyle) as being complex, relational properties of consciousness (and its concrete embodiment in brain processes); in fact, they could be something like irresolvable structural properties that appear simple precisely because they mark a limit of our sensory resolution. (2) In order to flesh this out, we must reject the Revelation Thesis—the thesis that acquaintance yields up all of the properties of sensory qualia. In particular, we can (and should) reject the idea that acquaintance tells us all of the categorial properties of sensory qualities. There is no good reason to believe that it does. Hence, they could fail to seem relational and yet still be relational. This is a solution to the “Grain Problem”—a problem arising from the fact that brain properties are “complex” and relational while sensory qualities (phenomenal colors, tones) do not seem to be. If we infer from the appearances then we cannot consistently hold that they are identical to brain properties. But we have no good reason for making that inference.<sup>34</sup> (3) It is not hard to understand why the sensory qualities would be integrated into a spatialized and “intentionally animated” grid that can serve as a “user interface” for us to deal with the external world, yielding a “transparent” manifold in Metzinger’s sense, a manifold we are built to systematically and automatically “see right through”—causing us to suffer from a sort of delusion of direct realism (see Williford et al. 2012; Williford 2013; Metzinger 2004, p. 163, and Revonsuo

2006). Finally, (4) appeals to the “Transparency Intuition” (in Tye’s sense of “transparency”) thus carry no serious weight. All the phenomenological data in question are accounted for by 1–3, and there are good independent lines of reasoning for each of these (that we do not have time to go into here).

I’ve argued that the notion of acquaintance, when interpreted in the rather minimal, phenomenological, and logico-ontological way proposed, is the proper notion for characterizing the relationship between consciousness and the differentiated but unified multimodal experiential manifold. Moreover, on the view proposed here, consciousness bears this same relation, generically understood, to itself.

If the episode of consciousness bears the relation to itself, then evidently there is something to which it bears that relation. But, non-trivially, we could not have the sort of direct evidence of its existence that we do have if consciousness were not self-acquainted—and acquainted with its own existence. And if the episode of consciousness bears the relation to the differentiated manifold that constitutes the surface that serves as its contact with a differentiated reality beyond it—i.e., if it bears it to a differentiated portion of itself—then there is something differentiated of which it is non-representationally aware. One is directly aware of the *difference* or *differentiation* even if one only, strictly speaking, *represents* what the things so differentiated happen to be or interprets them as being such and such (mental, physical, surfaces of objects, internal sense data, quotidian objects, etc.). In other words, I can see that red is not blue even if I do not know what colors are exactly, or if they are in physical space or only in a virtual space in my brain. One does not *merely* represent this difference or differentiation. One is acquainted with oneself and with the differentiation one contains. Of course, one is also acquainted with the apparently intrinsic properties that mark these internal differences, but again, this need not mean that the properties are in fact non-relational and simple. In fine, we are self-acquainted and acquainted with a differentiated manifold and thus, at some level, with real differences in the mind, the

<sup>34</sup> I’ve argued this is in a bit more detail in Williford (2013). For relevant background ideas see Williford (2005 and 2007). For a discussion of the Revelation Thesis see e.g., Stoljar (2006, ch. 11) and Goff (forthcoming).

world, or world-mind boundary.<sup>35</sup> The acquaintance relation consciousness bears to itself is, generically speaking, identical to the relation it bears to sensory qualia (or hyle)—which are taken here as ultimately just transient modifications in the unfolding embodiment of consciousness. It is important to understand that this does not imply that there is a special type of sensory quality (a “me-ish” quale) peculiar to consciousness. It is as diaphanous as G.E. Moore said. Remember that the acquaintance in an instance of acquaintance with phenomenal red is identical with the acquaintance in an instance of acquaintance with phenomenal C#, even though phenomenal red and C# are utterly heterogeneous.

One might reasonably ask for a more substantive definition or account of acquaintance. The definition given relies on phenomenology and logic and is otherwise quite empty. But this is as it should be, in my view. Any further account of the nature of acquaintance, of what the acquaintance relation *is*, will be the result of empirical inquiry and a well-supported *a posteriori* identification.

## 5 Self-acquaintance, subjective character, and individuation

Earlier I briefly noted that at the phenomenological level we should probably not construe subjective character fundamentally as a matter of “mineness” or a “sense of self” where the latter is thought of as a sense of oneself as an owner of experiences. It is not that I do not think this description contains a grain of truth; I do. The worry, though, is that if we go this route, we might come to the conclusion that subjective character involves acquaintance with a haecceity—Zahavihood and Gallagherhood once again. Here I want to consider the same issue from a more ontologically oriented point of view.

We are indeed individuated and aware of ourselves (something individuated). And we can be aware of ourselves *as* distinct individuals and owners. But this does not at all entail the doctrine of haecceities immediately present to con-

sciousness—for-me-ness or me-ishness as a special property that no one else can share. Rather, subjective character is a common form that all conscious states have; but having this form does not alone make something the individual it is, evidently. It may be that in virtue of which we can be *aware* of ourselves as individuals, but it is not that in virtue of which we *are* the individuals we are. Yes, there is a determinate individual (somehow construed) that is acquainted with itself. No, this does not necessarily mean that it is acquainted with that *in virtue of which* it is individuated. That could be whatever it is that individuates physical objects. Or, perhaps, nothing is metaphysically individuated *by* anything else. But it ought to be clear that simply in being aware of myself I need not be privy to anything non-trivial about my metaphysical individuation conditions.<sup>36</sup>

You are aware of your consciousness as something individual. You are a self-aware individual, if you prefer. But this does not mean that your subjectivity consists in being directly aware of *what* individuates you or the very property in virtue of which you are the individual you are. Or, perhaps, one may be aware of this property or set of properties, but only in the guise of being an individual that is thus and so. The “thus and so” part (all your contingent properties, your “facticity”) is radically changeable. You need not have been thus and so. (You could have been a contender! And if only you’d been rich!) You can also be aware that you are a particular instance. So, yes, you can become aware of your particularity. But everybody is aware of their own particularity. And it is, in a way, an empty and non-material (in the “formal vs. material” sense) property. It’s not as if my particularity has a special something that yours lacks and vice versa. Hence, I would not be able to tell, by phenomenological intuition alone (or in any other way for that matter), which of the infinitely many duplicate and near-duplicate worlds I am in (cf. Elga 2004). Am I in the world in which one of Napoleon’s buttons had a bit of his blood on them the morning of the Battle of Jena or in the world in which that was

<sup>35</sup> I have considered our acquaintance with a differentiated manifold *qua* mind-world boundary in more detail in Williford (2013).

<sup>36</sup> I have briefly argued this before in Williford (2011b). I was pleased to find that a similar line of argument was pursued by the eleventh-century Buddhist philosopher Ratnakīrti; see Ganeri (2012, p. 217).



not the case? I cannot tell by introspection, yet, depending on the correct answer, I am one type of individual (and of course, one token of uncountably many of that type) and not of the other type (which type also contains uncountably many individual counterparts of mine). I am individuated, and I know that; I belong to just one of these worlds. But I do not have complete access to my individuation conditions or the conditions, if there are any, that determine that this individual is in one world as opposed to another. I have uncountably many counterparts who feel exactly the same way because, to speak loosely, they don't know that they are not me; none of us can tell the difference. I cannot locate my Homeworld on the map of worlds that contains my relevant counterparts.

It is a mistake, then, to make subjective character depend on the sense of individuality; this reverses the proper order of explanation. Self-acquaintance and concrete instantiation yield the sense of individuality, and they do it again and again in many places and in the same way. Evidently, the contingent filling that experience and history infuse into the formal shell of conscious subjectivity is not relevant at the level we are concerned with. Hence, it can also be metaphysically, not just phenomenologically, misleading to use terms like “for-me-ness,” “mineness,” “me-ishness,” etc. That is to make something derived seem like something basic. The basic things are self-acquaintance (“reflexivity”) and actual, concrete instantiation or constitution. The sense of individuality comes from these, not the other way around.

Of course, if you are a real, concrete individual, you are individuated. But individuation is evidently not self-acquaintance. The latter is, however, required if one is to get the *sense* of being an individual, to know, feel, and be concerned with oneself *as* an individual. If we generally equated self-acquaintance with something's being the individual it is, then we'd have to hold either that every individuated thing in the cosmos is self-acquainted and conscious, or that conscious things have one type of metaphysical individuation conditions, and non-conscious things another, for very obscure reasons. Moreover, we either must not take subjective character to be a

univocal notion or must resort to some sort of hopeful brute resemblance nominalism about subjective character and maintain that we cannot not really know that, say, I, *qua* subject, am in any meaningful sense like you, *qua* subject. This is not a very good dilemma to be in.<sup>37</sup> I think the more plausible view is that self-acquaintance is not the source of the individuation of consciousness but rather something that both concretely depends upon individuation and enables the knowledge of individuality and, consequently, self-location in surrounding spaces.

It is misleading, then, both phenomenologically and ontologically to refer to subjective character principally as “mineness” or “me-ishness” or “for-me-ness,” even though subjective character is one of the bases of the sense of individuality. We should not think of self-acquaintance (and subjective character) as anything more than this relation all episodes of consciousness bear to themselves. It is a perfectly uniform structure and a kind of universal—in that sense, supposing one is some sort of realist about universals, there is indeed some identical thing that unifies all episodes (or subjects) of consciousness, namely the very property of being self-manifesting; but we are all distinct instances. Thus, in a very special and non-Vedantist sense, we could say that there are many instances of consciousness but only one subject, with some instances connected to each other and grouped together in other important ways as well. But there is no substantial self. In this regard, I am with Hume, Sartre, Parfit, Strawson, Metzinger, the Buddhists, and other “non-egological” theorists of consciousness. Note that this does not mean that consciousness is “anonymous” in the sense of “subjectless.” Every stream of consciousness has its transient subject (*viz.*, itself) but that is not a substantial self.

## 6 Self-acquaintance, intrinsic properties, and physicalism

Should we really regard self-acquaintance as a relational matter? Is it really a matter of some

<sup>37</sup> Previous episodes of consciousness normally connected to the present episodes (the ones producing this document) found themselves trying to live with the latter horn of the dilemma in the flawed Williford (2005).

sort of thing standing in a relation to itself? On the one hand, there is no special problem either logically or phenomenologically speaking with the idea of something relating to itself in this way. Appearance is appearance-to. That's relational. There is no *a priori* reason why something could not appear to itself. It does not lead to a regress.<sup>38</sup> One should put aside misleading and question-begging spatial analogies—consciousness is not like a knife trying to cut itself. Advocates of self-acquaintance will claim, opposing one analogy with another, that it is more like a candle's flame illuminating itself by emission while it illuminates other things by the reflection of its light; it does not require another candle flame for it to be illumined.<sup>39</sup> Moreover, one must remember to exclude from one's mind the sort of objectification and description-based cognition that normally overlays the phenomenal manifold. We are talking about the sphere of immanence, to speak Husserlian, and not about intentional objects or constituted objectivities given via *Abschattungen*. Again, we are talking about immediate self-acquaintance, not the representation of oneself as being such and such. It is indeed more like the emission of light than the reflection of light, if we must pick an analogy.

Nevertheless, even if we accept the relational construal and remember that it is an immediate and direct relation not mediated by concepts or descriptions, we still have a problem. It is not as if conscious episodes just happen to be self-manifesting. The property of being self-manifesting is not something that a thing can have and then not have—like changing coats of paint. It is of the very essence of a conscious episode. This is not an external relation to itself or one mediated by convention or history or anything else. Hence, it must have some set of intrinsic properties in virtue of which it is self-manifesting. Thus, the Heidelberg School, Michel Henry, and Dan Zahavi, I'll

concede, win on this ontological point. Dieter Henrich, Manfred Frank, Henry, and Zahavi have all maintained that self-manifestation could not be a relational matter (e.g., [Henrich 1971, 1982](#); [Frank 2002, 2007](#); [Zahavi 1999](#); [Henry 1973](#)). And they are very close to being right. I think, however, that it is more accurate to say that even if it is a relational matter, it is not an *external* relation we are dealing with. So there must be something about the internal structure of consciousness that grounds the relation. In short, as Henrich and Frank have long said, there must be some intrinsic property in virtue of which episodes of consciousness (out of all other things in the world) are self-manifesting. What could this property be? Are we left with something that cannot be physical, or, even if it is physical, is nevertheless irreducible in some sense?

It may seem now that David Rosenthal is having his revenge.<sup>40</sup> In effect, I have been arguing against the extrinsicalist view—the view that something's being conscious has to do with external relations the thing stands in—be those external relations to other mental states or external relations to historically distant states of affairs or to other parts of one's cognitive apparatus. Now, to our chagrin, it seems we are left with something explanatorily basic. At this point we are left with two problematic strategies. We could go the panpsychist route ([Strawson 2006](#)): It's no surprise that we're conscious if everything is! Or if, as I do, one thinks (after Locke in a similar context) that “every sleepy nod doth refute” this, we can hold that only certain physical complexes instantiate this particular property (or set of properties). This will mean either some form of property dualism or some form of identity theory (possibly with its “Harder Problem”; see [Block 2002](#)). If one does not want to be a dualist or a panpsychist, what can be said?

Here is the sort of approach that seems most attractive (to me, anyway). We want to hold that consciousness is indeed some sort of physical process. It's not, however, just a matter of the satisfaction of some functional role. I

<sup>38</sup> This is demonstrable. First, obviously, there is no logical problem with reflexive relations. Second, it requires special and highly questionable premises to generate another regress here. See [Williford \(2006\)](#). See also [Kriegel \(2009, p. 124\)](#) and [Janzen \(2008, p. 110\)](#).

<sup>39</sup> The knife blade and candle flame competing analogies loom large in the Indo-Tibetan debate on this issue. Clearly, the analogies will be found, by opponents and proponents, to be exactly as plausible as the views they encode.

<sup>40</sup> Though even Rosenthal's own view was pushed into being (or always was) problematic in this regard, as noted above.

think it also has a functional role. But it is not *in virtue of* playing that role that something is consciousness; rather, consciousness is suitable for that role because of its properties.<sup>41</sup> In principle, many different things could play that role (at least if we specify it entirely in behavioral terms). Or, at least, this is an open question. Consciousness has a functional role, but it is not to be identified with just any arrangement of elements that can play that role as causally and behaviorally specified. There is some special, distinctive physical process that is consciousness. It plays its functional role in virtue of its having the properties it does and *not vice versa*. But then does some version of Russellian Monism start to seem attractive (see e.g., [Stoljar 2006](#), ch. 6 and [Pereboom 2011](#), chs. 5 and 6)? Am I saying that the functional role is just being (contingently) satisfied by a (somehow) unified and self-manifesting group of qualia? Or something *wild* like that?

Here we play the same sort of trick we played when dealing with the Grain Problem. Consciousness is self-acquainted, but we are also, as Fumerton and Fales would say, acquainted with acquaintance; we are given givenness ([Fumerton 1985](#), pp. 57–58 and [Fales 1996](#), pp. 147–148). The relation does not seem complex or to involve many layers of relational structures. But we cannot infer from this appearance that it is in fact such a simple relation. Again, its not seeming complex does not, without controversial and implausible completeness assumptions, entail its being simple. Moreover, once we realize that normal consciousness involves a great many intricately related aspects—at least (non-contingently) differentiated unity and temporality, and (contingently) animation functions operating on a differentiated sensory manifold, iterations of these functions, pattern extractions, etc.—we have all the more reason to suppose that there is complicated machinery hidden from our introspective view. In fact, it will be noted in a Sartrean and Moorean vein, that consciousness, both as acquaintance relation and subject-relatum, seems mightily empty. Once we realize that Revelation theses fail, then we no longer need read this ap-

pearance as “consciousness *qua* acquaintance relation appears simple.” Rather, we read it as “consciousness *qua* acquaintance relation *does not appear* complex.” These are, in many cases, phenomenologically indistinguishable, but they are logically different.<sup>42</sup> The first reading, coupled with an infallibility thesis (or with just a strong presumption in favor of the deliverances of naïve introspection), leads to the view that acquaintance is simple. But the other requires a Revelation (or completeness) thesis to get the same result. Revelation is, again, totally implausible. And even if we were to assume infallibility, we have no a priori reason to favor one interpretation of the phenomenological data over the other—the “seeming non-P” vs. “not seeming P” formulation. We do, however, have plenty of *a posteriori* inductive reasons for preferring the latter: It does not seem complex, but it is (or at least could be for all we can tell phenomenologically).

Since we have an extremely limited resolution when it comes to penetrating into the nature of consciousness by introspective means, we are quite free to adopt another strategy. We can accept an *a posteriori* identity theory. Consciousness is identical to some sort of recurrent physical process unfolding in the brain. Fundamentally, what we get from introspection is a sort of structure and some irresolvables—the sensory qualities—that are like reflections of the materials in which the form or structure is instantiated. Since we have rejected Revelation (completeness) theses, we can accept that sensory qualities (and the acquaintance relation itself) are complex and involve layers of relations even though they do not seem this way (just as the headless woman<sup>43</sup> in the famous illusion does not seem to have a head—absence of appearance is transformed into the appearance of an absence; see [Armstrong 1968](#) and [1973](#)).

<sup>42</sup> They are phenomenologically indistinguishable in the way that the stream of consciousness’s being temporally continuous is, plausibly, phenomenologically indistinguishable from consciousness’s being punctate or discrete, or in the way in which consciousness’s seeming free from causal determination is phenomenologically indistinguishable from its simply not seeming determined (because the causal relations are inaccessible, as Spinoza suggested).

<sup>43</sup> See the following links:  
<https://www.youtube.com/watch?v=LXOqD5B5Sxc>  
<http://www.deceptology.com/2010/10/headless-woman-illusion.html>

<sup>41</sup> Here I am in considerable agreement with [Langsam \(2011, ch. 3\)](#).

We can use what structure we are aware of, however, to build models to guide our search for the neural correlates of consciousness. One thing we see is that the (only apparently simple) acquaintance relation involved is such that whenever  $xRy$ ,  $xRx$ ; while it is not the case that if  $xRy$ , then  $yRy$  (in the case where  $y$  is a sensory quality or manifold thereof). And we have some idea of what the qualities in the manifold could be—e.g., limits of resolution or irresolvables operated on by a spatializing filter. We can also see that spatial projection, integration of multimodal information, temporality, and the modulation of attention are involved (along, of course, with more advanced things like intentional animation, cognitive filtering and reprocessing, and poise for action). We have a self-manifesting totality containing a unified and spatialized but differentiated manifold. Consequently, we do need to look for processes that can do information integration and binding, but that is only necessary, not sufficient. We need to look as well at processes that spatialize the multimodal (and multidimensional) information (see [Williford et al. 2012](#)).

This does not at all mean we are looking for a little room in the brain that has patches of red, yellow, blue, and green mental paint in it. Rather we must look for more abstract correspondences. In the case of the sensory qualities, we are possibly looking for higher-order relations between fairly complex structures, structures that can transiently be pulled into and “rendered” by the core process. Basically, this panoply of contrast-related irresolvables gets generated in a real-time and transient fashion, now occupying this virtual “location”, now occupying that, depending on a whole host of input factors (head orientation, background, conceptualization, etc.). These “locations” map onto (we hope) real physical space at a certain scale, but it is not a matter of finding a “bubble within a bubble.” It is a matter of an abstract correlation of structure. The isomorphisms (or homomorphisms) could be there even if the internal “space” of experience is entirely virtual, a kind of computational “movie in the brain” to use another phrase of Damasio’s. Assuming the principle that the positive and critically evalu-

ated set of phenomenological descriptions gives us not just the way consciousness seems, but the way it in fact is, along with our identity postulate, we can be sure that something in the brain has a structure corresponding to this, no matter how transformed by “layers of abstraction” it may be.

What is more, self-acquaintance will demand that we explore models in which real reflexivity can be encoded. Hofstadter’s model is one of these.<sup>44</sup> But following D. Rudrauf and further encouraged by D. Bennequin, I have moved in the direction of considering projective geometrical models. There is no space to go into this here, but suffice it to say that there exist mathematical frameworks that allow us to conceptualize and investigate more deeply the self-acquaintance-related features of consciousness by considering the interplay of the space we project and the origin of the projection (see [Williford et al. 2012](#) and [Rudrauf et al. ms](#)).

The goal of such work would be the refinement of mathematical models of the structure of consciousness. Upon the achievement of that end, we would then try to determine how such models could possibly be physically realized in the brain. Once we can say what the physically detectable signatures of such a realization might be, then we could one day meaningfully test such theories. Were we to verify the existence of such a structured process in the brain, explaining consciousness would reduce to explaining how the process is realized—what parts have to be in what order doing what and at what time scale.

It will always seem to be a brute fact, at some level, that consciousness is physical process  $X$ , however  $X$  gets fleshed out. But we’ll just get used to it, as long as there is some somewhat intelligible bridge (in this case provided by mathematical models) from the lived phenomenon to its brain correlate. We’ll get used to it just as we’ve gotten used to water being  $H_2O$ . It could be that there will be multiple ways to implement such a process. Sup-

<sup>44</sup> While Hofstadter’s Gödel-inspired model might be problematic (both in terms of physical implementation and in terms of the strong mathematical realism it might presuppose), it is certainly in the right class of models we should be considering. See [Hofstadter \(2007\)](#).

pose, just for example, that it has to do with generating certain types of fields and that multiple substrates, not just brains, can generate and support the relevant sorts of fields. Then consciousness will be, to that degree, multiply realizable. Suppose it is a matter of realizing a certain computational organization. Then, in effect, implementing a certain program will be equivalent to being conscious; and if machines made from different substrates or with different architectures can run the program, consciousness will be multiply realizable in the sense of computational functionalism. Your particular consciousness then, as you know and love it, would be just the concrete running of the program in your particular brain.

We might wonder, in such a case, what it is to “run” a program or to “have” a certain structure or to “instantiate” such an arrangement or system of fields or whatever. Of course, this is a quite general metaphysical problem that we should not confuse with any problem specific to consciousness. However, given that we are acquainted with our own individual existence, it seems that somehow its instantiation makes its very instantiation available or manifest in some non-representational way. This is rather peculiar. If we are going to be physicalists who are nonetheless responsible to the phenomenology, however, this is what we have to accept, or so I have argued. Something is conscious if it has a certain internal structure and attendant dynamical profile. Being conscious *is* having that structure and profile. We will never be able to explain *why* that is the case because it is simply a confusion to think that identities like this admit of explanation; they can only be discovered (Papineau 2002, ch. 3). We must, of course, give evidence in favor of the relevant identity claim; uncovering such evidence is the goal of scientific research on consciousness. Our choice is between this sort of view and the view that *there is something else, something non-physical* that just *is* consciousness. Of course, we’d never be able to explain why *that* is the case either. So in the absence of compelling arguments for dualism or panpsychism, Occam’s Razor would lead us, as Smart pointed out so long ago, to embrace an identity theory.

The identity theory only adumbrated here would be neither a crude type-type identity theory nor a causal-role functionalist token-token identity theory where the realizers do not matter at all. Since any concrete consciousness is a marriage of form and matter (and the self-appearance of that marriage), and since there no doubt are physical constraints on what sorts of materials can be put into that form, we want to identify consciousness with neither a specific type of material (or “wonder tissue” in Dennett’s phrase) nor with an abstract, disembodied form that seems trivially realizable by practically any set of elements—since purely abstract isomorphisms may be a dime a dozen.<sup>45</sup> In other words, we need a non-eliminativist and non-idealist account of what it is to *really* realize a structure, instantiate a form, or, as the case may be, to *really* run a program or compute a function. To my knowledge, no one currently has such an account.

At bottom, this is just the old metaphysical problem of the Methexis—the relation of universals to particulars or of form to matter. When I am feeling optimistic, I imagine that I’ve reduced the problem of consciousness to another, more general (as well as ancient and probably insoluble) metaphysical problem. We may not know what it is for matter to really and mind-independently take on a certain form, but it is hardly an implausible metaphysics that says that this happens. It is arguably this type of metaphysical view that would best explain the success of applied mathematics, engineering, and the sciences: they are successful because the world really does have (or approximate) the relevant mathematical structures—these are *in re* structural universals. This seems to be a commitment of scientific realism. But perhaps we will never get beyond a rather crude operationalism when we empirically investigate such matters; perhaps the metaphysical nature of property instantiation will forever remain obscure to us. That should not, however, discourage us from carrying on such empirical investigations in the case of consciousness. Even if there will be

<sup>45</sup> For discussion, see Chalmers (1996) and Buechner (2008, ch. 3).

a residual metaphysical mystery, it is a general one, not one specific to consciousness.

The main point here, and the concluding one, is that consciousness could be self-acquainted, where this is not a matter of external relations, and still some form of relatively non-mysterious (hylomorphic) physicalism could be true. One might balk at the idea that this would not be a matter of external relations, especially if we go the computational functionalist route. But think of it like this: If we are realists about the implementation of computational structures, then even though the structures involve parts and elements, there is still a unity to the pattern as implemented. It is, in a certain sense, an indivisible whole that is not just the mereological sum of its parts. Analogously, the circle has its own structure and characteristic properties even though it is made of points. What we really need, and may never have (but who knows?) is a theory that tells us when we have a real, concrete unified whole, (where this is *not* simply a functional or conventional characterization but is a matter of more basic physical relations) and when we have unities and wholes (and instantiations of structures and properties) that are only conventionally real.

Suppose then that we adopt a sort of realism about computational (or otherwise structural) wholes, which we have some independent reason to do. Circles have remarkable properties, *qua* circles, even if they are made up of points. Concrete circular things approximate these. Simultaneous cycles have certain number-theoretic properties just *qua* cycles regardless of *what* they are cycles of (e.g., reproducing cicadas and cicada predators, see [Baker 2005](#)). Likewise, for the concrete implementation of consciousness, it is surely the case that certain elements must be put into a certain arrangement, realizing a certain structure and dynamics. This would not mean, however, that consciousness *as such* is to be identified with either those elements or the arrangement abstractly conceived. Rather it is the *concretely implemented* organization of those elements *qua* whole. In virtue of being an instance of that form or structure, it has

certain properties. One of these could be the property of being self-manifesting. That property could itself be a complex relational property having a certain unity. The account sketched here presupposes a certain realism about the instantiations of mathematical and computational structures—that there are determinate, mind-independent facts of the matter about this. We cannot go further into this rather large and complicated metaphysical hornet’s nest. Suffice it to say that a real, unified, concretely instantiated structure could, in a certain sense, be relational and have components even if it is, in another sense, an intrinsic property.

## 7 Conclusion

I have argued that the best way to characterize subjective character is in terms of self-acquaintance and not, for various reasons, in terms of Higher-Order, Same-Order, or Privileged-Object representation. I argued that every episode or stream of consciousness is acquainted with itself, and not with a self in some other sense—a homunculus, substance, or haecceity. This is, I maintain, the best way to make sense of the intuition of subject-object polarity and the Humean intuition that we do not find a self-entity. Moreover, one’s sense of being an individual is a consequence of self-acquaintance and concrete existence and not to be conflated with subjective character as such. Such conflation leads to potentially misleading descriptions of subjective character (as “mineness”) and, if taken literally, to metaphysically and epistemologically undesirable consequences. We are individuated and self-acquainted, and that is enough to allow us to derive the sense of self or “mineness”; but self-acquaintance is not itself what individuates us, nor does it necessarily make us aware of what does.

Nevertheless, I conceded to Henrich, Frank, Henry, and Zahavi (among others) that consciousness must have some intrinsic (or internal relational) property in virtue of which it is self-acquainted. But I argued that this does not nullify the appropriateness of de-

scribing subjective character as being a matter of a very complex relation, though it does not seem to be so complex.

Finally, I argued that the position advanced here is not incompatible with a form of (hylomorphic) physicalism. Sensory hyle, the acquaintance relation itself, the self-manifesting episodes, could all be brain processes and properties. On the phenomenological side, this gains plausibility once we take to heart the incompleteness of introspection (and of pre-reflective self-awareness as well): not seeming complex and relational does not entail not being complex and relational. On the ontological side, I argued that even some form of computational functionalism could be true. But, generally, the important thing to remember is that consciousness is the marriage of form and matter. It cannot be simply equated with either. This opens up space for multiple realizability, but it might also mean that not just any old substrate will do. It's an open question. The metaphysical commitment behind this position is just some form of realism about structural universals and their mind-independent instantiation conditions, which is arguably a commitment of scientific realism in any case. Absent dualism, panpsychism, or idealism, that is what we will have to accept, I believe. (Eliminativism is, of course, a non-starter.)

We do not need a theory of the Methexis, however, in order to attempt to find the neural correlates (correlation conceived of as indicating identity here) of consciousness by building mathematical models of the phenomenology and figuring out how the brain might implement the structures so modeled. In fact, just such an approach is quite in line with scientific practice generally: We know that the world we investigate with our relatively crude means is, in multiple ways, a play of matter and form even if we do not really know what the Matter ultimately is, what Forms are, and how the latter come to live in the former.

## Acknowledgements

Different parts of this material were presented at many places over several years. I would like

to thank audiences at the Johannes Gutenberg-Universität Mainz (seventeenth meeting of the MIND Group), the Berlin School of Mind and Brain, the Institut Jean Nicod, ZiF, SMU, the SSPP, TCU, and Tucson TSC for relevant discussions. I would like to thank these institutions and the symposia organizers (and in particular, Thomas Metzinger and Jennifer Windt; Manfred Frank, Marc Borner, Andreas Heinz and Anna Strasser; Brad Thompson and Philippe Chuard; Pete Mandik, Rik Hine, and Blake Hestir; and David Chalmers). Thanks to the College of Liberal Arts and the Department of Philosophy and Humanities at the University of Texas-Arlington for research and travel funding in this connection. I should thank (in alphabetical order) Katalin Balog, Daniel Bennequin, Jacob Berger, Alexandre Billon, Marc Borner, Philippe Chuard, Christian Coseru, Justin Fisher, Manfred Frank, Brie Gertler, Robert Howell, Tomis Kapitan, Bob Kentridge, Chad Kidd, Alex Kiefer, Uriah Kriegel, Greg Landini, Stefan Lang, Pete Mandik, Thomas Metzinger, Charles Nussbaum, David Papineau, Gerhard Preyer, Harry Reeder, David Rosenthal, Amber Ross, David Rudrauf, Susan Schneider, Miguel Sebastian, Charles Siewert, Anna Strasser, Brad Thompson, Keith Torausky, Michael Tye, Josh Weisberg, and Dan Zahavi for discussions, questions, criticisms, suggestions, etc., that were in one way or another of help to me in relation to the material presented here. In the same regard, I should thank two anonymous reviewers from the MIND Group for helpful feedback on an earlier version of this article; their feedback helped me to see some of my less-than-admirable tendencies as a writer of philosophy, even if it did not enable me to correct all their manifestations. Special thanks to Ying-Tung Lin of the MIND Group for her help. Special thanks to Trish Mann, Swathi Prabhu, Emma Nwokonko, and Anya Williford for help with the references. And very special thanks, once again, to Thomas Metzinger and Jennifer Windt for launching and managing this unique and ambitious project and to the Barbara-Wengeler-Stiftung for its support.

## References

- Armstrong, D. (1968). The headless woman illusion and the defence of materialism. *Analysis*, 29 (2), 48-49. [10.1093/analys/29.2.48](https://doi.org/10.1093/analys/29.2.48)
- (1973). Epistemological foundations for a materialist theory of the mind. *Philosophy of Science*, 40 (2), 178-193. [10.1086/288514](https://doi.org/10.1086/288514)
- Baars, B. (1997). *In the theater of consciousness: The workspace of the mind*. New York, NY: Oxford University Press.
- Baker, A. (2005). Are there genuine mathematical explanations of physical phenomena? *Mind*, 114 (454), 223-238. [10.1093/mind/fzi223](https://doi.org/10.1093/mind/fzi223)
- Balog, K. (2012). Acquaintance and the mind-body problem. In C. Hill & S. Gozzano (Eds.) *New perspectives on type identity: The mental and the physical* (pp. 16-42). Cambridge, UK: Cambridge University Press.
- Berger, J. (2013). Consciousness is not a property of states: A reply to Wilberg. *Philosophical Psychology* (ahead-of-print), 1-14. [10.1080/09515089.2013.771241](https://doi.org/10.1080/09515089.2013.771241)
- Block, N. (1995). On a confusion about a function of consciousness. *Behavioral and Brain Sciences*, 18 (2), 227-247. [10.1017/S0140525X00038188](https://doi.org/10.1017/S0140525X00038188)
- (2002). The harder problem of consciousness. *The Journal of Philosophy*, 99 (8), 391-425. [10.2307/3655621](https://doi.org/10.2307/3655621)
- (2011). The higher order approach to consciousness is defunct. *Analysis*, 71 (3), 419-431. [10.1093/analysis/anr037](https://doi.org/10.1093/analysis/anr037)
- Buechner, J. (2008). *Gödel, Putnam, and functionalism: A new reading of representation and reality*. Cambridge, MA: MIT Press.
- Butchvarov, P. (1979). *Being qua being: A theory of identity, existence, and predication*. Bloomington, IN: Indiana University Press.
- (1998). *Skepticism about the external world*. Oxford, UK: Oxford University Press.
- Byrne, A. (2001). Intentionalism defended. *Philosophical Review*, 110 (2), 199-240. [10.1215/00318108-110-2-199](https://doi.org/10.1215/00318108-110-2-199)
- Cappelen, H. & Dever, J. (2013). *The inessential indexical*. Oxford, UK: Oxford University Press.
- Carruthers, P. (2000). *Phenomenal consciousness: A naturalistic theory*. Cambridge, UK: Cambridge University Press.
- (2005). *Consciousness: Essays from a Higher-Order Perspective*. Oxford, UK: Oxford University Press.
- Caston, V. (2002). Aristotle on consciousness. *Mind*, 111 (444), 751-815. [10.1093/mind/111.444.751](https://doi.org/10.1093/mind/111.444.751)
- Chalmers, D. (1996). Does a rock implement every finite-state automaton? *Synthese*, 108 (3), 309-333. [10.1007/BF00413692](https://doi.org/10.1007/BF00413692)
- (2003). The content and epistemology of phenomenal belief. In Q. Smith & A. Jokić (Eds.) *Consciousness: New philosophical perspectives* (pp. 220-272). Oxford, UK: Oxford University Press.
- Clark, A. (1989). The particulate instantiation of homogeneous pink. *Synthese*, 80 (2), 277-304. [10.1007/BF00869488](https://doi.org/10.1007/BF00869488)
- Coseru, C. (2012). *Perceiving reality: Consciousness, intentionality, and cognition in buddhist philosophy*. Oxford, UK: Oxford University Press.
- Crowell, S. (2011). Ideality of nature: Jan Patočka on reflection and the three movements of human life. In I. Chvatík & E. Abrams (Eds.) *Jan Patočka and the heritage of phenomenology* (pp. 7-22). Berlin, GER: Springer.
- Dainton, B. (2000). *Stream of consciousness: Unity and continuity in conscious experience*. London, UK: Routledge.
- (2008). *The phenomenal self*. Oxford, UK: Oxford University Press.
- Damasio, A. (1999). *The feeling of what happens: Body and emotion in the making of consciousness*. New York, NY: Harcourt.
- (2010). *Self comes to mind*. New York, NY: Pantheon Books.
- Dehaene, S. & Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework. *Cognition*, 79 (1), 1-37. [10.1016/S0010-0277\(00\)00123-2](https://doi.org/10.1016/S0010-0277(00)00123-2)
- Dretske, F. (1995). *Naturalizing the mind*. Cambridge, MA: MIT Press.
- Dreyfus, G. (2011). Self and subjectivity: A middle way approach. In M. Siderits, E. Thompson & D. Zahavi (Eds.) *Self, no self?: Perspectives from analytical, phenomenological, and Indian traditions* (pp. 114-156). Oxford, UK: Oxford University Press.
- Edelman, G. & Tononi, G. (2000). *A universe of consciousness: How matter becomes imagination*. New York, NY: Basic Books.
- Elga, A. (2004). Defeating Dr. Evil with self-locating belief. *Philosophy and Phenomenological Research*, 69 (2), 383-396. [10.1111/j.1933-1592.2004.tb00400.x](https://doi.org/10.1111/j.1933-1592.2004.tb00400.x)
- Fales, E. (1996). *A defense of the given*. New York, NY: Rowman & Littlefield.
- Frank, M. (2002). Self-consciousness and self-knowledge: On some difficulties with the reduction of subjectivity. *Constellations*, 9 (3), 390-408. [10.1111/1467-8675.00289](https://doi.org/10.1111/1467-8675.00289)



- (2004). Fragments of a history of the theory of self-consciousness from Kant to Kierkegaard. *Critical Horizons*, 5 (1), 53-136. [10.1163/1568516042653567](https://doi.org/10.1163/1568516042653567)
- (2007). Non-objectal subjectivity. *Journal of Consciousness Studies*, 14 (5-6), 5-6.
- Fumerton, R. (1985). *Metaphysical and epistemological problems of perception*. Lincoln, NE: University of Nebraska Press.
- Ganeri, J. (2012). *The self: Naturalism, consciousness, and the first-person stance*. Oxford, UK: Oxford University Press.
- Gennaro, R. (2005). The HOT theory of consciousness: Between a rock and a hard place? *Journal of Consciousness Studies*, 12 (2), 3-21.
- (2006). Between pure self-referentialism and the (extrinsic) HOT theory of consciousness. In U. Kriegel & K. Williford (Eds.) *Self-representational approaches to consciousness* (pp. 221-248). Cambridge, MA: MIT Press.
- (2012). *The consciousness paradox: Consciousness, concepts, and higher-order thoughts*. Cambridge, MA: MIT Press.
- Gertler, B. (2011). *Self-knowledge*. London, UK: Routledge.
- (2012). Renewed acquaintance. In D. Smithies & D. Stoljar (Eds.) *Introspection and consciousness* (pp. 93-128). Oxford, UK: Oxford University Press.
- Goff, P. (forthcoming). Real acquaintance and physicalism. In P. Coates & S. Coleman (Eds.) *Phenomenal qualities: Sense, perception and consciousness*. Oxford, UK: Oxford University Press.
- Goldman, A. (1993). Consciousness, folk psychology, and cognitive science. *Consciousness and Cognition*, 2 (4), 364-382. [10.1006/ccog.1993.1030](https://doi.org/10.1006/ccog.1993.1030)
- Harman, G. (1990). The intrinsic quality of experience. *Philosophical Perspectives*, 4, 31-52. [10.2307/2214186](https://doi.org/10.2307/2214186)
- Henrich, D. (1971). Self-consciousness, a critical introduction to a theory. *Man and World*, 4 (1), 3-28. [10.1007/BF01248576](https://doi.org/10.1007/BF01248576)
- (1982). Fichte's original insight. In D. Christensen (Ed.) *Contemporary German philosophy: Volume 1* (pp. 15-53). University Park, PA: Penn State University Press.
- Henry, M. (1973). *The essence of manifestation*. Den Haag, NL: Nijhoff.
- Higginbotham, J. (2003). Remembering, imagining, and the first person. In A. Barber (Ed.) *The epistemology of language* (pp. 496-533). Oxford, UK: Oxford University Press.
- (2010). On words and thoughts about oneself. In F. Recanati, I. Stojanovic & N. Villanueva (Eds.) *Context-dependence, perspective and relativity* (pp. 253-282). Berlin, GER: De Gruyter.
- Hofstadter, D. (2007). *I am a strange loop*. New York, NY: Basic Books.
- Hopp, W. (2011). *Perception and knowledge: A phenomenological account*. Cambridge, UK: Cambridge University Press.
- Horgan, T. & Kriegel, U. (2007). Phenomenal epistemology: What is consciousness that we may know it so well? *Philosophical Issues*, 17 (1), 123-144. [10.1111/j.1533-6077.2007.00126.x](https://doi.org/10.1111/j.1533-6077.2007.00126.x)
- Howell, R. (2013). *Consciousness and the limits of objectivity: The case for subjective physicalism*. Oxford, UK: Oxford University Press.
- Janzen, G. (2008). *The reflexive nature of consciousness*. Amsterdam, Netherlands: John Benjamins.
- Kapitan, T. (2006). Indexicality and self-awareness. In U. Kriegel & K. Williford (Eds.) *Self-representational approaches to consciousness* (pp. 379-408). Cambridge, MA: MIT Press.
- Kidd, C. (ms). The idols of inner-awareness: Towards disjunctive self-representationalism. 5th Online Consciousness Conference.
- Kiefer, A. (2012). Higher-order representation without representation. 104th Annual Meeting of the Southern Society for Philosophy and Psychology in Savannah, GA.
- Kriegel, U. (2006). The same-order monitoring theory of consciousness. In U. Kriegel & K. Williford (Eds.) *Self-representational approaches to consciousness* (pp. 143-170). Cambridge, MA: MIT Press.
- (2007). The phenomenologically manifest. *Phenomenology and the Cognitive Sciences*, 6 (1-2), 115-136. [10.1007/s11097-006-9029-8](https://doi.org/10.1007/s11097-006-9029-8)
- (2009). *Subjective consciousness: A self-representational theory*. Oxford, UK: Oxford University Press.
- (2011). *The sources of intentionality*. Oxford, UK: Oxford University Press.
- (Ed.) (2013). *Phenomenal intentionality*. Oxford, UK: Oxford University Press.
- Kriegel, U. & Williford, K. (Eds.) (2006). *Self-representational approaches to consciousness*. Cambridge, MA: MIT Press.
- Lane, T. & Liang, C. (2011). Self-consciousness and immunity. *Journal of Philosophy*, 108 (2), 78-99.
- Langsam, H. (2011). *The wonder of consciousness*. Cambridge, MA: MIT Press.

- Levine, J. (2001). *Purple haze: The puzzle of consciousness*. Oxford, UK: Oxford University Press.
- Lockwood, M. (1993). The grain problem. In H. Robinson (Ed.) *Objections to physicalism* (pp. 271-292). Oxford, UK: Oxford University Press.
- Lycan, W. (1996). *Consciousness and experience*. Cambridge, MA: MIT Press.
- Lycan, W. (ms). A simple point about an alleged objection to higher-order theories of consciousness. <http://www.unc.edu/~ujanel/A%20Simple%20Point%20about%20an%20Alleged%20Objection%20to.pdf>
- Macphail, E. (1998). *The evolution of consciousness*. Oxford, UK: Oxford University Press.
- Mandik, P. (2009). Beware of the unicorn: Consciousness as being represented and other things that don't exist. *Journal of Consciousness Studies*, 16 (1), 5-36.
- (forthcoming). Conscious-state anti-realism. In C. Munoz-Suarez & F. De Brigard (Eds.) *Content and consciousness revisited*. Berlin, GER: Springer.
- Metzinger, T. (2004). *Being no one: The self-model theory of subjectivity*. Cambridge, MA: MIT Press.
- Millikan, R. (1990). The myth of the essential indexical. *Noûs*, 24 (5), 723-734. [10.2307/2215811](https://doi.org/10.2307/2215811)
- (1995). *White queen psychology and other essays for Alice*. Cambridge, MA: MIT Press.
- Moore, G. (1910). The subject-matter of psychology. *Proceedings of the Aristotelian Society*, 10, 36-62.
- Neander, K. (1998). The division of phenomenal labor: A problem for representational theories of consciousness. *Noûs*, 32 (S12), 411-434. [10.1111/0029-4624.32.s12.18](https://doi.org/10.1111/0029-4624.32.s12.18)
- Northoff, G. (2013). Brain and self - a neurophilosophical account. *Child and Adolescent Psychiatry and Mental Health*, 7 (28), 28-28. [10.1186/1753-2000-7-28](https://doi.org/10.1186/1753-2000-7-28)
- Papineau, D. (2002). *Thinking about consciousness*. Oxford, UK: Oxford University Press.
- Pereboom, D. (2011). *Consciousness and the prospects of physicalism*. Oxford, UK: Oxford University Press.
- Prüfer, T. (1975). An outline of some Husserlian distinctions and strategies, especially in the Crisis. *Phänomenologische Forschungen*, 1, 189-204.
- Revonsuo, A. (2006). *Inner presence: Consciousness as a biological phenomenon*. Cambridge, MA: MIT Press.
- Rosenthal, D. (1986). Two concepts of consciousness. *Philosophical Studies*, 49 (3), 329-359. [10.1007/BF00355521](https://doi.org/10.1007/BF00355521)
- (1997). A theory of consciousness. In N. Block, O. Flanagan & G. Güzeldere (Eds.) *The nature of consciousness: Philosophical debates* (pp. 729-753). Cambridge, MA: MIT Press.
- (2005). *Consciousness and mind*. Oxford, UK: Oxford University Press.
- (2011). Exaggerated reports: Reply to Block. *Analysis*, 71 (3), 431-437. [10.1093/analys/anr039](https://doi.org/10.1093/analys/anr039)
- (2012). Higher-order awareness, misrepresentation and function. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367 (1594), 1424-1438. [10.1098/rstb.2011.0353](https://doi.org/10.1098/rstb.2011.0353)
- Rudrauf, D., Bennequin, D., Landini, G. & Williford, K. (ms). Phenomenal consciousness has the form of a projective 3-space under the action of the general projective linear group.
- Sebastian, M. (forthcoming). Experiential awareness: Do you prefer it to me? *Philosophical Topics*.
- Shoemaker, S. (1968). Self-reference and self-awareness. *Journal of Philosophy*, 65 (19), 555-567. [10.2307/2024121](https://doi.org/10.2307/2024121)
- Smullyan, R. (1984). Chameleonic languages. *Synthese*, 60 (2), 201-224. [10.1007/978-94-017-1592-8\\_11](https://doi.org/10.1007/978-94-017-1592-8_11)
- Stoljar, D. (2006). *Ignorance and imagination: The epistemic origin of the problem of consciousness*. Oxford, UK: Oxford University Press.
- Strawson, G. (2006). Realistic monism: Why physicalism entails panpsychism. In A. Freeman (Ed.) *Consciousness and its place in nature: Does physicalism entail panpsychism?* (pp. 3-31). Charlottesville, VA: Imprint Academic.
- (2009). *Selves: An essay in revisionary metaphysics*. Oxford, UK: Oxford University Press.
- (2011). *The evident connexion: Hume on personal identity*. Oxford, UK: Oxford University Press.
- Thiel, U. (2011). *The early modern subject: Self-consciousness and personal identity from Descartes to Hume*. Oxford, UK: Oxford University Press.
- Tononi, G. (2014). Consciousness: Here, there, but not everywhere. 20th Anniversary Toward a Science of Consciousness April 21-26, 2014 Tucson, Arizona.
- Tononi, G. & Koch, C. (2008). The neural correlates of consciousness. *Annals of the New York Academy of Sciences*, 1124 (1), 239-261. [10.1196/annals.1440.004](https://doi.org/10.1196/annals.1440.004)
- Tye, M. (1995). *Ten problems of consciousness: A representational theory of the phenomenal mind*. Cambridge, MA: MIT Press.
- (2000). *Consciousness, color, and content*. Cambridge, MA: MIT Press.
- (2011). *Consciousness revisited: Materialism without phenomenal concepts*. Cambridge, MA: MIT Press.
- Weisberg, J. (2008). Same old, same old: The same-order representation theory of consciousness and the division of phenomenal labor. *Synthese*, 160 (2), 161-181. [10.1007/s11229-006-9106-0](https://doi.org/10.1007/s11229-006-9106-0)

- (2011a). Misrepresenting consciousness. *Philosophical studies*, 154 (3), 409-433. [10.1007/s11098-010-9567-3](https://doi.org/10.1007/s11098-010-9567-3)
- (2011b). Abusing the notion of what-it's-like-ness: A response to Block. *Analysis*, 71 (3), 438-443. [10.1093/analys/anr040](https://doi.org/10.1093/analys/anr040)
- (2012). On HOTs and HOTIEs: Higher-order thoughts, indexed essentially. 10th Biennial Toward a Science of Consciousness April 9-14, 2012 Tucson, Arizona.
- (2014). *Consciousness*. Cambridge, MA: Polity Press.
- Wilberg, J. (2010). Consciousness and false HOTs. *Philosophical Psychology*, 23 (5), 617-638. [10.1080/09515089.2010.514567](https://doi.org/10.1080/09515089.2010.514567)
- Williams, P. (2000). *The reflexive nature of awareness: A tibetan madhyamaka defence*. Delhi, India: Motilal Banarsidass Publishers.
- Williford, K. (2004). Moore, the diaphanousness of consciousness, and physicalism. *Metaphysica*, 5 (2), 133-153.
- (2005). The intentionality of consciousness and consciousness of intentionality. In G. Forrai & G. Kampis (Eds.) *Intentionality: Past and future* (pp. 143-155). Amsterdam, NL: Rodopi.
- (2006). The self-representational structure of consciousness. In U. Kriegel & K. Williford (Eds.) *Self-representational approaches to consciousness* (pp. 111-142). Cambridge, MA: MIT Press.
- (2007). The logic of phenomenal transparency. *Soochow Journal of Philosophy*, 16, 181-195.
- (2011a). Auto-representacionalismo y los problemas de la subjetividad. *Cuadernos de Epistemología*, 5, 39-51.
- (2011b). Pre-reflective self-consciousness and the autobiographical ego. In J. Webber (Ed.) *Reading Sartre: On phenomenology and existentialism* (pp. 195-210). London, UK: Routledge.
- (2013). Husserl's hyletic data and phenomenal consciousness. *Phenomenology and the Cognitive Sciences*, 12 (3), 501-519. [10.1007/s11097-013-9297-z](https://doi.org/10.1007/s11097-013-9297-z)
- Williford, K., Rudrauf, D. & Landini, G. (2012). The paradoxes of subjectivity and the projective structure of consciousness. In S. Miguens & G. Preyer (Eds.) *Consciousness and subjectivity* (pp. 321-353). Frankfurt, GER: Ontos Verlag.
- Zahavi, D. (1999). *Self-awareness and alterity: A phenomenological investigation*. Evanston, IL: Northwestern University Press.
- (2005). *Subjectivity and selfhood: Investigating the first-person perspective*. Cambridge, MA: MIT Press.
- (2006). Thinking about (self-) consciousness: Phenomenological perspectives. In U. Kriegel & K. Williford (Eds.) *Self-representational approaches to consciousness* (pp. 273-295). Cambridge, MA: MIT Press.

---

# Explaining Subjective Character: Representation, Reflexivity, or Integration?

A Commentary on Kenneth Williford

[Tobias Schlicht](#)

---

While Williford puts forward a self-reflexive account of subjective character, which identifies the subject of experience with episodes (or the stream) of consciousness, an alternative account is defended here that identifies the subject of experience with the whole organism. On this latter approach, a mental representation is conscious if its neural substrate is integrated into the overall neuronal state underlying the conscious state of the organism at that time. This approach avoids an important problem arising for Williford's theory, namely the individuation of episodes. This problem is elaborated in greater detail.

## Keywords

Consciousness | Integration | Phenomenal character | Representationalism | Subject of experience | Subjectivity

## Commentator

[Tobias Schlicht](#)  
tobias.schlicht@rub.de  
Ruhr-Universität Bochum  
Bochum, Germany

## Target Author

[Kenneth Williford](#)  
williford@uta.edu  
The University of Texas  
Arlington, TX, U.S.A.

## Editors

[Thomas Metzinger](#)  
metzinger@uni-mainz.de  
Johannes Gutenberg-Universität  
Mainz, Germany

[Jennifer M. Windt](#)  
jennifer.windt@monash.edu  
Monash University  
Melbourne, Australia

## 1 Introduction

The starting point for this commentary on Williford's article is the *commitment* to subjective character as a defining feature of consciousness. Subjective character is what *makes* a conscious experience conscious, i.e., what *all* conscious experiences have in common in virtue of which we call them conscious. [Kriegel \(2009, p. 1\)](#) has offered a distinction between the *qualitative character* and the *subjective character* as two

important aspects of any conscious experience. If you have a phenomenally conscious sensation of red, then there is something that it is like for you to have it. On the one hand, having this experience feels like *this* (where *this* quality distinguishes it from feeling a sensation of pain, say). On the other hand, it feels like something *for you* (i.e., it is subjective *in the same sense* that all of your other conscious experiences are sub-

jective). Qualitative character is the *distinguishing* mark of conscious experiences with regard to each other; subjective character is the *common mark* of all my conscious experiences. This latter aspect has also often been referred to as the me-ishness, ipseity or mine-ness of consciousness (Block 1995; Zahavi 1999).

Williford recognizes that, sadly, not all philosophers theorizing about consciousness share this commitment to subjective character, and that some formulations of it in terms of mine-ness are misleading in giving rise to objectionable implications about essential entities (Metzinger 2011). But for the purposes of this commentary we can leave aside such controversies and instead start with a shared commitment to this feature, on which this commentary will exclusively focus.<sup>1</sup> As a constraint on a theory of subjective character, Williford maintains that it has to respect (1) the relational structure of consciousness, and (2) the Humean intuition that one of the relata, the subject, remains somewhat invisible and is at least not constituted by a special (additional) entity. His solution, in short, is to peacefully combine these two intuitions by *identifying* the subject with (an episode of or) the stream of consciousness, which is itself reflexively self-aware. A further claim is that this account is supposedly “compatible with physicalism” (Williford this collection, p. 1). I do not address this aspect of Williford’s rich paper in this commentary, mostly for reasons of space but also because I think that the putative truth of physicalism should not put any a priori constraints on a theory of consciousness.

<sup>1</sup> One might argue that this move is already problematic since it looks like a *petitio principii*. But I simply take it as an analysis of Nagel’s phrase that *there is something it is like for the organism* to experience something red, say. As a characterization of phenomenal consciousness this is almost unanimously accepted in the field. What it picks out according to the present analysis is a variant aspect that differs in different experiences (qualitative character), and an invariant aspect that remains identical across different experiences (subjective character). I do not have enough space here to argue in detail for this analysis. A further reason for distinguishing both aspects, subjective and qualitative character, is the phenomenal observation that we can become *conscious of ourselves as the identical subject* in contrast to the constantly changing stream (or ensemble) of conscious representations. Here, the qualitative differences of the multiple representations we have at a time do not matter. What matters here is that they are related to myself such that I can call them and experience them as mine (cf. Schlicht 2011).

My commentary is thus structured as follows. In the [second](#) section, I will recapitulate Williford’s take on subjective character and point to problems with his identification of the subject with the stream (or episodes) of consciousness. In the [third](#) and [fourth](#) sections, I will present an alternative way of conceptualizing the subject in the context of a theory of consciousness that also satisfies the constraints mentioned above. On this alternative view, a mental representation is conscious (i.e., it exhibits its subjective character) if it is integrated in the right way into the overall conscious state of the organism. This overall state includes representations of the state of the organism. By way of integration, all *conscious* representations are something *for* the organism that is identified as the subject of experience. This alternative, which is an instance of an integration-theory, has the advantage both of bypassing the problems that seem to beset Williford’s account and of being not only compatible with but also supported by the best empirical hypotheses about consciousness currently available. I will sketch an argument for this view and attempt to answer possible objections to the premises of this argument.

## 2 Williford on subjective character

Williford’s aim is to characterize the subjective character of consciousness in a way that accounts “for both the Humean intuition that the subject-relatum is, in some sense, invisible and that, nevertheless, consciousness has a subject-object relational structure that is phenomenally manifest and non-inferentially knowable” (this collection, pp. 10-11). There are three constraints on an account of subjective character, according to Williford: (a) conscious experiences are relational in having both a subject- and an object-pole; (b) the subject-pole is not constituted by some additional, irreducible, or otherwise special entity; (c) the subject-pole must be something that is nevertheless manifest in consciousness, not hidden from it.

Note that it is not an option for someone taking subjective character seriously to agree that phenomenal consciousness is relational, in-

volving a subject-pole, but at the same time holding that this pole is forever “*hidden*” (p. 11). This does not work because such an account could not explain subjective character. After all, we consider subjective character as real *only* because it supposedly shows up phenomenally: I experience my conscious states as *mine*.<sup>2</sup>

Williford’s commitment to a subject (subject-pole) rules out the possibility that experiences may be free-floating entities, not being enjoyed by *anyone*. Phenomenal consciousness is supposed to be relational through-and-through, directed *at* some object and existing *for* some subject: “anything that phenomenally appears, appears to someone or something” (p. 9).<sup>3</sup> In general, Williford attempts to capture both the intentionality and the subjectivity of consciousness in the slogan that every experience involves the “appearance of something to something” (p. 9), where the latter refers to subjectivity. He leaves the notion of “*appearing* or of *phenomenally manifesting* undefined” (p. 10), but in order for what he says to make sense we have to take it to be just another way of saying that something is *phenomenally conscious*: it is “just the appearance to/in consciousness of something” (p. 10).

In order to meet the constraints he set for himself, Williford identifies the subject with the stream of consciousness or with (some complex or rich) episode of consciousness (p. 10). This identity claim then leads to the situation that the subject-pole of the consciousness-relation *appearing* (or being manifest) in the conscious episode is the episode itself. The subject-pole is thereby manifest, i.e., consciously experienced, but not separable as an entity from the conscious episode in question, and thus it is—in a sense—invisible. But it is only invisible in the sense that there is no additional entity that ac-

counts for the subject-pole. In order to meet the constraints mentioned above, Williford therefore defends “the view that consciousness is self-manifesting” (p. 10), i.e., an episode or stream of consciousness *appears to itself* no matter what else is manifest to consciousness (some perceived object, say).

Partly because Williford subscribes to the Humean intuition that we do not find a self, or a “self-entity, me-haecceity, me-ish quale, or subject-relatum” (p. 10) if we turn to our stream of conscious experiences, he is led to the identification of the subject-relatum with the stream of consciousness itself. Although the conscious episode appears itself *in* the episode, consciousness is self-reflexive, yet not self-representing. The relevant difference between an unconscious and a conscious episode is not due to some form of representation. Rather, the conscious episode contains an internal relational (intrinsic) property that is responsible for the episode’s being *acquainted with itself*.<sup>4</sup> Subjective character is thus supposedly “the self-acquaintance of every instance of consciousness” (p. 1), which these instances exhibit in virtue of “some internal relational property” (p. 1). The subject of experience, being *identical* to the episode of consciousness, is self-acquainted. But although consciousness is self-reflexive, the claim is not that a mental episode becomes conscious through an act of reflection directed at it (p. 10). This is an impossible path when it comes to explaining subjective character, since an act of reflection presupposes that what it reflects upon is already *mine* in the relevant sense to be explained (Frank 2007; Zahavi 1999). Reflection can discover but not bring into being a self-referential conscious state.

Now, the stream (or episode) of consciousness exhibits subjective character in the sense that the stream itself is manifest within the stream so that the relationality constraint is met, although no additional entity need be introduced in order to play the subject-role. Therefore, the Humean invisibility-constraint is met as well. This is more or less the positive

<sup>2</sup> One way to put this with respect to sensations like hunger is to say that, since they are related to me in such an unmediated sense, it is impossible to be mistaken about the subject undergoing such sensations (Shoemaker 1968).

<sup>3</sup> This claim is defended especially in opposition to what Williford calls F-theories, or varieties of first-order representationalism such as Tye’s (1995) PANIC-theory, which arguably neither accepts nor explains subjectivity so understood. Higher-order and same-order accounts at least accept this feature of consciousness, which they—mistakenly—attempt to explain in terms of representation.

<sup>4</sup> Thus, the property of being conscious (and thus subjective) is not bestowed upon the episode by some external property, like a higher-order thought directed at (or representing) it (Rosenthal 2005).

story as far as I have understood it. The main philosophical problem for Williford's account is to formulate criteria as to how to individuate an episode. This problem leads to a dilemma for his account that is spelled out in more detail below.

If we follow Williford and identify the subject with complex conscious episodes (or even *the whole* stream of consciousness), then subjective character only seems to arise for complex episodes, and not for any of the episode's parts or elements: "[t]he episode is a unified whole, the differentiated qualities and objects appearing in/to it are like its parts [...]" (pp. 10-11). Since he emphasizes that all episodes have parts (*ibid.*), I take it that a single sensation of red, say, consequently does *not* count as an episode, because it can hardly be separated into parts; then it can instead always appear only as an element of an episode which is in turn a "unified whole". On the other hand, Williford also emphasizes that, trivially, everything always also is an improper part of itself. On this reading, a single sensation of red *could* be an episode. This gives rise to the following options regarding the individuation of episodes that can be put in terms of a dilemma:

1. *If a single sensation of red is too simple to count as an episode*, then all that Williford's theory can explain is why the complex episode as an emergent whole (having single experiences as its parts or elements) is conscious. It *cannot* explain what makes an individual element of this whole episode (or stream), a sensation of red say, conscious. But the varieties of representationalism (which he criticizes) aim to explain exactly this feature of consciousness. A problem with this first horn of the dilemma is thus that we need to answer the question whether or not such single sensations can be conscious independently of being an element of a larger episode.
  - a) *If individual sensations can be conscious independently*, then the question arises as to whether they can be conscious *without* thereby exhibiting subjective character (given subjective character only arises on

the level of whole episodes). This is not what Williford should accept since he takes subjective character to be a defining feature of consciousness; there is no consciousness without subjective character. So if an individual sensation of red could be conscious then it could be so only by exhibiting subjective character. This seems to lead us to Zeki's theory of "micro-consciousness" (Zeki & Bartels 1998; Zeki 2007) according to which every individual node of a perceptual system (visual, auditory etc.) can generate an "atom" of consciousness independently. This is an extreme version of what Bayne (2010) calls an "atomistic" approach to consciousness, standing in contrast to more "holistic" approaches:

"Theorists that adopt an atomistic orientation assume that the phenomenal field is composed of 'atoms of consciousness'—states that are independently conscious. Holists, by contrast, hold that the components of the phenomenal field are conscious only as the components of that field. Holists deny that there are any independent conscious states that need to be bound together to form a phenomenal field. Holists can allow that the phenomenal field can be formally decomposed into discrete experiences, but they will deny that these elements are independent atoms or units of consciousness." (Bayne 2010, pp. 225-226)

The problem with such atomistic approaches is really the phenomenon of the unity of consciousness, i.e., that such individually conscious units would need to be bound together to a much larger all-encompassing unified "phenomenal field", as Bayne puts it, in order to account for what we actually experience. But then we should expect there to be a *mechanism* responsible for such phenomenal binding, a mechanism that we also should expect to break down occasionally under certain circumstances; but there is no evidence for

such a mechanism. The phenomenal unity of consciousness seems to be a deep feature of consciousness just like subjective character, in the sense that it cannot break down and that phenomenal consciousness cannot occur without it. I agree with Bayne's point here (cf. Schlicht 2007), and I think that Williford would not be prepared to take Zeki's route either. At least there is no indication in the text that would support this reading. Alas, Williford also sets aside the important issue of the unity of consciousness, which arises given the unresolved problem of providing criteria for the individuation of episodes.

- b) So we are left with the alternative that *individual sensations cannot be conscious independently*. For an individual element to become conscious (and to exhibit subjective character) it must then be *integrated* into a larger (cumulative) episode. What's needed then is a theory (and a mechanism) explaining how such integration into an episode takes place. However, then we are left with an alternative view regarding the question of what is responsible for a representation's being conscious, namely some kind of integration-theory. In fact, that is the path I will recommend (and elaborate in more detail) below in section 3. The general idea is that phenomenally-conscious representations are those that are adequately integrated into a global state (we may call it an episode). My worry with regard to Williford's account is simply that once we have such an integration-account, there is no need for his additional story in terms of self-reflexivity in order to explain subjective character. Since subjective character is (taken to be) a defining feature of conscious experience, an account that informs us about how individual sensations become conscious will also inform us about how they acquire subjective character: through integration.
2. But that's not the end of the story. Williford simply could say that a sensation of red may

be a conscious episode. So far, we have discussed the problem of individuating episodes on the assumption that a single sensation of red cannot count as an episode. Now we have to discuss the consequences of the assumption that *a single sensation of red may count as an episode*. This leads to two further possibilities.

- a) One could accept such minimal episodes despite the fact that this concession gives rise to a multiplicity of (streams and consequently) conscious subjects. Although it's metaphysically (somewhat) extravagant, this is a perfectly coherent position to take. Indeed, it seems to be akin to Strawson's theory of the self, according to which a self lasts only as long as an individual state (or episode) of consciousness (Strawson 1997). But this view flies in the face of experience. For one thing, it is inadequate to explain an important aspect of consciousness, namely what we may call, following Kant, the *(empirical) consciousness of the identity of oneself as subject*: "I am [...] conscious of the identical self in regard to the manifold of the representations that are given to me in an intuition because I call them all together *my* representations, which constitute *one*" (B134). What he means is that, at least in non-pathological cases, I can become conscious of myself *as* the single, (synchronically as well as diachronically) identical subject vis-à-vis my diverse experiences. I *never* identify myself with one or many of my conscious representations (or episodes for that matter). Rather, I distinguish myself from them *as the subject* who has them when I self-ascribe them. And this empirical consciousness of an identical subject is possible, according to Kant, *because* all my *conscious* experiences are already self-related. I can already call them *mine* because they exhibit subjective character simply by being phenomenally conscious. Kant, famously and notoriously, tried to account for this consciousness of self by simply postulating a *transcendental unity* of apperception in which this is sup-



posed to originate. If Strawson’s view were correct, then Kant would presumably reply by pointing to a natural, yet implausible consequence: “I would have as multicoloured, diverse a self as I have representations of which I am conscious” (CpR B134). Accepting this horn of the dilemma therefore has the consequence that we would now need a story that helps us make sense of how the subject of the sensation of red is related to the subject that is identified with an auditory sensation of a loud sound, etc. In effect, this would lead to a binding problem for the multitudinous “subjects” of experience, since in my view, we cannot be content with a multiplicity of conscious subjects. I also think that Williford might not be satisfied with such an outcome, since he never entertains the possibility of multiple subjects in his essay.

- b) Therefore—again, on the hypothesis that a single sensation of red counts as an episode—one could argue that *the multiplicity of conscious episodes has to be overcome in favor of one (unified) stream of consciousness*. This calls, again, for an integration mechanism that produces such a unity. Though I can understand why one would now identify this resulting integrated single stream of consciousness with the subject of experience, I don’t see any motivation to identify the episode “single sensation of red” with a subject of experience, if a more complex combination of episodes is needed anyway.

I conclude that the problem of individuating episodes either leads to the acceptance of implausible views like Zeki or Strawson’s theories of consciousness and self or to the need for an integration account that explains how individual elements are combined into the one global conscious experience. The claim I would like to put forward is that once we have such an integration account, Williford’s proposal becomes superfluous, because what it is intended to explain is then already explained by the integration account.

### 3 Integration vs. representation

When the aim is to provide an account of the difference between a representation’s being phenomenally conscious and it’s being unconscious many philosophers are drawn to some form of *representationalism*. This is motivated in part by the prospect of reducing the problem of consciousness to the problem of intentionality or *representation* (Tye 1995; Dretske 1995; Rosenthal 2005; Lycan 1996; Metzinger 2003; Kriegel 2009; Kriegel & Williford 2006). But many of those who are dissatisfied with a representational criterion argue that the difference is due to some sort of *integration* (Dehaene 2014; Van Gulick 2004; Edelman & Tononi 2000; Damasio 2010; Metzinger 1995; Kant 1999; Schlicht 2011). Such integration may eventually result in a higher-order or more complex representational state. In that sense, the two accounts do not mutually exclude each other. But they give different answers to the question of what is responsible for the representation exhibiting the feature of being conscious. To put forward both a representational condition and an integration mechanism would amount to wearing a belt as well as suspenders. Williford’s paper demonstrates that other theories are also possible. He favors self-reflexivity as the core feature a representation must exhibit in order for it to be conscious.

In the first part of his paper, Williford scrutinizes all dominant varieties of representationalism, especially with respect to their explanatory power regarding the subjective character of conscious experiences. His case against first-order, higher-order, and same-order or self-representationalism is solid, and I have nothing to add in this regard (cf. also Schlicht 2008b; Vosgerau et al. 2008).<sup>5</sup>

The basis for answering the question as to which conditions have to be met by a single sensation of red in order for it to be conscious and subjectively experienced is the observation

<sup>5</sup> I disagree with respect to what Williford calls P-Theories, according to which a “privileged object” is represented which makes all the difference between conscious and unconscious representations. Williford interprets Damasio’s theory in this way, but although various representations (of the body especially) play an important role in Damasio’s theory (as in most other theories), this is not the whole story (see fn. 7).

that the organism in question is already conscious in the creature-sense. This general consciousness (or state of vigilance) admits of degrees (from deep coma to wakefulness) and is one of the conditions for being able to enjoy a sensation of red at all (Dehaene et al. 2006). Empirical evidence points to the assumption that the neural structures in the brain supporting this state contain the relevant structures monitoring and regulating the homeostatic balance of the whole organism. Damasio (1999, 2010) calls these structures “proto-self”-structures, the biological forerunner of that which we eventually experience as a sense of self. He assumes that the brain can only perform these functions of monitoring and regulating if the overall state of the whole organism is represented in the brain.

In addition to representations of the organism, the brain is assumed to produce representations of (objects in) the external world. Given the limited capacity of conscious perception and memory systems, such representations stand in competition (Koch 2004). The basic idea of integration-theories is that some of these competing representations, like a sensation of red, are conscious because they are integrated into a more global state that also contains the structures responsible for creature-consciousness. Van Gulick (2004) has sketched such an integration-theory, based on ideas already to be found in Metzinger (1995):

The basic idea is that lower-order object states become conscious by being incorporated as components into the higher-order global states (HOGS) that are the neural and functional substrates of conscious self-awareness. The transformation from unconscious to conscious state is not a matter of merely directing a separate and distinct meta-state onto the lower-order state but of “recruiting” it into the globally integrated state that is the momentary realization of the agent’s shifting transient conscious awareness. (Van Gulick 2004, pp. 76-77)

In other words, a single sensation of red is consciously experienced if the neural activation

pattern supporting this sensation is integrated in the right way into the neural basis representing the overall state of the organism, the “dynamic core” in Edelman’s words (Edelman & Tononi 2000).<sup>6</sup>

Importantly, the integration mechanism (which is what has to be determined empirically in this framework)—synchronous oscillations, say—is not only responsible for producing a coherent single experiential state of the organism; it also thereby conveys subjective character to the integrated individual representations. If this idea is combined with Damasio’s (1999) notion of proto-self-structures, then integration facilitates a strong connection between the substrate of an individual sensation (of red, say) and the biological structure representing the organism in the brain.<sup>7</sup> Of course, just like on all other theories, the hard problem is not addressed head-on, i.e., it is not explained *why* activation of these structures feels like something at all. All that can be provided (at this stage anyway) is a coherent story of how all these aspects hang together. But one advantage of the present integration-account is that by establishing a connection between the organism (as represented in the brain) and its object-representations we can make sense of the important fact that all conscious representations feel like something *for the organism*. The organism provides, as Damasio puts it, a “haven of stability and invariance” (1999, p. 142, p. 153; see also Metzinger 2003, p. 161), i.e., just what we need in order to account for subjective character. For remember that subjective

6 Another way to think of this is along the lines of the “Global Neuronal Workspace Model” in which attentional mechanisms determine which of the neural coalitions are integrated (Dehaene et al. 2006). But means other than attention are possible.

7 Williford discusses Damasio’s theory under the label of a P-Theory as a variety of representationalism and finds it wanting. Of course, representations of various sorts, especially of the organism, play an important role in Damasio’s theory (as in many other theories). But I do not share Williford’s interpretation that it is these (special) representations as such that are responsible for consciousness. Various representations (or maps, as Damasio also calls them) have to be integrated in the right kind of way in order for there to be something it is like for the organism. Therefore, I do not consider Damasio’s theory a version of representationalism since there, the mechanism responsible for consciousness is not representation but integration of body representations with object representations via recurrent activations in so-called “convergence zones” (Damasio 1994, p. 95-96, 162).

character is the feature that remains stable across different representations, while qualitative character is the feature that distinguishes different representations from each other. So in order to get an account of subjective character started, we have to look for the point of “*maximal invariance of content* in the conscious model of reality”, as Metzinger (2003, p. 134) puts it. Metzinger agrees that this invariance is most likely due to the organism and its bodily structures represented in the brain, since it is invariance (or maintenance of homeostatic balance) that keeps the organism alive. Another advantage of this view is that it does so without introducing a questionable new entity and by avoiding Williford’s phenomenologically counterintuitive claim that the stream of consciousness should be identified with the subject of experience. In this commentary, I cannot argue in detail for this positive alternative but hope that these sketchy comments give the reader a general idea of what it aims at. Since I am dissatisfied with Williford’s identification of the subject of experience with the stream or an episode of consciousness, let me now finally turn to an argument for a different conceptualization of the subject.

## 4 The subject as organism

My alternative claim is that we should simply identify the subject with the organism. This section is an attempt to support this bold claim. The premises of the argument focus on analyses of the structures of phenomenal consciousness and intentionality:

*Premise 1 (phenomenal consciousness):*

Phenomenal consciousness is characterized by there being something that it is like *for a subject* to be in that state. In this minimal sense, consciousness is relational and requires the assumption of a *subject-pole* of experience.

*Premise 2 (intentionality):*

The structure of intentionality is such that a *subject* is directed (via some psychological act or attitude like believing, desiring, perceiving etc.) at a content, object, or state of affairs. Inten-

tionality is quasi-relational since at least the *subject* must exist, although the intentional object need not exist.<sup>8</sup>

*Premise 3 (subject identity):*

The subject that is intentionally directed is *identical* to the subject for whom there is something that it is like to be in a given mental state.

*Premise 4 (embodied cognition):*

Many intentional attitudes (like perceiving, grasping, emoting) are *embodied* and can be ascribed only to an embodied agent, i.e., to the whole organism.

*Conclusion:*

The *subject* for which there is something it is like to be in a given mental state and the *subject* that is intentionally directed at a content or object is the *organism*.

### 4.1 Elaboration of the premises

*Premise 1: Phenomenal consciousness*

First of all, it is interesting to note that Nagel’s initial characterization of consciousness in terms of *there being something that it is like* is already concerned with *the organism* as the entity *for which* there is something that it is like: After having noted the diversity of beings capable of conscious experience which may lead to very different kinds of conscious experience, Nagel argues that “no matter how the form may vary, the fact that an organism has conscious experience *at all* means, basically, that there is something it is like to *be* that organism, [...] something it is like *for* the organism” (1974, p. 436). So, given that the philosophical community seems to have agreed to refer to Nagel’s slogan in order to characterize phenomenal consciousness in the first place, they should seriously consider Nagel’s talk of the organism as the subject of experience. But apart from this ob-

<sup>8</sup> A reviewer pointed out that different thinkers had different opinions about *what* is intentionally directed: the mental state, the psychological act, or the thinker etc. As will become clear below, I do not share the view that a mental state is itself directed, but favor the view that a creature of some sort is directed at something *via* an act or attitude. A great advantage of this view is that such attitudes are not limited to mental states like beliefs and desires (as traditionally held), but it also allows also for motor intentional attitudes like grasping or holding etc., i.e., essentially bodily ways of being directed (premise 4). For details see Schlicht (2008a).

servation, all that is stressed in the first premise is the relational character of phenomenal consciousness, much in the sense of one of the commitments defended in Williford’s paper. The reasons for holding this are mainly phenomenological: it simply appears that way. And we are all aiming at a theory of why this is so. Williford’s elaboration of the relational structure of consciousness in terms of the genitive and dative of manifestation captures the intuition expressed in this premise very well. Thus, there is not much room for disagreement here.<sup>9</sup>

### *Premise 2: Intentionality*

In his canonical elaboration of the structure of intentionality, Subject—Intentional Mode—Content, Tim Crane (2001, p. 31) admits that he does not provide an account of the first relatum, “because the nature of the subject is not something that is within the scope of this book (strange as that may seem)”. Yet, as far as intentional states are concerned, the assumption that attitudes are not free-floating entities but come along with a thinker, perceiver, or believer is rather uncontroversial. What’s controversial is how we should characterize the subject and what kind of commitment is implied in the “acceptance” of a thinker, perceiver, or believer.<sup>10</sup>

### *Premise 3: Subject identity*

In a way, this premise is at the same time trivial and important. First of all, if one accepts premises 1 and 2, then it is natural to accept premise 3, if only because the alternative would lead to a multiplicity of subjects, giving rise to questions regarding the relations between them. I discussed this option above in section 2. There are many debates about the relation between consciousness and intentionality, but there is hardly any debate about the relation between the subjects of each. So in a way, this premise simply states the obvious, given premises 1 and 2. But it is plausible to accept it even independently of these premises as the default position. One important reason for this is that there are many conscious experiences that are both phenomenal and intentional—perceptual experi-

ences, for example. If I am looking at a red tomato, then my conscious experience presents me with an object in the external world at which I am thereby visually directed. But there is also something that it is like for me to see the tomato if I am phenomenally conscious of it. Since it would be odd to claim that there are two subjects involved here—one being intentionally directed and one being conscious of the tomato—the default position is that it is one the same subject that is intentionally directed and phenomenally conscious. Second, despite the discussion among analytic philosophers in the last fifty years, it is not clear that phenomenal consciousness and intentionality can be separated from each other so easily anyway. In fact, proponents of phenomenal intentionality (or cognitive phenomenology, see Bayne & Montague 2011) like Searle (1992), Strawson (2004), Pitt (2004), Horgan & Tienson (2002), Kriegel (2013) and others argue to the contrary. Again, then the premise simply states the obvious.

But this premise also is important because once we commit to it, we can follow either premise 1 or 2 in our investigation to see whether we can formulate constraints on the nature of the subject based on either consciousness or intentionality. This is the job of premise 4, which accepts lessons from recent investigations into ways of being intentionally directed.

### *Premise 4: Embodied Cognition*

Cognitive Science has recently been dominated by discussions on the so-called 4Es, i.e., embodied, embedded, enactive, and extended cognition. These notions characterize four important ways in which our current theorizing about cognition departs from classical cognitive science. They are more or less independent of each other and can be accepted and rejected in isolation.<sup>11</sup> This is not the place to elaborate in detail all four of them, especially because for the purposes of this argument only the feature of embodiment is important. Many of our psychological acts, like *perceiving*, being *emotionally* directed at or affected by something or other, performing *intentional actions*, etc., are embodied

<sup>9</sup> I support this premise in more detail in Schlicht (forthcoming). For the purposes of this commentary it is sufficient to note the agreement on the intuition that phenomenal consciousness is relational.

<sup>10</sup> Again, I argue for this premise in Schlicht (forthcoming).

<sup>11</sup> An exception may be the intricate connection between cognition being embodied and (therefore) being embedded.

in the sense that features of an organism's non-neural body contribute importantly—be it causally or even constitutively—to the execution of these cognitive acts (Wilson & Foglia 2011).

A plausible claim defended by enactivists is that even a basic cognitive act like perceiving involves many bodily movements like eye-, head- and whole-body movements when looking at or focusing on an object, or when jointly attending to an object with someone else (Noë 2004). This can be accepted independently of more radical claims regarding the usefulness of representations typically put forward by enactivists (Hutto & Myin 2013). What's more, a bulk of empirical evidence has accumulated that supports the important role of the body and bodily actions for psychological acts:

- a) Facial expressions and bodily postures are arguably constitutive elements of feelings and their expression. Many theories of emotion such as multifactorial models (e.g., Scherer 2009; Welpinghus 2013) usually include as one component a bodily feature. Moreover, eye- and head-movements count among the constitutive and content-determining elements of visual perception (Noë 2004).
- b) Research on mirror neurons has demonstrated the intricate relation between perceiving and acting in the sense that the same neural structures are employed for the execution and observation or recognition of intentional acts and emotional expressions (Rizzolatti & Sinigaglia 2008; Keysers 2013). Controversial debates about the role of mirror neurons for social cognition notwithstanding, it is fair to say that from a neural perspective, perception and action have to be considered as constituting one single complex system. We develop motor programs for the performance of certain actions and reuse these programs in our observation of others when they perform such actions. These motor programs contain goal-directed representations with a bodily format (Goldman & de Vignemont 2009) that are crucially different from the propositional format of a belief, say.
- c) What's more, lessons from studies of pathological conditions like visual form agnosia (Mil-

ner & Goodale 1995) suggest that we can be directed at an object in a purely motor-intentional way, thereby demonstrating a “bodily understanding” (Kelly 2002) of an object that is not based on concepts and cannot be put into appropriate words.

Generalizing these (and many other) points (see e.g., Gallagher 2005) leads to a paradigm shift with regard to our understanding of the subject of intentionality: intentionality is not restricted to propositional attitudes; an embodied agent, i.e., an organism,<sup>12</sup> has many sensorimotor, affective, and cognitive means to be directed at objects and states of affairs. This way of understanding the structure of intentionality allows us to capture many more phenomena that clearly fall under the name of intentionality as directedness, e.g., reaching for and grasping an object.

All the premises taken together yield the conclusion that there is one subject capable of intentionality and consciousness that can be identified with the organism (not with the stream of consciousness), characterized by a variety of cognitive capacities allowing for a range of intentional attitudes—some of which are affective,<sup>13</sup> others sensorimotor,<sup>14</sup> and still others are of sophisticated cognitive<sup>15</sup> varieties. The overall state of the subject, being the whole organism, is represented in the brain. This representation contains information about its body, its interior milieu, etc., such that all representations having to do with the organism's interaction with objects can be coupled to or integrated with the representations monitoring and regulating the state of the organism in the

<sup>12</sup> Talk about embodied agents is broader than talk about organisms. The biological constraints on full-blown cognitive and conscious agents are currently unknown. Whether artificial cognitive systems are possible depends on the limits set by such constraints. In this paper, I cannot address this point.

<sup>13</sup> One of Brentano's examples in his famous passage on intentionality being the mark of the mental is love, in which someone is loved. This example cannot be adequately captured by restricting intentionality to propositional attitudes which can be formulated using “that-clauses”.

<sup>14</sup> Many forms of being intentionally directed are sensorimotor, e.g., all that has to do with perception and action, this being the biologically primary form of intentionality (Searle 1983, p. 36).

<sup>15</sup> Most cognitive varieties of intentionality are sophisticated and propositional, like beliefs and desires, which can be put into sentences containing “that-clauses”, e.g., Ken believes that physicalism is true.

brain. On the basis of such couplings, it is in principle possible to make sense of the idea that object-representations become subjective in the sense of being something *for* the organism.

A caveat: this does not amount to an explanation of how consciousness arises in the first place, or of *why* integrated representations are experienced at all. But hardly any theory of consciousness has properly addressed this hard problem so far (Chalmers 1996). The limited claim of this commentary is that we can at best make sense of the subjective character of phenomenal consciousness if we adopt an integration-theory as outlined above and regard the subject for which there is *something that it is like* as the whole organism. As I concluded in the first part, depending on how he is going to individuate episodes—a problem which he has not yet solved—, Williford seems to be in need of such an integration-account anyway. Therefore, this sketch of an alternative should be appealing for someone taking subjective character seriously.<sup>16</sup>

## Acknowledgements

I would like to thank Jennifer Windt, Thomas Metzinger, and two anonymous reviewers for very helpful comments on an earlier draft of this paper. I am also grateful to Kenneth Williford, who provided such a stimulating and challenging paper and once again to Jennifer Windt and Thomas Metzinger for the chance to contribute this commentary.

This work is funded by the Volkswagen Foundation.

## References

- Bayne, T. (2010). *The unity of consciousness*. Oxford, UK: Oxford University Press.
- Bayne, T. & Montague, M. (2011). *Cognitive phenomenology*. Oxford, UK: Oxford University Press.
- Block, N. (1995). On a confusion about a function of consciousness. *Behavioral and Brain Sciences*, 18 (2), 227-287. [10.1017/S0140525X00038188](https://doi.org/10.1017/S0140525X00038188)
- Chalmers, D. J. (1996). *The conscious mind*. Oxford, UK: Oxford University Press.
- Crane, T. (2001). *Elements of mind*. Oxford, UK: Oxford University Press.
- Damasio, A. (1994). *Descartes' error*. New York, NY: Avon Books.
- (1999). *The feeling of what happens*. New York, NY: Basic Books.
- (2010). *Self comes to mind*. New York, NY: Basic Books.
- Dehaene, S. (2014). *Consciousness and the brain. Deciphering how the brain codes our thoughts*. New York, NY: Viking.
- Dehaene, S., Changeux, J.-P., Naccache, L., Sackur, J. & Sergent, C. (2006). Conscious, preconscious, and subliminal processing: A testable taxonomy. *Trends in Cognitive Sciences*, 10 (5), 1-8. [10.1016/j.tics.2006.03.007](https://doi.org/10.1016/j.tics.2006.03.007)
- Dretske, F. (1995). *Naturalizing the mind*. Cambridge, MA: MIT Press.
- Edelman, G. M. & Tononi, G. (2000). *A universe of consciousness: How matter becomes imagination*. New York, NY: Basic Books.
- Frank, M. (2007). Non-objectal subjectivity. *Journal of Consciousness Studies*, 14 (5-6), 152-173. [10.1007/s11098-011-9837-8](https://doi.org/10.1007/s11098-011-9837-8)
- Gallagher, S. (2005). *How the body shapes the mind*. Oxford, UK: Oxford University Press.
- Goldman, A. & de Vignemont, F. (2009). Is social cognition embodied? *Trends in Cognitive Sciences*, 13 (4), 154-159. [10.1016/j.tics.2009.01.007](https://doi.org/10.1016/j.tics.2009.01.007)
- Horgan, T. & Tienson, J. (2002). The intentionality of phenomenology and the phenomenology of intentionality. In D. J. Chalmers (Ed.) *Philosophy of mind. Classical and contemporary readings* (pp. 520-533). Oxford, UK: Oxford University Press.
- Hutto, D. D. & Myin, E. (2013). *Radicalizing enactivism. Basic minds without content*. Cambridge, MA: MIT Press.
- Kant, I. (1999). Critique of pure reason. In A. Wood & P. Guyer (Eds.) Cambridge, UK: Cambridge University Press.

<sup>16</sup> The limits of this commentary do not permit an exhaustive discussion of possible objections to this account, but I discuss it at greater length in Schlicht (forthcoming).

- Kelly, S. (2002). Merleau-Ponty on the body: The logic of motor intentionality. *Ratio*, 15 (4), 376-391.  
[10.1111/1467-9329.00198](https://doi.org/10.1111/1467-9329.00198)
- Keysers, C. (2013). *The empathic brain*. Groningen, NL: Social Brain Press.
- Koch, C. (2004). *The quest for consciousness*. Englewood, CL: Roberts & Company.
- Kriegel, U. (2009). *Subjective consciousness*. Oxford, UK: Oxford University Press.
- (2013). *Phenomenal intentionality*. Oxford, UK: Oxford University Press.
- Kriegel, U. & Williford, K. (Eds.) (2006). *Self-representational approaches to consciousness*. Cambridge, MA: MIT Press.
- Lycan, W. (1996). *Consciousness and experience*. Cambridge, MA: MIT Press.
- Metzinger, T. (1995). Faster than thought. Holism, homogeneity and temporal coding. In T. Metzinger (Ed.) *Conscious experience* (pp. 425-461). Thorverton, UK: Imprint Academic.
- (2003). *Being no one*. Cambridge, MA: MIT Press.
- (2011). The no-self alternative. In S. Gallagher (Ed.) *The oxford handbook of the self* (pp. 279-296). Oxford, UK: Oxford University Press.
- Milner, A. D. & Goodale, M. A. (1995). *The visual brain in action*. Oxford, UK: Oxford University Press.
- Nagel, T. (1974). What is it like to be a bat? *Philosophical Review*, 83 (4), 435-450.
- Noë, A. (2004). *Action in perception*. Cambridge, MA: MIT Press.
- Pitt, D. (2004). The phenomenology of cognition or what is it like to think that p? *Philosophy and Phenomenological Research*, 69 (1), 1-36.  
[10.1111/j.1933-1592.2004.tb00382.x](https://doi.org/10.1111/j.1933-1592.2004.tb00382.x)
- Rizzolatti, G. & Sinigaglia, C. (2008). *Mirrors in the brain*. Oxford, UK: Oxford University Press.
- Rosenthal, D. (2005). *Consciousness and mind*. Oxford, UK: Oxford University Press.
- Scherer, Klaus R. (2009). The dynamic architecture of emotion: Evidence for the component process model. *Cognition & Emotion*, 23 (7), 1307-1351.  
[10.1080/02699930902928969](https://doi.org/10.1080/02699930902928969)
- Schlicht, T. (2007). *Erkenntnistheoretischer Dualismus. Das Problem der Erklärungslücke in Geist-Gehirn-Theorien*. Paderborn, GER: Mentis.
- (2008a). Ein Stufenmodell der Intentionalität. In P. Spät (Ed.) *Zur Zukunft der Philosophie des Geistes* (pp. 59-91). Paderborn, GER: Mentis.
- (2008b). Selbstgefühl. Damasio's Stufentheorie des Bewusstseins und der Emotion. In E. Düsing (Ed.) *Geist und Psyche* (pp. 337-369). Würzburg, GER: Königshausen & Neumann.
- (2011). Non-conceptual content and the subjectivity of consciousness. *International Journal of Philosophical Studies*, 19 (3), 489-518. [10.1080/09672559.2011.595197](https://doi.org/10.1080/09672559.2011.595197)
- (forthcoming). Selves, or something near enough. In S. Grätzel (Ed.) *Buddhism and Bioethics*.
- Searle, J. R. (1983). *Intentionality*. Cambridge, MA: MIT Press.
- (1992). *The rediscovery of the mind*. Cambridge, MA: MIT Press.
- Shoemaker, S. (1968). Self-reference and self-awareness. *Journal of Philosophy*, 65 (19), 555-567.  
[10.2307/2024121](https://doi.org/10.2307/2024121)
- Strawson, G. (1997). The self. *Journal of Consciousness Studies*, 4 (5-6), 405-428.
- (2004). Real intentionality. *Phenomenology and the Cognitive Sciences*, 3 (3), 287-313.  
[10.1023/B:PHEN.0000049306.63185.0f](https://doi.org/10.1023/B:PHEN.0000049306.63185.0f)
- Tye, M. (1995). *Ten problems of consciousness*. Cambridge, MA: MIT Press.
- Van Gulick, R. (2004). Higher-order global states. In R. Gennaro (Ed.) *Higher-order theories of consciousness* (pp. 67-92). Amsterdam, NL: John Benjamins.
- Vosgerau, G., Schlicht, T. & Newen, A. (2008). Orthogonality of phenomenality and content. *American Philosophical Quarterly*, 45 (4), 329-348.
- Welpinghus, A. (2013). *Emotions as natural and social kinds*. (Unpublished)
- Williford, K. (2015). Representationalisms, subjective character, and self-acquaintance. In T. Metzinger & J. M. Windt (Eds.) *Open MIND*. Frankfurt a. M., GER: MIND Group.
- Wilson, R. & Foglia, L. (2011). Embodied cognition. *The Stanford Encyclopedia of Philosophy*. E. N. Zalta (Ed.) <http://plato.stanford.edu/entries/embodied-cognition/>
- Zahavi, D. (1999). *Self-awareness and alterity. A phenomenological investigation*. Englewood, IL: Northwestern University Press.
- Zeki, S. (2007). A theory of micro-consciousness. In M. Velmans & S. Schneider (Eds.) *The Blackwell companion to consciousness* (pp. 580-588). Oxford, UK: Blackwell.
- Zeki, S. & Bartels, A. (1998). The autonomy of the visual systems and the modularity of conscious vision. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 353 (1377), 1911-1914.  
[10.1098/rstb.1998.0343](https://doi.org/10.1098/rstb.1998.0343)

---

# Individuation, Integration, and the Phenomenological Subject

A Reply to Tobias Schlicht

[Kenneth Williford](#)

---

Tobias Schlicht argues that subjective character derives from the integration of mental states into a complex of representations of the organism and that therefore there is no need to try to account for subjective character in terms of “reflexivity” or self-acquaintance, as I do. He further argues that the proper subject of consciousness is the whole organism and not the episode or stream of consciousness, as I maintain. He maintains that his account solves problems about the individuation and synchronic unity of conscious mental states that mine does not. While I agree that we need an account of the individuation of episodes of consciousness and an account of the synchronic and diachronic unities of consciousness (something I bracketed in my paper), I disagree that making the organism into the phenomenological subject of consciousness helps with these problems. However, I am willing to concede that the organism is the subject of consciousness in some non-phenomenological sense.

## Keywords

Conscious vs. unconscious mental states | Individuation | Integration | Organism | Phenomenological subject | Reflexivity | Self-acquaintance | Unity of consciousness

## Author

[Kenneth Williford](#)  
williford@uta.edu  
The University of Texas  
Arlington, TX, U.S.A.

## Commentator

[Tobias Schlicht](#)  
tobias.schlicht@rub.de  
Ruhr-Universität  
Bochum

## Editors

[Thomas Metzinger](#)  
metzinger@uni-mainz.de  
Johannes Gutenberg-Universität  
Mainz, Germany

[Jennifer M. Windt](#)  
jennifer.windt@monash.edu  
Monash University  
Melbourne, Australia

## 1 Introduction

In his insightful commentary on my contribution to *Open MIND*, Tobias Schlicht argues for the following claims: (1) The subject of conscious episodes should be identified with the organism whose episodes they are ([Schlicht this collection](#), pp. 2, 8-9). (2) Once we understand how non-conscious mental states (perceptions, thoughts, etc.) become conscious by being integrated into the underlying organismal creature-consciousness, we will have understood all that is important about how a

conscious state is endowed with subjective character ([Schlicht this collection](#), pp. 5-6). And, (3), such an account would obviate an account like mine, since there would be no need to imagine that individual episodes of consciousness have a sort of self-contained subjective character (which I construe in terms of “reflexivity” or self-acquaintance)—instead, their subjective character would just derive from their integration into the underlying creature-consciousness, which *ipso facto*



makes the organism to be the subject-pole of the episode (Schlicht [this collection](#), pp. 5-8).

Part of claim (3), as stated, (the part that begins with “since”) is my interpretation of Schlicht, since he does not spell out the claim in great detail, given limitations of space. So my arguments directed at that interpretation may not target exactly what Schlicht had in mind. But claims (1) and (2) are stated very clearly (Schlicht [this collection](#), pp. 5-6, 8-9). In this reply, I will take issue with these three claims and discuss some of Schlicht’s other claims in relation to them.

## 2 The phenomenological subject and the organism

I readily admit (and did so in the contribution) that the claim that the subject of consciousness is the episode (or stream) of consciousness itself is rather counterintuitive.<sup>1</sup> However, part of this counterintuitiveness can be ameliorated easily enough. To begin with, I should have made clearer that I was talking about what I like to call the “phenomenological subject” of consciousness. I did use the phrase, but I did not explain it and should have done so. The phenomenological subject is just *that to which* the objects of phenomenal consciousness *seem* to appear. In other words, granted that consciousness seems to have a subject-object, relational structure, the phenomenological subject is just the subject-pole of conscious experience in so far as it is given (reflectively as well as pre-reflectively).

Now, suppose we interpreted the Humean “no-self” intuition in the strongest possible way. In that case, we would conclude that there is no phenomenological subject at all. As I argued in the paper, I think the Humean intuition is not to be dismissed. However, I do think that the subject-object polarity of consciousness is a datum and not projected or inferred. If one is already sympathetic with the idea that consciousness is always aware of itself (is its own “secondary object”, as Brentano put it (see

Brentano 1995, pp. 128ff.), or is always non-positionally conscious of itself, as Sartre put it (see Sartre 2004, p. 8)), then it is not much of a stretch to identify this feature of “reflexivity” with subjective character and, if we must reify, make the episode or stream into the phenomenological subject. As long as one understands that by “subject” here, I just mean the subject-pole of conscious experience, a pole that one is phenomenally conscious of, then one will get my meaning. Given a commitment both to some version of the Humean intuition and to the intuition that one is phenomenally aware that consciousness has a subject-object polarity, one will need to resolve the tension between these in some way. Self-representationalism and self-acquaintance theories can do this in a very elegant way, I argued, a way that neither first-order nor higher-order representationalisms can. All of this is compatible with the evident fact that we normally experience ourselves as having a body in space that bears various relations to objects in space; but the phenomenological subject of consciousness should not, in my view, be identified with the body or with a representation of the body.

One might, of course, use “subject of consciousness” in different ways. One might, for example, mean “that organism or system to which we attribute consciousness” or “that which is the substrate of consciousness in an organism”. We might then speak of the “metaphysical subject” or “ontological subject” of consciousness, rather than the “phenomenological subject”. The metaphysical subject of consciousness need not appear to or be represented in consciousness. I have nothing against the idea that the organism (or a set of sub-processes of it) is the metaphysical subject of consciousness. I grant, moreover, that we normally speak in such a way that the grammatical subject of attributions of consciousness (or conscious mental states) is a noun that refers to an organism. We say things like “Skipper, my dog, sees his food coming” or “The bird saw me walking toward it and became frightened”. Indeed, insofar as consciousness is a property of or process going on in the brain of organism, there is nothing erroneous about such attributions. However, given the

<sup>1</sup> Indeed, after hearing me present a version of the target paper (at TCU in 2014), Michael Tye told me that even on a charitable reading, the claim lacks a truth value. At least, that’s what I understood him to mean.

falsity of animism and the commitment to what I like to call encephalism (the view that consciousness resides in the brain), the ontology of consciousness cannot just be read off of the grammar of such attributions, not that Schlicht is suggesting that it could.

It would, however, be more accurate to say in the Skipper case that there is a process of phenomenal visual consciousness having such-and-such representational content and being connected in such-and-such a way with Skipper's volitional, appetitive, and motor systems (all, of course, related to Skipper's organismic homeostatic systems) going on in Skipper's brain when, in normal conditions, the food is presented to him. Of course, I just referred to Skipper *qua* organism multiple times in reformulating this apparently simple attribution, but that just has the effect of roughly localizing the conscious process and, of course, connecting it to Skipper's behaviors and functions as an organism. No one should deny that consciousness, as it has arisen in organisms with an evolutionary origin, has a biological function, though it is highly debatable that consciousness should be defined or analyzed in terms of such functions. It may well be that it serves these organismic functions but could exist in substrates that do not have them or need them. In fact, I would put my money on the claim that artificial consciousness is possible in systems whose homeostatic functions can be carried out in a way that its consciousness does not contribute to *at all*. But that is a debate for another time.<sup>2</sup>

The counterintuitiveness of claiming that the phenomenological subject of consciousness is the episode or stream of consciousness might derive in part from the oddness such a view would seem to introduce into our quotidian attributions of conscious mental states, if we were to try to make our ways of speaking match this theory. It would be rather odd indeed to say, "Skipper's current episode of consciousness sees the food coming". But the view I defend does not really legitimate such locutions. Those attributions run together the phenomenological and

metaphysical notions of "the subject". The sense of counterintuitiveness that comes from saying "the episode sees..." (etc.) stems from the fact that the episode is *not* the metaphysical subject.

When we make normal attributions of conscious mental states to a creature, we encode information about the location or individuation of the conscious episode (and this gets construed as "ownership"—it's Skipper's seeing), information about the representational content and modality or attitude the episode involves (food and seeing, respectively, in this case), and a sort of folk theory about the relational structure of consciousness. That folk theory puts whole organisms or agents, as it were, "behind" the conscious mental state, as the point of view or subject pole from which the experience emanates or, to use another metaphor famously attacked by Dennett (see 1991, ch. 5), as the spectator in the "Cartesian Theater". That folk theory is hopelessly homuncular, it seems to me. It offers no analysis of what a subject is and gives no hint as to what the real conditions of unity are for either organisms or subjects. (By contrast, self-representational and self-acquaintance theories try to preserve what is right about the Cartesian Theater intuition while avoiding a commitment to homuncularism.)

When we say "Skipper sees" we do not really mean that Skipper is, as it were, behind some sort of internal telescope looking out of his eyes or at some internal screen, though the first theory of seeing that some kids come up with is indeed the homuncular and regressive one according to which there is a little person in our heads looking at just such an internal screen. It seems that what is encoded in the folk theory implicit in our normal conscious mental state attributions is something like a homuncular projection of our third-person experience of other conscious organisms onto the organism's first-person experience. In other words, we see Skipper with his excited behavior and the food out there in front of him, some distance from his body; we then imagine that this relationship that we see "sideways on" (to borrow a phrase from John McDowell, see e.g., McDowell 1994, pp. 34-36) is, so to speak, rotated 90 degrees

<sup>2</sup> It is also possible that something other than consciousness could carry out most (but not all) of the functions consciousness performs in us and other organisms; this is also a debate for another time.

and moved inside Skipper’s head—with Skipper as an irreducible agent assuming the position behind the eyepiece of his internal periscope.

That may be a bit too fanciful an exercise in the conceptual archaeology of folk-psychological mental state ascriptions, but the main point is just that our normal attributions of conscious mentality seem to run together generally accurate information about individuation or location (“ownership”), content, and attitude with a naïve and homuncular picture of subjective character. So, yes, it is true, I would say, that the organism is the subject of consciousness in the sense that conscious episodes (so far anyway) take place in organisms (actually in their brains) and causally depend upon organismic metabolic and homeostatic processes for their existence. In this sense, the organism is the *metaphysical* subject of consciousness, and this is properly reflected in our usual mental state attributions. However, I would emphatically (perhaps even hysterically) deny that the whole organism could be the *phenomenological* subject of consciousness. This is something also reflected in our usual attributions, but this is because the metaphysical and phenomenological subjects are simply conflated by folk psychology. The whole organism could not be the phenomenological subject for two reasons.

First, if one agrees with me, as Schlicht seems to, that subjective character and the subject-object polarity are phenomenally manifest even in pre-reflective consciousness, but one adds to this the claim that the phenomenological subject is the organism, then it would seem to follow that we are always aware of ourselves *qua* organism when we are consciously aware of anything—since, again, all consciousness by hypothesis has subjective character. Now, this could mean either that we represent ourselves *qua* organism or that we are acquainted with ourselves (and we are, in fact, organisms). Surely we do not, at the level of consciousness, represent ourselves *qua* organisms all the time, unless all one means by that is that consciousness has a biological function of some sort (that is, in, say, the teleofunctional sense, consciousness is “about” the organism and its ongoing relationship to the world). The latter is

undoubtedly true, but that is not a phenomenological characterization of subjective character; rather it is a thesis about the function of consciousness and its relation to organismal homeostasis. Of course, one could make an identity claim according to which subjective character (as experienced) really just is the suitably integrated representation of the organism, but this then would mean that one is embracing some form of P-theory (a theory according to which conscious representational states are distinct from non-conscious ones in part because they target some privileged object, e.g., the organism, a substantial self). If the claim is taken to mean that the organism is self-acquainted, then I might be willing to agree depending on the spin one puts on that claim.

One might just mean that there is some sub-process of the organism that is self-acquainted (that is self-manifesting or directly phenomenally self-representing, if one prefers). If that is all that is meant by the claim, then I can agree. Something like this is exactly the position I defend in the paper. After all, the central claim was just that consciousness is self-acquainted. And it was an unstated assumption of the paper that consciousness is a sub-process of the brain, and the brain a part of the organism. If, on the other hand, one means that the *whole* organism is directly acquainted with itself, this seems to me to be either an unexamined endorsement of commonsense, homuncular ways of making conscious mental state attributions (criticized above) or the claim that consciousness *necessarily* involves the entire organism.

The latter disjunct seems as false to me as the former. Yes, the prolonged existence of consciousness depends on the prolonged operation of the essential metabolic and homeostatic functions. And, indeed, almost certainly the metabolic functions that support synaptic transmission and some other basic neuronal processes are *sine qua non* for consciousness as it happens to be implemented in human and animal brains. But none of the specific means whereby our bodies support these functions have to be in place, it seems to me.

We can have artificial hearts and artificial respiration. In principle, we could offload all the

metabolic processes outside those internal to the nervous system to non-natural machines. And we could even, in principle, replace the natural generation of essential neurotransmitters with their artificial synthesis and, possibly, artificial projections for distributing them in the brain properly. Homeostasis could be maintained artificially and, in principle, without the relevant brainstem nuclei needing to do anything anymore (unless, of course, some of their activities just *happen*, for totally contingent evolutionary reasons, to be constitutively necessary for the occurrence of consciousness in brains like ours).

In short, it is certainly physically possible (though technologically beyond our current means) to keep a brain alive and operating in a “vat”! As long as we maintain those processes that are the neural correlates of (are identical to, in my view) consciousness, there would be consciousness in that brain in that vat. I am sorry, but all the evidence indicates that encephalism is true. And it seems to me to be a sort of externalist fetishism to think that consciousness literally extends beyond the brain (save by intentionality and causal interfacing). As Dan Lloyd says, we *already are* brains in vats! The cranium is the vat! (See [Lloyd 2004](#) pp. 244-245.)

Haven’t we learned from dreams, hallucinations, ALS, direct brain stimulation, and locked-in syndrome that consciousness does not need anything but the relevant brain functions to exist? Of course, the functioning brain depends upon a properly functioning body, but this does not mean that consciousness should be identified with those (other) bodily functions in some way. If we go this route, adding bodily correlates to neural correlates, when the latter causally and distally depend on the former, what is stop us from adding everything the body depends upon to our list of correlates (the gravitational constant, the bonding properties of molecules, the stability of the proton, etc.)?

True, at a certain level of analysis, it is hard to say precisely where the nervous system ends and the rest of the body begins. But then the same can be said for the body and the rest of the world (especially given that we routinely

appropriate, by breathing and eating, parts of that world). The boundaries are fuzzy at a certain scale, but this does not mean we should say there are no boundaries at all. Moreover, causal dependencies and interdependencies are myriad, but the causal relation is not the parthood relation—we cannot infer from “X causally depends on Y” that Y is a part of X. Human consciousness, as it is currently implemented, causally depends on respiration, but this does not mean that respiration is part of consciousness or that the physiological correlates of respiration are also correlates of consciousness.

If it is true that consciousness resides in the brain and depends on our specific homeostatic and metabolic processes only contingently (and I mean physically contingently, not merely logically contingently), then saying that consciousness requires an organism as a subject-pole starts to look a little fishier. This brings me to the second main reason why I would reject the claim that the whole organism is the phenomenological subject of consciousness.

What, after all, is an organism? It is clearly not just a collection of parts. We can agree with the Aristotelian tradition that it is a functional unity. We can agree that it is a system that, in virtue of its form or organization, is able to give rise to temporal successors (and I don’t necessarily mean offspring) that maintain that form or organization (at least for a while and at least within the range of conditions it evolved to live in). The matter always changes, but the form remains, from cradle to grave.

The organism takes matter from its environment to keep its processes going. And relative to its function of maintaining homeostasis (thereby giving rise to temporal successors that have the same organization it does) and to the scale at which those functions are, so to say, visible, we can truly say that the organism behaves like a goal-directed *whole* with interdependent parts and processes (and organs—the heart needs the lungs; the lungs need the heart; the kidneys need the stomach, etc.). This is all fine and dandy. But it is clear that organs are not to be identified with the whole organism. Skipper is not Skipper’s heart, though without it (or some suitable artificial replacement) little

Skippy would soon cease to be an organism at all and eventually become soil (or parts of other organisms).

More to point here, the brain is not the organism. Consciousness resides in the brain. We could, in principle, preserve consciousness simply by preserving the functioning brain. If this is true, then you do not need an organism for there to be consciousness, you just need a suitable *organ*—the brain.

I would never deny, of course, that we normally represent ourselves as having a body and relating to a world through that body. No doubt about it. Moreover, I believe it is metaphysically necessary that any consciousness be embodied in some substrate and that this embodiment configures consciousness in a way that is phenomenologically accessible to a certain extent. And, of course, we are organisms with a certain natural history. All of this “facticity” does indeed configure our consciousness to one degree or another. If all the organismal claim comes down to is that conscious beings are necessarily acquainted with their own contingent embodiment in a certain manner, then I will wholeheartedly agree. If it means that, contingently, consciousness evolved out of and is still connected to basic homeostatic functions (in some way), I will regard the claim as a not implausible hypothesis to be investigated.

This last claim is not, it seems to me, exactly what Damasio (see [Damasio 2010](#), ch. 2) and philosophers who follow him on this point, like [Charles Nussbaum \(2003\)](#) and apparently Schlicht himself, mean. They want to say something stronger. Like Francisco Varela (who possibly influenced Damasio on this point, see [Damasio 1999](#), p. 347; cf. [Varela 1979](#); [Maturana & Varela 1980](#)), they seem to want to connect consciousness essentially in the constitutive sense to the kinds of processes that are involved in homeostasis and the very emergence of an internal organism/non-organism distinction. It may well be that the subject/object distinction apparent in consciousness is, in evolutionary terms, some sort of extrapolation of this more basic distinction. It is clear, however, that, whatever the exact relation, the organism/non-organism distinction in, say, the immune sys-

tem, cannot just be identified with the subject/object distinction in consciousness. On Damasio’s view (as I understand it anyway) consciousness arises out of multiple, integrated layers of representation of the organism/non-organism distinction, where this representation itself has a certain regulatory function.

I certainly agree with [Schlicht \(this collection, p. 7\)](#) that, for Damasio, organismal and objectual representations have to be integrated in the right way for there to be something it is like for the organism, and I did not mean to suggest otherwise. This does not, however, make Damasio’s theory a non-representationalist theory (no more than the poise requirement makes Tye’s theory non-representationalist (see [Tye 1995](#), p. 138, [2000](#), p. 62)). As long as representation is considered a necessary condition for consciousness, the theory is representationalist, by my lights. And since the relevant representations, in Damasio’s theory, include, centrally, representations of the organism, it still qualifies as a P-theory in my sense—the organism is a privileged object. The representations that, when integrated, constitute consciousness, must include representations of the organism, according to the theory. Whatever else is represented in conscious mental states, on this sort of view, the organism certainly is. And the organism (ultimately in its guise as the “core self”) could thus serve as the phenomenological subject of consciousness.

I do understand how such a theory attempts to capture subjective character. In effect, it bundles it up into an object of a special sort that is always represented (one way or another) in any conscious mental state. For various reasons (e.g., the Fichte-Shoemaker Regress, see [Henrich 1982](#); [Frank 2004](#); [Frank 2007](#), pp. 157ff.; [Shoemaker 1968](#)), I do not think that such a theory can do the trick. Briefly, it is not enough simply to represent some object that you just happen to be. That is not sufficient to ground the manifest indexicality (the “I am this, here, now” aspect) of our conscious experience. Also, it is somewhat puzzling to require that consciousness representations have to have some specific type of content. Consciousness seems to be so flexible in this regard, that is

odd to think that there is such a “magic” object of representation. By contrast, I view subjective character (“reflexivity”) as a formal or structural feature of consciousness and not as a matter of representing some object or other (whether consciously or unconsciously)—including “the organism”. In fact, I believe such views, while on the right track relative to views that disregard subjective character, get the cart before the horse. *The representation of oneself as a self or an particular organism depends upon reflexivity, and not the other way around.*

For an organism (call it O) to benefit from representing an organism interacting with the world and with other organisms, there must be some way that it encodes that *it* is O (and not anything else). In effect, it needs a “you are here” (or rather “I am here”) dot on its map, a kind of “fixed-point” (see [Ismael 2007](#)). The mapper needs to know where *it* is on the map it has made. On pain of regress, it cannot derive a representation like “I must be here and not there, this organism on the map and not that one” without having some antecedent, unmediated self-reference or primitive self-knowledge (again, see [Shoemaker 1968](#)). Without this direct self-reference, the best we could hope for is a system that just happens to control itself by representing something that resembles itself in the relevant ways. Such a system might as well be controlling an exact duplicate in a duplicate room next door. We do not get manifest indexicality, self-location, or subjective character out of this. A system built up around reflexivity or direct self-reference (or primitive self-knowledge, if you prefer), would have all the control functions of a system that lacks it as well as these other features of subjectivity. Following a suggestion by Metzinger (personal communication), though with a certain modification, I would be happy to call this kind of reflexivity “prepersonal”. As such, it is the basis for one’s conscious representation of oneself as a person, organism, or anything else for that matter; but it is not essentially the representation of a person or an organism. Rather, it is the reflexivity of a process that happens to be housed in an organism (and in an organ within that organism) and that allows that organism to self-locate in a

multiplicity of spaces (physical, social, semantic, etc.).

I would add that that reflexivity could not itself be purely representational, as I argued in the paper. It very well could be, however, that reflexivity was first achieved in the evolution of organismic control systems and that these control systems have everything to do with the maintenance of homeostasis, though this is a contingency. *That* does not seem implausible to me at all. But that is a hypothesis about the evolutionary origin of self-acquaintance, not an account of what self-acquaintance is or how it is routinely generated and supported in the brain.

### 3 Unity, individuation, and integration

This brings me to Schlicht’s second and third claims. [Schlicht](#) is absolutely right, of course, to press me on the need for an account of the synchronic and diachronic unities of consciousness and for an account of the individuation of episodes and streams of consciousness ([this collection](#), p. 5). I bracketed such worries because I have not worked out any such accounts to my own satisfaction. I do, however, disagree with [Schlicht](#) on the idea that regarding the organism as the subject of consciousness can help with individuation ([this collection](#), p. 6). It is, in my view, not much easier to specify the metaphysical individuation conditions for an organism than it is to do so for an episode of consciousness. And it is problematic to assume that the brain has, so to speak, figured out what these conditions are for us. It is true that the brain must regulate a certain set of functions and processes in order to facilitate the maintenance of homeostasis; and I can even grant that doing so involves “representation” in a teleosemantic or functional-role sense. But this is orthogonal to any issues about the metaphysical individuation of organisms. While it may be easier to say what sort of processes an organism must involve (see above) and to roughly localize those processes than it is to say when one episode of consciousness begins and another ends, it is no easier to provide the *ultimate* metaphysical individuation conditions that ground the identities of the more basic physical processes that

both organisms and consciousness depend upon. At the end of the day we are always left turning our spades with the thought that there just exists a plurality of things in the cosmos—this proton (or bare particular or property or location) is not that one—end of story!

In the case of consciousness there is no *special* problem of metaphysical individuation, if what we are talking about is the fact that these episodes over here are in “my” head (in this brain), and those over there are in “your” head (in that brain). From a purely epistemic point of view, it does indeed seem to be the case that we individuate episodes of consciousness by reference to individuated organisms. But so what? This is a mere contingency. If, say, we saw conscious processes (or their correlates) first, and only with great effort could we locate the organism to which the processes were attached, we would individuate organisms by referring to the streams of consciousness that “own” those organisms. There is what we might call the “epistemic relativity of individuation”. This does not mean that there are no mind-independent facts about what ultimately individuates things. It just means that, since we have no access to what the ultimate individuators are, no particular way we individuate something should be regarded as privileged. We are guided by practical and interest-relative considerations. We might as well talk about the acorns’ squirrel rather than the squirrel’s acorns, but squirrels are more entertaining to us.

In any case, for any physical process, once you drink the metaphysical individuation Kool-Aid, you won’t come back to normal. At some point you’ll just find yourself saying “this is not that”. And anything you scratch (from universes, to stars and planets, to organisms, to molecules and particles) will fall apart in this connection. I believe there are real unities in nature and that conscious mental states are such real unities (it follows that I think certain brain processes are too), but from the epistemic and conceptual point of view Nagarjuna and the Madhyamikas seem to be right—we cannot pinpoint the basic individuators anywhere, they just seem to dissolve upon analysis one way or another (see e.g., Westerhoff 2009). I believe

such individuators exist but that they are inaccessible to us. And, as I argued in the paper, it is a serious confusion to think that in being self-acquainted you *ipso facto* have access to what it is that ultimately individuates you. You are aware of the unity and individuality of your episodes of consciousness, but those features, in turn, could depend on other unities and individuation conditions that you have no access to (cf. Williford 2011, pp. 202-203).

I am as happy as the next bloke to claim that conscious mental episodes arise and eventually give way to the next conscious mental episodes. Perhaps they overlap somehow to form a stream. Or perhaps they are punctate and we could empirically determine their temporal boundaries; I do not know. It is, however, clear that my conscious episode of reading H.P. Lovecraft for nostalgia’s sake before bed is not the conscious episode of eating yogurt for breakfast that I began the day with. But is it the same *organism* that eats in the morning and reads in the evening? It is indeed, given our normal (possibly partly pragmatic) individuation criteria for organisms. One would not have to be Heraclitus, however, to notice that the being that started the day is *quite different* from the one that unwisely decided to read Lovecraft before bed. And at a certain very fine-grained level of analysis, it is a *radically* different being.

We say things like “I ate yogurt this morning” and “I read Lovecraft at the age of 14” and “I am reading Lovecraft right before bed tonight”. We take this “I” to refer to the same organism and to the same “autobiographical self” to use Damasio’s term or same “ego” in Sartre’s sense of the word. But as Sartre pointed out (2004, pp. 7-9), no such temporally extended and dubitable entity could be entirely present in consciousness (so as to serve as its subjective character). Instead we have only a set of processes that remain more or less constant (and of course a causal-historical chain that is not broken). But the whole causal-historical chain does not exist at the present moment. It cannot be packed into a single conscious episode (though it could, of course, be represented in one). Nor can it, as a real pole of identity, exist

throughout all conscious episodes. There is no transcendental ego. Nor is there any transcendental organism. An appeal to the organism does not, just by itself, help us with the diachronic unity problem or the individuation problem. And though the brain may somehow unconsciously always “represent” us as organisms (in perhaps the teleosemantic sense of “represent”), it is evident that we are not always conscious of ourselves as organisms (whatever exactly that is supposed to mean). Yet subjective character is there whenever consciousness is.

Thus, I do not see how the organismal theory helps with either the individuation problem for conscious episodes and streams or the related diachronic unity problem. But it is also not obvious to me how the organismal theory could help us with the synchronic unity issue either—the issue of how different phenomenal and representational contents get integrated into whole, unified conscious mental states or episodes. I fully agree with Schlicht that a brain process (e.g., an “unconscious perception” of something) can be integrated into consciousness, making a new whole, and then possibly slip out again. And I too like Edelman’s and Tononi’s “Dynamic Core” idea (see [Edelman & Tononi 2000](#), Part IV) as a way of conceptualizing this integration and dis-integration. And I would emphatically reject the “micro-consciousness” idea of Zeki (see [Zeki 2007](#)). It does not seem to help to imagine many consciousnesses in the brain that somehow meld together to make a bigger one. (At least I hope it is not like that!)

In connection with this, I, like most people, believe that normally there is only one stream of consciousness existing in a given brain (with split-brain cases perhaps being an exception). I prefer the idea that consciousness is a type of process that has certain generic, essential structural features (temporality, subjective character, phenomenal character, representational character) and certain variable features (this comes down to variability of phenomenal and representational characters at different times). Due to some constantly fluctuating integration process, the phenomenal and representational characters of consciousness are al-

ways in flux, while temporality and subjective character remain invariant. Moreover, phenomenal and representational characters are such that they can, so to speak, expand and contract.

I can be hearing Bach’s *Musikalisches Opfer* while staring at a Jackson Pollock painting. I can then close my eyes so that only the beauty remains in my consciousness. If I open them again, then the visual horror will be reintegrated into it. When I closed my eyes there was “contraction”; when I opened them again, “expansion” back to the “size” the experience was before I closed them. There must indeed be something that accounts for this integration process and the resulting synchronic, differentiated unity of consciousness.

I completely agree with Schlicht that we need a theory that allows us to understand how something enters consciousness and how it gets integrated into a whole with other things that have already entered consciousness, but I do not know what that theory. Moreover, I do not know if it is possible to have a conscious experience of but one sensory quality in one modality (say, the auditory consciousness of a pure C tone) without any other thoughts, imaginings, perceptions, or anything else. I would say, though, that if one could, that episode of consciousness would still have subjective character (and temporality).

Though I agree that we all need an integration story, it is quite unclear to me that that story alone will give us a story about subjective character, unless, like [Van Gulick](#) (see e.g., [2004](#); cf. [Metzinger 1995](#)), one thinks that subjective character as a kind of reflexivity emerges out of integration. That may be, though I have never understood how, exactly, that is supposed to happen on Van Gulick’s view (though I do have some sense of what he means). It is not clear to me how this is supposed to happen on [Schlicht’s](#) view either (see [this collection](#), p. 11). But I would be quite pleased if an account like this could be made to work, since deriving reflexivity from integration would, it seems to me, be a theoretical advance.

In any case, for me, subjective character as reflexivity is a phenomenological given. It is



part of the data set from which I begin. It is my “Phenomenological Muse”. I could be wrong about its importance, of course, but until I am shown that, I will explore the model space that is appropriate to that intuition and leave it to empirical testing to determine whether or not the Muse was lying to me. (I will add, in a purely psychologistic vein, that many philosophers are allergic to reflexivity for purely irrational reasons. They just find it odd or too complicated or too puzzling. So they see it as an advantage if they can offer an account that gets around the need for it. For me this is like taking Marlon Brando out of *Apocalypse Now* or preferring decaf coffee to the real deal.)

As I see it, subjective character is like a universal or form. It is just the “reflexive” structure of consciousness. It is not to be reified into an entity (or refried like a bean, for that matter). If we think of it as an entity, we will find ourselves puzzling over questions about momentary subjects and how all these different subjects relate to each other over time and at a time. This is a confusion in my view. Yes, our normal use of language and the naïve ontology it encodes demand entities and substances to correspond to our nouns! But consciousness is not an entity or substance. It is a process; it is more like a wave than the medium the wave requires. It has a certain structure and a certain dynamical profile. Subjective character is, like temporality, an ever-replicated form that, in my view, is necessary for all consciousness. There is no subject *entity* strictly speaking. When there is an episode of looking back down the tunnel of previous conscious episodes that are connected in the normal way to that very episode of conscious looking, individuated subjective character is always seen. Just in terms of subjective character, all the episodes are qualitatively identical. This helps reinforce the illusion of a stable, continuous subject entity.<sup>3</sup> Again, there is no such entity. There is just a common form or structure living in the many different tokens. After Parfit (1984) and the Buddhists, we might say that this at once helps dissolve the thing we

once thought so substantial and important and draws us closer to other tokens, no matter what stream they happen to be in.

Finally, it seems to me that Schlicht ([this collection](#), p. 7) must take “creature-consciousness” to be more fundamental than “state-consciousness” (or “episode-consciousness”), whereas I would adhere to the usual idea that phenomenally conscious creatures are just those that host episodes (or states) of consciousness. I don’t see how reversing this order helps.

## 4 Conclusion

To recapitulate: (1) I do not see what is gained, either in relation to the individuation problem or the unity and integration problems, by regarding the organism as the phenomenological subject of consciousness. (2) I understand how P-theories attempt to do this by making the organism part of the representational content of every episode of consciousness, but I do not find those theories plausible or helpful, even if we stress the integration aspect of the theories (which does not make them cease to be P-theories). (3) I was less clear on how Schlicht thinks that an integration theory could account for subjective character if it deviates from the Damasio-style theory or from Van Gulick’s HOGS model, which latter I have also always found a little hard to understand, though I am in sympathy with it. (4) I would emphatically deny the existence of Zeki-style micro-consciousnesses; rather, I believe there is (normally) only one stream of consciousness per brain—and that stream can “expand and contract” as more or less gets integrated into it. (5) We do need an account of how unconscious processes get integrated into consciousness and of both diachronic and synchronic unity; but I am not prepared to offer such an account at present. (6) Regardless of how such an account goes, I take reflexivity (self-acquaintance) to be an essential structural feature of all consciousness; and I take it to be a phenomenological datum. All streams of consciousness are immediately aware of themselves, and that is the foundation of all other forms of self-representation, autobiographical cognition, and so on. (7) This reflexivity is subjective

<sup>3</sup> See Hofstadter 2007, ch. 7, “The Epi Phenomenon”, for a nice analogy in this connection: the illusion of a marble in the center of a box of envelopes arises just from the stacking of the envelopes (with their repeated structure).

character (for-me-ness), but it is a mistake to turn this structural feature into a kind of entity or homunculus. Thus in saying that the episode is the phenomenological subject, I am offering a non-homuncular account of the subject of consciousness. This ought to reduce a little bit of the weirdness of my claim that the episode is the phenomenological subject. (8) In other senses of “subject”, it is undoubtedly correct to say that the subject of consciousness is the organism, since it is (so far) organisms that have consciousness. However, strictly speaking, consciousness is a sub-process of the organism and lives in one of its organs—the brain. (9) Since we could, in principle, have conscious, functioning brains without the rest of the organism, it seems to follow that the organism is not the phenomenological subject—unless one adopts a P-theory according to which the privileged object we represent is just the organism we happen to be; but see (2) above.

## Acknowledgments

I would like to thank Thomas Metzinger and Jennifer Windt and the MIND team, once again, for organizing this volume. Thomas and Jennifer also gave me very valuable feedback on a draft of this reply—thank you both for that! I would also like to thank Tobias Schlicht for writing such a gracious, challenging, and stimulating commentary. I hope he knows that I honestly do not think he is irrationally allergic to reflexivity or that he embraces a fetishistic externalism! Cheers, Tobias!

## References

- Brentano, F. (1995). *Psychology from an empirical standpoint* (trans. A.C. Rancurello, D.B. Terrell and L.L. McAlister). London, UK: Routledge.
- Damasio, A. R. (1999). *The feeling of what happens: Body and emotion in the making of consciousness*. New York, NY: Harcourt.
- (2010). *Self comes to mind*. New York, NY: Pantheon Books.
- Dennett, D. C. (1991). *Consciousness explained*. Boston, MA: Little.
- Edelman, G. M. & Tononi, G. (2000). *A universe of consciousness: How matter becomes imagination*. New York, NY: Basic books.
- Frank, M. (2004). Fragments of a history of the theory of self-consciousness from Kant to Kierkegaard. *Critical Horizons*, 5 (1), 53-136.
- (2007). Non-objectal subjectivity. *Journal of Consciousness Studies*, 14 (5-6), 152-173.
- Henrich, D. (1982). Fichte’s original insight. In D. Christensen (Ed.) *Contemporary German Philosophy, Vol. 1* (pp. 15-53). University Park, PA: Penn State University Press.
- Hofstadter, D. R. (2007). *I am a strange loop*. New York, NY: Basic Books.
- Ismael, J. (2007). *The situated self*. Oxford, UK: Oxford University Press.
- Lloyd, D. E. (2004). *Radiant cool: A novel theory of consciousness*. Cambridge, MA: MIT Press.
- Maturana, H. R. & Varela, F. J. (1980). *Autopoiesis and cognition: The realization of the living*. Berlin, GER: Springer.
- McDowell, J. (1994). *Mind and world*. Cambridge, MA: Harvard University Press.
- Metzinger, T. (1995). Faster than thought: Holism, homogeneity, and temporal coding. In T. Metzinger (Ed.) *Conscious experience* (pp. 425-461). Thorverton, UK: Imprint Academic.
- Nussbaum, C. (2003). Another look at functionalism and the emotions. *Brain and Mind*, 4 (3), 353-383.  
[10.1023/B:BRAM.0000005469.62248.00](https://doi.org/10.1023/B:BRAM.0000005469.62248.00)
- Parfit, D. (1984). *Reasons and persons*. Oxford, UK: Oxford University Press.
- Sartre, J. P. (2004). *The transcendence of the ego: A sketch for a phenomenological description* (trans. Brown, A.). London, UK: Routledge.
- Schlicht, T. (2015). Explaining subjective character: Representation, reflexivity, or integration? In T. Metzinger & J. M. Windt (Eds.) *Open MIND*. Frankfurt a. M., GER: MIND Group.

- Shoemaker, S. (1968). Self-reference and self-awareness. *Journal of Philosophy*, 65 (19), 555-567.
- Tye, M. (1995). *Ten problems of consciousness: A representational theory of the phenomenal mind*. Cambridge, MA: MIT Press.
- (2000). *Consciousness, color, and content*. Cambridge, MA: MIT Press.
- Van Gulick, R. (2004). Higher-order global states (HOGS) An alternative higher-order model. In R. Genaro (Ed.) *Higher-order theories of consciousness* (pp. 67-92). Amsterdam, NL: John Benjamins.
- Varela, F. J. (1979). *Principles of biological autonomy*. New York, NY: North Holland.
- Westerhoff, J. (2009). *Nagarjuna's Madhyamaka: A philosophical introduction*. Oxford, UK: Oxford University Press.
- Williford, K. (2011). Pre-reflective self-consciousness and the autobiographical ego. In J. Webber (Ed.) *Reading Sartre* (pp. 195-210). London, UK: Routledge.
- Zeki, S. (2007). A theory of micro-consciousness. In M. Velmans & S. Schneider (Eds.) *The Blackwell companion to consciousness* (pp. 580-588). Oxford, UK: Blackwell.