Ergo AN OPEN ACCESS JOURNAL OF PHILOSOPHY

Verbal Disputes in the Theory of Consciousness

JOSEPH GOTTLIEB
Texas Tech University

The primary aim of a theory of consciousness is to articulate existence conditions for conscious states, i.e. the conditions under which a mental state is conscious rather than unconscious. There are two main broad approaches: The Higher-Order approach and the First-Order approach. Higher-Order theories claim that a mental state is conscious only if it is the object of a suitable state of higher-order awareness. First-Order theories reject this necessary condition. However, both sides make the following claim: for any mental state M of a subject S, M is conscious iff there is something it is like for S to be in M. This is the *Nagelian Conception* of consciousness. Taking the Nagelian Conception as a starting point, I contend that the best rationalizing explanation for the ways in which Higher-Order and First-Order theorists contribute to their dispute is to see those contributions as consistent responses to two distinct questions.

> If we agree . . . not to worry about which mental phenomena deserve the honorific title 'consciousness,' it may seem that there is nothing left about which Block and I disagree. We might even agree to apply the term 'conscious,' in a special sense, to states that exhibit only thin phenomenality.
>
> David Rosenthal (2002a: 660)

1. The Transitivity Principle

The relationship between what makes a mental state conscious and awareness of our mental states is a contested one. Consider:

1. Michael's perception of the tree is conscious.
2. Michael is aware of his perceiving the tree.

Contact: Joseph Gottlieb <joseph.gottlieb@ttu.edu>

http://dx.doi.org/10.3998/ergo.12405314.0005.012

According to *Higher-Order* (HO) theories of consciousness, (1) entails (2). This is because a mental state's being conscious consists in its subject being suitably aware of it. HO theorists codify (2)'s entailment by (1) as *the Transitivity Principle*, or TRANSITIVITY for short:

> **TRANSITIVITY** A mental state is conscious only if one is in some way aware of that mental state.

Different HO theories will have different things to say about the way we are aware of our conscious states. For instance, Higher-Order Thought theorists (e.g., Rosenthal 2005) say that Michael is aware of perceiving the tree by having an appropriate higher-order thought (a 'HOT') about his perceptual state, while Higher-Order Perception theorists (e.g., Lycan 1996) will appeal to an appropriate higher-order perception (a 'HOP').

By contrast, *First-Order* (FO) theories of consciousness deny that (1) entails (2), and so reject TRANSITIVITY.[1] Beyond having this shared belief in the falsity of something, FO theories are a motley crew. One can be a naïve realist (e.g., Martin 2004), a qualia realist (e.g., Shoemaker 1994; Block 1996), or a representationalist (e.g., Dretske 1995; Tye 2000) and be a FO theorist. My interest, however, is in a certain subset of FO theorists. I am interested in FO theorists that purport to give a positive account of what it is for mental state to be conscious that is in direct conflict with the HO-theoretic approach. By this measure I count Ned Block, Michael Tye, Fred Dretske, and Jesse Prinz as FO theorists in the relevant sense.

I am also only interested in a mental state's being *phenomenally* consciousness. So (2)'s entailment by (1) is shorthand for (2)'s entailment by (1*): Michael's perception of the tree is phenomenally conscious. In this way, my focus is on what Block (2011) calls 'ambitious' HO theorists; if a HO theorist might be read as giving an account of some *other* form of consciousness (e.g., introspective consciousness), then the FO theorist might have no qualm with *that* sort of HO theorist. By the ambitious measure I count David Rosenthal, Uriah Kriegel, Rocco Gennaro, and Josh Weisberg (amongst others) as HO theorists in the relevant sense.[2]

In our epigraph, David Rosenthal suggests that his disagreement with Ned Block—and presumably other FO theorists—comes down to a disagreement about what phenomenon deserves the honorific 'consciousness.' Rosenthal

1. There are other, subtly different, formulations of TRANSITIVITY that are orthogonal to the HOT and HOP implementations of TRANSITIVITY mentioned here. Although these alternative formulations have their virtues, they would unnecessarily complicate the present discussion, especially given that FO theories reject those formulations too. For a discussion of these other formulations in a different context, see Gottlieb (2016).

2. It is less obvious that Armstrong (1968) and Lycan (1996) are ambitious HO theorists.

was likely being facetious.[3] Nonetheless, I'll argue that Rosenthal's speculation, made in jest or not, is right: the dispute over TRANSITIVITY, I contend, is verbal. By exploring the actual practice of current HO and FO theorists and the semantics of conscious state ascriptions, this paper attempts to build a model of what, at bottom, is going on the dispute between HO and FO theorists. What makes the model successful is that it provides the best rationalizing explanation for the ways in which parties from both sides of the aisle contribute to the dispute over TRANSITIVITY. This picture will strike many as highly revisionary, and because of that, the burden of proof is high. To that end, I close by attempting to speak to the intuition that HO and FO are engaging in a debate that is well worth having, even if—contra the standard picture—it is not at the level of literal expression of conflicting semantic contents.

2. The Nagelian Conception of Consciousness

We begin with the Nagelian Conception of consciousness, or 'TNC':

> **TNC** For any mental state M of a subject S, M is conscious iff there is something it is like for S to be in M.

TNC plays a curious role in the philosophical literature. On the one hand, its presence is rife. Block (a FO theorist) tells us that,

> Phenomenal consciousness is experience; what makes a state phenomenally conscious is that there is something "it is like" (Nagel 1974) to be in that state. (2002: 206)

And Michael Tye (another FO theorist) says,

> A mental state, then, may be said to be phenomenally conscious just in case there is something it is like to undergo the state. . . . (1997: 290)

Likewise, according to Rosenthal (a HO theorist),

> When one lacks conscious access to a state, there is literally nothing it's like for one to be in that state. Without access to a state one has no first-person perspective on it, and so there is nothing it's like to be in it. As

3. Although we occasionally see others at least taking the idea seriously. See, e.g., Byrne (2004), a FO theorist, and Kriegel (2009), a HO theorist. I discuss Kriegel's take in Section 3.4.

> Thomas Nagel has insisted, what matters for consciousness is that there be something it's like "for the organism" (Nagel 1979: 166). And there will be something it's like for the organism only if the organism has conscious access to the relevant state. (2000: 275)

And Uriah Kriegel (another HO theorist) says,

> Phenomenal consciousness is the property mental states have when, and only when, there is something it is like for their subject to undergo them, or be in them.[4] (2006: 58)

So despite the disagreement over TRANSITIVITY, TNC is common currency for HO and FO theorists alike. On the other hand, HO theorists are not only keen to point out that TNC entails or at least strongly suggests TRANSITIVITY, but that this fact is obvious. It is so obvious, we are told, that to think otherwise is to abuse or misunderstand TNC. Here are two examples.

The first concerns an exchange between Block (2011a) and HO theorist Josh Weisberg (2011). Recapitulating Block's criticism of HO theories, Weisberg tells us,

> [According to Block,] [m]y defense entails that the what-it's-like ness occurring during an episode of HO misrepresentation 'is fake'! There is no first-order state, and so . . . there is no property of what-it's-like-ness. . . . And by this logic, the HO theorists cannot explain why there is an important difference between those really having conscious pains (replete with real what-it's-like-ness) and those merely representing themselves as having conscious pains (possessing only fake what-it's-like-ness). Block concludes that there is no way to explain why pain matters on the HO approach. And that means that the approach is 'about consciousness in a *merely technical sense* of the term.' This is not meant as a compliment. (2011: 439, emphasis added)

Weisberg is not convinced:

> According to Block's reading, pains matter even if the subject is in no way aware of them. Consider a person in such a state. She will not be aware of being in pain. She will deny that anything is wrong. . . . But according to Block, the pains that matter are still present. The subjects

4. Kriegel (2009) is a self-representationalist. Self-representationalism counts as a version of HO theory since it entails TRANSITIVITY, and appeals to higher-order content, i.e., content whose subject matters involves representations or representational properties.

> might even be in agonizing pain—that is, they might possess states with the what-it's-like-ness property of agonizing pain—even though there's nothing it's like for the subjects. I contend that, whatever is going on here, to characterize such episodes as conscious is to use a '*merely technical sense* of the term.' (2011: 440, emphasis added)

By Weisberg's lights, to characterize mental states of which a subject is wholly unaware as conscious—to reject TRANSITIVITY—is to employ a 'merely technical sense of the term.' Block didn't mean this as a complement. I doubt Weisberg did either.[5]

A second example comes in Block's discussion of visual extinction. Visual extinction is caused by damage to the parietal lobe. Extinction subjects can report stimuli on either side of their visual field so long as the stimuli are not presented simultaneously. If both stimuli are presented at once—one on the unimpaired side of the visual field and one on the impaired side—such subjects can only report the stimulus on the unimpaired side. Perception of the stimulus on the unimpaired side thus 'extinguishes' perception on the impaired side. Block (2001: 203) makes three claims of interest here: (i) that extinction subjects do not know—indeed cannot know—or be aware, of their perceiving the stimuli in the impaired portion of their visual field; (ii) that TNC is true; and (iii) that despite (i) and (ii), extinction subjects nonetheless undergo conscious experiences with respect to the impaired portion of their visual field.[6] Rosenthal regards (iii) as wildly implausible. He says,

> Block . . . allows that phenomenality can occur not only without one's knowing it, but in cases in which one would firmly deny its occurrence. This does not fit comfortably, however, with the explanation of phenomenality as "what it is like to have an experience." It's important to distinguish this somewhat special use of the phrase what it's like to describe subjectivity from its more general, non-mental use. There is something it's like to be a table, or even to be this very table. What it's like to be a table . . . is roughly something's having characteristic features of tables.
>
> But this is of course not what's involved in talking about what it's like to have an experience. As Nagel stressed. . . . what it's like to have an experience is what it's like for the individual that has the experience. When a person enjoys the taste of wine . . . there is something it's like *for that person*. . . .

5. The backstory concerns a version of the 'empty HOT' objection to the HOT brand of HO theories, which Block deems fatal. Nothing to follow turns on these details.

6. See also Block (2008: 291). Block is not alone amongst FO theorists in endorsing (iii). See, for instance, Dretske (2006: 167).

> Not so in cases of visual extinction; there is nothing it's like for an extinction subject to have a qualitative experience of the extinguished stimuli. That's why seeing visual extinction as the having of phenomenality without one's knowing it does not fit comfortably with the explanation of phenomenality in terms of what it's like to have an experience. (Rosenthal 2002a: 656)

Now compared to Weisberg, it is perhaps less clear that Rosenthal is charging Block with linguistic abuse; after all, he says that allowing for phenomenality in states of visual extinction "does not fit comfortably" with the Nagelian usage. That doesn't have the same linguistic ring as saying that someone's usage of a term is a merely technical usage. But in fact this is not all Rosenthal is saying. The comparison with what it's like to be a table suggests something stronger, viz., that Block's usage does not even *remotely* fit comfortably. To my ear, that ought to strike one as at least hinting at the possibility of linguistic abuse. So there is a sense in which Rosenthal is making essentially the same point as Weisberg: if there is something it is like to see extinguished stimuli on Block's account, then Block is abusing TNC. Block is not just wrong about visual extinction. He is *obviously* wrong. He simply *doesn't get* TNC. That's quite a claim.[7]

In sum, the presumption—at least for the HO theorist—is this: if one thinks that TRANSITIVITY is false, one cannot have TNC in mind. Such is the real import of Weisberg's right-back-at-you to Block. Weisberg is not being flippant. He is insisting that Block must be talking about something else entirely given TNC. These observations raise the following possibility in an acute form: could HO and FO theorists disagree about TRANSITIVITY because they disagree, perhaps implicitly, about the meaning of TNC?

7. Perhaps feeling pulled by such reactions, some FO theorists have floated the idea of preserving something like TRANSITIVITY by saying that conscious mental states are mental states we are 'automatically' aware of, in a way akin to how one automatically laughs one's laugh or dance's one dance (e.g., Block 2007; Tye 2009). At least in Block's case, it is hard to see what this could come to aside from a highly *deflationary* view of HO awareness, since it won't amount to any form of perceptual or cognitive access of the state in question, and will not—as the extinction case shows—be the sort of awareness that results in knowledge or being able to express that knowledge in verbal report. It's also important to note that this option has never been fully and explicitly endorsed by FO theorists, or unpacked in any meaningful detail, save from the obvious truth that it is the sort of awareness that, if real, is reported by means of an internal accusative. Indeed, Tye actually *denies* that the relevant sense of 'experiencing an experience' consists in being conscious *of* that experience, saying that "supposing otherwise is no more plausible than supposing that if I have a laugh at a joke and in so doing I laugh a laugh at the joke, I am laughing at my laugh" (2009: 5). In any event, the point is this: for our purposes, 'HO awareness' picks out non-deflationary HO awareness-of. Awareness of a mental state will thus imply subject-level access to that state, or at least subject-level access to some features of that state. Mental states that are not even accessible, like those at play in visual extinction, are not objects of (non-deflationary) HO awareness.

3. The HO-FO Dispute is Verbal

3.1. Preliminaries

Let a dispute be any exchange that involves (at least) two parties appearing to disagree over a sentence *s*, where one endorses *s*, and the other endorses ¬*s*. Finding conditions that are jointly necessary and sufficient for a verbal dispute is not easy. Luckily, a sufficient condition will do. Here I follow Eli Hirsch. According to Hirsch, two parties A and B are in a verbal dispute over *s* if A and B are engaged in a sincere prima facie dispute, but, given the correct view of linguistic interpretation, A and B will agree that each "speaks the truth in its own language" (2011: 239).[8]

Now we saw that all parties agree that TNC is true. So, if we substitute TNC for 'is conscious' in TRANSITIVITY, we get the following disputed sentence:

> **NAGELIAN TRANSITIVITY** A mental state is like something for its subject only if its subject is in some way aware of that mental state.

This allows us to frame our main thesis with more precision. I'll argue that HO and FO theorists disagree about the truth-value of TRANSITIVITY because HO and FO theorists are having a (tacit) semantic disagreement about the meaning of the expression 'like something for its subject' in NAGELIAN TRANSITIVITY.[9]

In advancing this claim, I don't assume that HO and FO theorists are faultless in every respect. A verbal dispute can arise due to one side misusing a portion of language that the other side is using correctly, and that very well may hold here. One side might just get things wrong about the meaning of TNC. Therefore, I make no assumptions as to which side, or whether any side, employs 'conscious,' 'what it is like,' or similar expressions in a deviant manner.

Such is my thesis. Here, in broad strokes, is how I defend it.

We have a disputed sentence: NAGELIAN TRANSITIVITY. The HO theorist says that it is true; the FO theorist says that it is false. We are looking for a charitable interpretation that makes each side right. This can come about in a number of ways, but one way—what Hirsch (2011: 162) calls "the simplest paradigm"—is

8. More accurately, it's an *imagined* public language. Hirsch thinks that semantic externalism commits us to this stipulation. But for doubts about the necessity of this stipulation, see Jackson (2013).

9. For HO and FO theorists to have a verbal dispute it only need be the case that they *take* the expression under dispute to have different meanings, whether or not this ultimately reflects any real difference in literal meaning for them. They need not be having what Chalmers (2011) calls a 'narrow' verbal dispute. I also assume that whenever speakers use an expression, they do so with beliefs about its meaning, be they tacit or explicit.

this. Take your disputed sentence *D*, and two undisputed sentences, U1 and U2. Then argue that one side holds that *D* is equivalent to U1 and the other side holds that *D* is equivalent to U2. Each disputant can then charitably conclude that, *in the other side's language*, one is speaking the truth with respect to *D* because the "asserted equivalence holds in that language" (Hirsch 2011: 162).[10]

That's the basic template I'll follow. The details unfold in three steps. The first step consists in the construction of competing languages, i.e., languages that assign different truth-conditions to NAGELIAN TRANSITIVITY. This requires competing semantics for 'what it is like'-sentences. Examples of 'what it is like'-sentences include NAGELIAN TRANSITIVITY, but also more common English sentences like:

(a) There is something it is like to smell garlic.
(b) There is something it is like for Michael to see the sunset.
(c) What is it like to hear Led Zeppelin?[11]

Appealing to recent work by Daniel Stoljar (2016), Section 3.2 sets out the two main competitors for a semantics for 'what it is like'-sentences: *the Operator View*, and Stoljar's preferred alternative, *the Affective View*. The results are *HO-English*, which assigns its meaning to 'what it is like'-sentences in accordance with the Operator View, and *FO-English*, which assigns its meaning to 'what it is like'-sentences in accordance with the Affective View.[12]

The second step, in Section 3.3, consists in arguing that HO theorists are most charitably interpreted as speaking HO-English, and that FO theorists are most charitably interpreted as speaking FO-English. I'll follow Hirsch in assuming

10. Karen Bennett's (2009: 40) case of someone who uses 'telephone' to refer to leprechauns is a nice example of this.

11. NAGELIAN TRANSITIVITY is unlike (a) and (b) as it does not include an explicit quantifier, and unlike (c) as it is an indicative, not an interrogative. And like (a) and (b) but unlike (c), NAGELIAN TRANSITIVITY does not explicitly use the term 'what.' Despite these differences, NAGELIAN TRANSITIVITY is still a 'what it is like'-sentence for our purposes.

12. The viability of Stoljar's project is predicated on the meaning of 'what it is like'-sentences not being a technical matter. On this sort of view, 'what it is like'-sentences involve terms that may look like ordinary terms but in fact have a distinct meaning determined—as Stoljar (2016: 1183) puts it—by the 'technocrats' (i.e., philosophers of mind). Talk involving 'what it is like' would be like talk about 'work' in physics. Call this view *the Technical Account* (e.g., Lewis 1995: 140). The Technical Account is highly problematic (see, e.g., Stoljar 2016; Farrell 2016). There are several issues, but to just mention one: if the Technical Account were true, we would expect the relevant technical terms to be explicitly introduced as such. This is because technical terms sound and look just like their ordinary counterparts. Yet as Jonathan Farrell (2016) notes, that's not what we find. And he is right: Nagel doesn't do this, and neither do Wittgenstein (1980), B. A. Farrell (1950), or Sprigge (1971) before him. ('What it is like'-talk, it turns out, predates Nagel.) I'm also setting aside views where TNC is meaningless (Hacker 2002), false, or trivial (Snowdon 2010). This is defensible given the ubiquity of TNC amongst HO and FO theorists.

that principles of charity are key to the norms of linguistic interpretation. Our focus will be on one principle in particular, viz., *charity to understanding*: ceteris paribus, we ought to assume that the speakers of a language don't make a priori or conceptual errors regarding uncomplicated matters (Hirsch 2011: 182).[13] The final step, in Section 3.4, consists in arguing that the equivalences that hold in our two languages are in fact undisputedly true (in HO-English) and undisputedly false (in FO-English).

3.2. The Semantics of 'What it is like'-Sentences

Stoljar (2016) discusses several accounts of the semantics of 'what it is like'- sentences. We'll only make use of two, and their presentation will be simplified.[14]

We begin with a single canonical 'what it is like'-sentence. Let CANONICAL be the sentence formed by swapping 'see the sunset' for a general perceptual act-object schema 'ϕ'ing x' in 'There is something it is like for Michael to see the sunset,' i.e., (b) above. This gives us:

CANONICAL There is something it is like for Michael to ϕ x.

Then, according to the Affective View:

The Affective View CANONICAL is true iff there is some way that Michael feels as a result Michael's ϕ'ing x.

The Affective View is Stoljar's preferred account. Here, 'what it is like'-sentences "assert an experiential relation between an event and an individual, where an experiential relation obtains just in case there is a way the individual feels in virtue of the event" (Stoljar 2016: 1176).

Now, if by 'feels' we mean something like a bodily sensation, then the Affective View is likely false. Consider the sentence 'There is something it is like for

13. My reliance on Hirsch's framework is therefore rather minimal. Indeed, the question of charitable interpretation is more tractable in the present case when compared to those in disputes about ontology. As we will see, HO theorists are rather forthcoming about what they mean by 'what it is'-talk.

14. Two notes. First, Stoljar is concerned with 'what it is like'-sentences in their stereotypical context of use. By 'stereotypical' Stoljar means the ways of using such sentences that speakers regard as routine, even if those ways are not mandated by their linguistic meaning. Stoljar uses the example 'he has all the qualities of his father' to convey this idea (2016: 1176). If we only consider this sentence's linguistic meaning, the sentence could express propositions about any qualities. But it is good qualities that are relevant in its stereotypical use. Second, Stoljar is concerned with explicating the linguistic meaning of 'what it is like'-sentences and the propositions—here, truth-conditions—expressed in a context by such sentences, not those conveyed (2016: 1164).

Mary to see red.' Stoljar points out that if the Mary in question is Frank Jackson's Mary—the physically omniscient scientist who from birth is locked in a black-and-white room—this sentence reports an experience (Stoljar 2016: 1181). But then, though the sentence is true, Mary needn't feel any way in virtue of having her experience. It's not as if she must have touched something, or have an emotion, for seeing red to be like something for her (Stoljar 2016: 1181). Likewise for CANONICAL: Michael need not feel any way as a result of seeing the sunset.

Stoljar (2016: 1181) offers two responses to this worry, each of which can be employed by a friend of the Affective View. His first response acknowledges that while some philosophers have employed a narrow notion of feeling, English speakers are more lenient. Consider the notion of feeling at play in 'I feel like class has taken forever.' It is not implausible to think this reports an experience. But the sense of feeling involves no appeal to bodily sensations or emotions.[15]

The second response requires a modification to the Affective View. The modification turns the Affective View into a disjunctive thesis: CANONICAL is true iff (i) there is some way that Michael feels as a result of Michael's ϕ'ing x or (ii) there is some way that x seems to Michael in virtue of his ϕ'ing x, where x is not itself Michael's state of ϕ'ing. This avoids the objection from Mary, since it is plausible that things seem some way when she sees something red, even though she doesn't feel any way (Stoljar 2016: 1182). Going forward, I will largely operate under the first response. However, when a move to the disjunctive version of the Affective View might make a difference, this will be noted.

Here, though, is the central point: on the Affective View, the 'for the subject'-phrase is (i) logically speaking an argument of the verb phrase 'is like', and (ii) semantically expressing affectiveness, i.e., some way the subject feels as a result of ϕ'ing. It is *in virtue of* Michael ϕ'ing x that there is something it is like for him. It is an additional claim to say that Michael is aware of his ϕ'ing x.

The second account is the Operator View:

The Operator View CANONICAL is true iff there is some way such that it seems to Michael that Michael's ϕ'ing x is that way.

The Operator View is so-called because it appeals to an operator—a function from sentences to sentences—viz., 'it seems to you that.' The Operator View accords with the Affective View on the logical structure of CANONICAL (and other 'what it is like'-sentences), but departs over how to interpret the 'for Michael' phrase of CANONICAL. On the Operator View, the 'for Michael' phrase is (i) logically speaking an operator, and (ii) semantically expresses what 'it seems to you that' expresses (Stoljar 2016: 1186).

15. See Brogaard (2012) for broader readings of 'feeling.'

Let me note two things about 'seems.' First, it is well known that there are different uses of 'seems' and related terms like 'look' and 'appear' (e.g., Chisholm 1957). I am using 'seems' in the phenomenal sense, not the epistemic sense. The hallmark of the epistemic sense of 'seems' and 'looks' is that they disappear in the presence of a defeater if the subject is rational. But that's not so when it comes to consciousness; if I have a conscious visual experience as of a rainbow-colored elephant, the way things look or seem to me won't change even if I find out by some non-experiential means that I am in fact hallucinating.

Second, if x seems some way to a subject S, then S is aware of x. I take this entailment to be straightforward, but it leaves open *how* S is aware of x. The Operator View, as I understand it, is non-committal. It can just be simple thing-awareness. Or it can be fact-awareness.[16] The former might be attractive to HOP theorists. The latter might be attractive to HOT theorists. These matters, however, are downstream; of concern is what HOP (or HOT) theorists say *qua* HO theorists, not what HOP (or HOT) theorists say *qua* HOP (or HOT) theorists.[17] Whether the relation picked out by 'it seems to you that' need involve a certain sort of belief or disposition to believe, or some sort of perceptual relation, is otiose.[18]

What *will* matter—and what should be keyed in on as we proceed—is how the view treats 'for the subject' as an operator ('it seems to you that') whose arguments are mental states and mental events. To foreshadow, just as Hirsch says that the culprit in disputes over ontology is 'there is,' the culprit in the dispute over TRANSITIVITY is 'for the subject.'

3.3. HO-English and FO-English

We can now define two languages: HO-English and FO-English. According to HO-English, a 'what it is like'-sentence (e.g., NAGELIAN TRANSITIVITY) has its truth-conditions in accordance with the Operator View. According to FO-English, a 'what it is like'-sentence has its truth-conditions in accordance with the Affective View. With this, we face two questions. First, ought the FO theorist interpret the

16. For the distinction between thing-awareness and fact-awareness, see Dretske (1999).

17. Eric Lormand (2004) takes 'seems' to implicate an inner sense specifically, resulting in a view akin to Lycan's HOP theory

18. In fact, there is even more latitude than this. I have formulated the Operator View in terms of a particular operator ('it seems to you that') because (as we'll see) this is how Rosenthal interprets TNC, and this is how Stoljar himself formulates the Operator View. But as Stoljar (2016: 1186, Footnote 26) also notes, the Operator View could be formulated in terms of other sentential operators such as 'Michael represents it as being the case that,' or non-sentential operators like 'Michael is aware of.' The bedrock claim for our purposes is simply that, on the Operator View, 'for the subject' express a proposition concerning *some* sort of awareness-relation that holds between a subject and a mental state or event.

HO theorist as speaking HO-English? Second, ought the HO theorist interpret the FO theorist as speaking FO-English? The answer to both questions is *yes*.

We can begin by adopting the standpoint of a FO theorist. If we are FO theorists trying to make sense of what HO theorists are saying when they assert NAGELIAN TRANSITIVITY, things are relatively easy. Rosenthal—again, arguably the progenitor of modern-day HO theory—all but outright admits the Operator View.[19] Here is Rosenthal on TNC: "As many, myself included, use that phrase, there being something it's like for one to be in a state is simply its seeming subjectively that one is in that state" (2011: 434).

So Rosenthal gives us an explicit (or close to explicit) commitment to the Operator View. There are also implicit commitments. The literature is riddled with HO theorists stressing the 'for' in 'what it is like for'—often literally, by underscoring or italicizing—as if to remind us that the term 'for' alone is supposed to tell us TRANSITIVITY must be true. Stoljar aptly calls these *emphatic arguments*. When Rosenthal expressed his bewilderment that Block could possibly entertain states of visual extinction being conscious (Section 2), he points to the 'for' in TNC. Similarly, Weisberg tells us that the 'for the organism' (our 'for the subject') in Nagel's 'something it is like for the organism' suggests "a connection to the rest of the mind, a mode of access by a sentient subject" (2011: 411). Joseph Levine makes the same point:

> the very phrase that serves to canonically express the notion of the phenomenal—'what it's like for x to . . .'—explicitly refers to the phenomenal state in question being 'for' the subject. Phenomenal states/properties are not merely instantiated in the subject, but are experienced by the subject. Experience is more than mere instantiation, and part of what that 'more' involves is a kind of access.[20] (2007: 514)

Rosenthal, Weisberg, and Levine are giving emphatic arguments. The emphatic argument doesn't work, however, without the Operator View.

To see why, note how the argument unfolds. First, we are given TNC as a premise, with an added emphasis on 'for the subject':

P1 For any mental state M of a subject S, M is (phenomenally) conscious iff there is something it is like *fo*r S to be in M.

19. Everyone that I know of who is committed to Operator View in print is a HO theorist. See also Hellie (2007a), Weisberg (2011), and Janzen (2011).

20. Levine (2006) denies that the awareness-of relation we bear to our mental states is representational. He is in this respect closer to Hellie (2007b).

From P1 we add an additional premise, and then conclude that some form of HO theory is true:

P2 There is something it is like *for* S to be in M only if S is aware of M.

∴ For any mental state M of a subject S, M is (phenomenally) conscious only if S is aware of M.

Why accept P2? Speaking for the HO theorist, the answer, I take it, comes from what 'for the subject' in TNC means. Otherwise, emphasizing the 'for' in the prepositional phrase 'for the subject' would count for little. That's why the label 'emphatic' is so fitting. The term 'for' is itself doing the work. What the HO theorists have in mind, then, is a semantic auxiliary thesis:

AUX There is something it is like *for* S to be in M only if M seems some way to S.

AUX bridges P1 and P2. Yet AUX is just the right-to-left direction of the Operator View. Indeed, Greg Janzen, another HO theorist, essentially admits that AUX is doing much of the work when he tells us how "the very language of the what-it-is-like formula, the words in it, suggests that it ought to be read as expressing a proposition about a subject's awareness of her own mental states" (2011: 83). So, as an interpretive matter, charity is not much of an issue when it comes to attributing HO-English to the HO theorist. To advance the Emphatic Argument, as so many HO theorists do, is to be committed to the Operator View. HO theorists, whether explicitly or implicitly, speak HO-English.

The result is that, in HO-English, NAGELIAN TRANSITIVITY is equivalent to 'A mental state seems some way to its subject only if its subject is in some way aware of that mental state,' by substitution of the expression 'seems some way to its subject' for the (original) 'is like something for its subject.' But now the FO theorist should agree that, if the HO theorist is speaking HO-English, she is speaking the truth in her own language when she asserts TRANSITIVITY and NAGELIAN TRANSITIVITY. For notice: if (i) a mental state is conscious only if there is something it is like for the subject to be in that mental state, and (ii), there is something it is like for the subject to be in that mental state only if that mental state seems some way to its subject, it will follow that (iii) a mental state is conscious only if its subject is aware of it. The move to (iii) is licensed by the assumption that if x seems some way to S, then S is aware of x.

Let us now adopt the standpoint of the HO theorist: is it charitable to interpret the FO theorist as not speaking HO-English? It seems that it is. From the standpoint of the HO theorist, the FO theorist is making an *a priori* error when

he denies NAGELIAN TRANSITIVITY. For given the Operator View, it follows a priori NAGELIAN TRANSITIVITY is true. It is plausibly an a priori truth that if a state of ϕ'ing is like something for Michael iff that state of ϕ'ing seems some way to Michael, then a state of ϕ'ing is like something for Michael iff Michael is in some way aware of that state of ϕ'ing.

Notice that this is an exceedingly reasonable application of charity to understanding. Requiring that we avoid attributing *any* a priori mistakes at all to a subject is undue (cf. Jackson 2013: 427). It is not hard to fathom how a rational subject might have trouble recognizing a complex a priori truth, like a mathematical theorem. But the relevant a priori entailment, from 'Michael's ϕ'ing x seems some way to Michael' to 'Michael is in some way aware of his ϕ'ing x', is not complex and does not seem hard to process. This is especially so since HO theorists can avail themselves of multiple forms of higher-order awareness (cognitive, perceptual, etc.). All that is being advanced, then, is the claim that Michael's ϕ'ing x seeming some way (to Michael) entails that Michael is *in some way* aware of his ϕ'ing.

This point is critical. Charity to understanding will force the HO theorist to re-interpret the FO theorist only if the HO theorist regards her own view as being discoverable without such difficult reasoning—only if, that is, she takes the truth of TRANSITIVITY to be not like other, more complex and difficult a priori truths. TRANSITIVITY follows trivially from TNC, given what HO theorist mean by TNC. Moreover, HO theorists are forthright that they *do* view TRANSITIVITY as a non-complex truth—that is, a truth that is not difficult to discern. Lycan has been quoted (Gennaro 2011: 29) as wondering whether the HO approach is "nearly trivially true." Rosenthal describes TRANSITIVITY as "pre-theoretic" datum (1997a: 156) and "intuitively obvious" (2000: 273). Kriegel (2003a: 106) says that "there is something artificial" in saying that a mental state can be conscious even if we are wholly unaware of that mental state. And Pessi Lyrra (2009: 68) remarks that TRANSITIVITY is treated as a *fundamentatum* for further theorizing.

What these considerations show, I venture, is that as HO theorists, we should reject the assumption that FO theorists mean what we mean by NAGELIAN TRANSITIVITY and other 'what it is like'-sentences, and instead look for a more charitable interpretation.

The more charitable interpretation is that FO theorists speak FO-English. The case for this is admittedly less straightforward. When speakers use an expression, they do so with beliefs about its meaning, or so I'll assume. With HO theorists, we were in luck, since they were fairly transparent with respect to what they take 'what it is like'-sentences to mean. Some HO theorist all but tell us they speak HO-English. Others give us arguments that are hard to make sense of unless they speak HO-English. That's not the case with FO theorists, although there are some exceptions. For instance, Jesse Prinz tell us that "phenomenal

consciousness . . . refers to mental states that feel like something" (2012: 4). That's an appeal to a crucial element of the Affective View—that of feeling some way—as essential to being in a conscious state.

Block gestures towards another aspect of the Affective View. Consider how he explains why (the extinction subject) GK can have conscious states that are still 'for him':

> [GK] genuinely has [a] face experience that he does not and cannot know about. Wait—is that a real possibility? What would make it *his* experience? The question about GK can be answered by thinking about the visual field. . . . If GK genuinely does experience the face on the left that he cannot report, then it is in his left visual field on the left side, and as such has relations to other items in his visual field some of which he will be able to report. The fact that it in his visual field shows that it is his experience. (2009: 1119)

There is no mention of affective relations or feelings here, but what Block says is nonetheless salient. We are told that GK has a conscious visual experience of a face. That means, by TNC, that there is something it is like for GK to see the face. Yet given that GK is *ex hypothesi* not aware of his seeing the face, how can there being anything it is like *for him*? The answer that Block suggests—that the face occurs in his visual field—is exactly what we would expect a FO theorist to say if she ascribed to the Affective View. For, according to the Affective View, all 'for the subject' does, as an argument of the verb phrase 'is like', is draw out the fact that there is a subject who is standing in an affective relation to a mental state—i.e., *for whom* there is something it is like. And that, in effect, is what Block is saying; GK's experience is 'for him' because the face occurs in *his* visual field. That's simply not enough on the Operator View. On the Operator View, GK's experience is not 'for him' at all.

These cases suggest that at least some FO theorists might speak FO-English. Can we say something stronger? Here is how I see things. There is a metasemantic pressure to charitably interpret natural language. We saw that it is not charitable to interpret FO-theorists as speaking HO-English. The metasemantic pressure goes further, though. The point isn't simply to interpret FO theorists so as to make their denial of NAGELIAN TRANSITIVITY (and TRANSITIVITY) reasonable—presumably there are a whole host of semantical accounts of 'what it is like'-sentences, however independently implausible, that would do the trick. We can allow (as noted earlier) that FO theorists (or HO theorists) are wrong about what 'what it is like'-sentences mean, but there is also a pressure to not interpret them as taking 'what it is like'-sentences to mean something they manifestly do not mean. That's why FO-English seems so plausible. It preserves a core notion many associate with (phenomenal) consciousness: affective relations.

But moreover, other options are lacking. Suppose we interpreted the FO theorist as speaking a language in which 'what it is like'-sentences have their meaning in accordance with what Stoljar (2016: 1184) calls *the comparative account*. In this language, CANONICAL is true iff there is some y such that Michael's ϕ'ing x resembles y. On the face of it, interpreting FO theorists in this way would be odd as resemblance (or similarity) doesn't seem like it has much to do with consciousness, or 'what it is like'-talk.[21] It would in addition commit the FO theorist to absurdities: by the comparative account, Michael would know what ϕ'ing x is like simply by knowing what ϕ'ing x resembles. But that's false. I likely won't be able to know what it's like to eat a clementine unless I have actually eaten a clementine. Yet I might still know (perhaps because I consulted an expert on the citrus species) that eating a clementine resembles eating a cross between a mandarin orange and a sweet orange (Lewis 1988; cf. Stoljar 2016: 1185). Or suppose we treat the FO theorist as departing from ordinary meaning, and instead making a technical stipulation about their use of 'what it is like'-sentences. The problem with this option is that there is no evidence for it. For presumably if FO theorists *were* employing 'what it is like'-sentences in a technical sense, they would tell us (cf. Footnote 12). But they don't.

Therefore, I think there is good cause to interpret FO theorists as speaking FO-English.[22] The next question, though, is whether the HO theorist would agree that NAGELIAN TRANSITIVITY (and thus TRANSITIVITY) is false in FO-English—that is, whether she would agree that one can feel some way in virtue of being in a mental state without one being aware of that mental state. I think there is evidence that she would.

3.4. Disputed and Undisputed Sentences

Let's begin by extracting from FO-English and HO-English two undisputed sentences U1 and U2, one true and one false, such that the FO theorist holds that U1 is equivalent to TRANSITIVITY and the HO theorist holds that U2 is equivalent to TRANSITIVITY. Let U1 be 'A mental state will be such that its subject feels some way in virtue of being in it only if its subject is in some way aware of it.' Let U2 be 'A mental state seems some way to its subject only if its subject is in some way aware of that mental state.'

I'll focus my efforts on this claim: HO theorists think that U1 is false. So I'll be

21. Remember: with Stoljar, we are in the first instance not interested in the linguistic meaning of 'what it is like'-sentences, but what propositions 'what it is like'-sentences stereotypically express. (See Footnote 14).

22. Different FO theorists will of course give different accounts of the nature of the affective relations described by 'what it is like'-sentences in FO-English. But that's perfectly consistent with the present point.

setting aside two points: (i) that FO theorists also think that U1 is false, and (ii) that both HO and FO theorists think that U2 is true. These are not problematic shortcuts. On (i), if the FO theorist is going to ascribe to the Affective View at all, and so claim that a mental state M of a subject S is conscious if and only if S feels some way in virtue of being M, she will grant that we need not be thereby aware of M. So we can assume that the FO theorist will agree that U1 is false. Likewise on (ii): we know that the HO theorist thinks that U2 is true, and given the considerations adduced above, it is hard to see how the FO theorist could claim otherwise.[23] Also, to make things easier, I'll take Rosenthal (one of our usual suspects) as my main illustrative example, although other HO theorists will also crop up. Yes, other HO theorists differ from Rosenthal *qua* HO theorist, but what's at issue is whether these are relevant differences—that is, whether these differences will make it the case that U1 is disputed rather than undisputed. I doubt they will be, since these differences primarily concern whether or not the higher-order awareness is perceptual, and whether or not there is a constitutive relationship between the state of higher-order awareness and its lower-order target. Neither point of contention is relevant here.[24]

To see why U1 is false on HO theory, we can start by recalling Rosenthal's discussion of Block's stance on visual extinction. Rosenthal, we saw, claimed that the experiences extinction subjects have with respect to the impaired portion of their visual field are conscious to the same extent tables are conscious—which is to say not conscious at all. Such experiences are, like tables, *like something*, insofar as they have various features and characteristics, but they are not, as TNC demands, like something *for* their subjects. Yet shortly thereafter we find Rosenthal striking a more conciliatory attitude, distinguishing between two kinds of 'phenomenality:'

> One . . . consists in the subjective occurrence of mental qualities, while the other kind consists just in the occurrence of qualitative character without there also being anything it's like for one to have that qualitative character. Let's call the first kind *thick phenomenality* and the second *thin phenomenality*. Thick phenomenality is just thin phenomenality together with there being something it's like for one to have that thin phenomenality. (2002a: 657)

23. What Rosenthal calls 'conscious mental states', Block (2001) calls conscious mental states with 'reflexivity,' and he agrees that reflexivity is just higher-order awareness.

24. I tend to think that the difference between self-representationalists (those who *do* think there is a constitutive relationship between the higher-order awareness and the lower-order target) and traditional HO theorists (those who *don't* think this) is not as significant as it is sometimes supposed. Kriegel (2009) says that conscious states are self-representational, but these states are actually comprised of two distinct vehicles. What motivates Kriegel to deny that there is a state for each vehicle—as would be the case on traditional HO theories—is just that the vehicles are bound together as a complex by some "psychologically real" relation.

What are we to make of this? Here are three things we know. First, 'phenomenality' is actually Block's term: "What is phenomenality? What it is like to have an experience. When you enjoy the taste of wine, you are enjoying gustatory phenomenality" (2001: 202). Second, Rosenthal wants to demote the phenomenality our mental states have in the absence of HO awareness to thin phenomenality. Third, Rosenthal denies that states with mere thin phenomenality are like anything for their subjects. This shouldn't come as a surprise. For given Rosenthal's adherence to the Operator View, a mental state will be like something for its subject only if that mental state seems some way. And a mental state will seem some way to its subject only if that subject is aware of that mental state—precisely what is missing in states with thin phenomenality alone.

So what can we conclude? More specifically, what is it for a mental state to have thin, as opposed to thick, phenomenality? Well, if 'phenomenality' is to have any purchase here whatsoever, it is plausible to think that what Rosenthal means is simply that while we can feel some way in virtue of being in states with thin phenomenality, those states are not truly conscious unless we are also aware of them. Yet then Rosenthal would certainly grant that U1 is false: we can feel some way in virtue of being in a mental state without our being aware of that mental state. So U1, it would seem, is indeed uncontested.

Consider too what Rosenthal says about pain:

> Bodily sensations such as pains can also occur without being conscious. For example, we often have a headache or other pain throughout an extended period even when distractions intermittently make us wholly unaware of the pain. (2002b: 411)

Pains are a type of feeling. In having a pain, one *ipso facto* feels some way. Given that "pains can . . . occur without being conscious," Rosenthal would indeed likely claim that U1 is false. Hence, once again, U1 appears to be undisputed.

I see two objections that might arise here. To say that Rosenthal would claim that U1 is false is not merely to say that there can be (say) unconscious pains, but that there can be unconscious *feelings*—for Rosenthal, feelings that I am in no way aware of having. The first objection suggests an alternate possibility: whatever Rosenthal intends to pick by 'pain', it is not a feeling, or a state with a *feel*. Perhaps it is just a mere first-order representation of tissue damage.

The second objection assumes that Rosenthal recognizes unconscious feelings, but points out that this does not mean that he would agree that the *subject* feels some way in virtue of being in such states. And this, the objection continues, is precisely what U1 demands; by the Affective View, to say that U1 is false is to say that the *subject* can feel some way in virtue of being in a mental state, despite being unaware of that mental state. Anything less won't cut it for U1.

I'll broach both worries in tandem. We might note right off the bat that whatever else we want to say about the supposed implausibility of our having feelings, and thus feeling some way, despite not being aware of that feeling, this is not a mantle Rosenthal needs to bear alone. For we already know that FO theorists must recognize this possibility. But bracket that. Ditch 'pain' altogether and focus on 'feeling' instead. Clearly there is a sense of the term 'feeling' where *we*—that is, subjects of experience—can feel some way and not be aware that we feel that way. Consider emotional feelings. I can be depressed or angry—I can *feel* depressed or *feel* angry—and not know that I am feeling these ways until a friend points out my aberrant behavior. Indeed, HO theorists would concur; discussion of such cases is common in the literature on HO theories, meant to illustrate the requirement that the requisite higher-order awareness cannot come about by conscious inference (e.g., Rosenthal 1997a). These attributions are *person-level* attributions; when your friend comes to you, she does not merely say that there is some sub-personal feeling, but rather that you feel some way. She does so because it is our feeling a certain way that explains our behavior. The HO theorist just says that we are not aware of the mental state in virtue of which we feel some way.

Thus there is nothing about the notion of 'feeling' as such that would make U1 contested. This suggests an argument from parity: if we are willing to make person-level attributions of emotions (a type of feeling) that we are unaware of, then we should also make person-level attributions of pains (*qua* feelings) of which we are unaware.

There is further evidence beyond pain states. Consider what Rosenthal says about perception:

> It is beyond dispute that perceiving occurs not only consciously, but without being conscious as well, in masked priming and other forms of subliminal perceiving. . . . Absent special assumptions about how to taxonomize mental qualities, the natural and indeed inevitable conclusion would be that, apart from the property of being conscious itself, whatever mental properties occur in conscious perceiving also occur in the subliminal, non-conscious case. (2015: 36)

Suppose I look at a pinkish sunset. I am in a visual state with a pinkish quality. Call this quality *Q*. In the present context, to say U1 is false is to say that I can instantiate *Q* (or be in state with *Q*), and thus feel some way (the way I feel when in *Q*-states), despite my not being aware of *Q* (or my being in a state with *Q*).

Now Rosenthal will not say that *there is something it is like* to (merely) be in a state with *Q* because of what he thinks this phrase means. However, that does not mean he would deny that we feel some way in virtue of being in a state

that has *Q*. The idea is that apart from being conscious itself, whatever mental properties occur in conscious perceiving also occur in the non-conscious case. *Q* is a sensory property. Taking this point in conjunction with Rosenthal's earlier concession that a state with *Q* still has thin phenomenality, a natural conclusion to draw is that we feel some way in virtue of being in a state with *Q*. True, the state with *Q* won't seem any way to its subject, and so for Rosenthal won't be conscious, until it's targeted by a suitable state of higher-order awareness. Nonetheless, the state with *Q* still has a pinkish quality, and it still has thin phenomenality, and it has all these things absent being suitably targeted. But if a pinkish quality has *any* phenomenality, yet we deny that that state with that quality seems any way, then it stands to reason that we must at least feel some way in virtue of being in that state. For otherwise talk of being in a state with pinkish quality and thin phenomenality loses all purchase.[25]

Kriegel (2003b) says something similar about *absent-minded perception*, as when one takes a long-distance drive and 'zones out' about the details on the road. Kriegel (2003b: 4) claims that not only does the absent-minded subject have unconscious perceptual states that represent the road, but that these states also have *qualia*—qualities like *Q*.[26] It's not unreasonable to interpret this in the same manner: what Kriegel must mean is that the absent-minded subject feels some way in virtue of being in such absent-minded perceptual states. And these states are absent-minded precisely because the subject is unaware of them. Only when they are not absent-minded—only when we are aware of them—are they conscious in the HO-theoretic sense.

Taken together, then, these considerations suggest the HO theorist would agree that U1 is false.

What if we switched to the disjunctive version of The Affective View, and focused on its second disjunct? Would that change things? Here, U1 becomes U1*: 'A mental state of which one is wholly unaware cannot be such that things seem some way in virtue of being in it.' Would the HO theorist say that U1* is false too? There is reason to think she would. Consider semantic priming in the lexical decision task (e.g., Dehaene et al. 1998). Subjects are presented (usually visually, but auditorily as well) with a mix of words and pseudo-words as targets preceded by non-conscious primes, which can either be semantically related (like *wolf* and *dog*) or unrelated (like *lawyer* and *turkey*) to the target words. That

25. In the background here is Rosenthal's Sellarsian *quality space theory* of perception. Unlike *consciousness-based* theories, where consciousness is the only way we can know about mental qualities, the quality space theory is a *perceptual-role theory*. On this view, mental qualities figure into perceptual discriminations, which may or may not be made consciously (Rosenthal 2010: 274). The upshot for our purposes is that *the very same qualities* that occur when we *are* aware of our mental states—qualities like *Q*—occur when we are *not* aware of our mental states.

26. Kriegel (2003b: 2) uses 'qualia' interchangeably with Rosenthal's 'sensory qualities' and 'mental qualities.'

subjects are faster to respond to semantically related prime-target pairs relative to unrelated pairs suggests that a stimulus can be categorized and conceptualized without our being aware of seeing that stimulus. But if this is so, then it's plausible that a stimulus can seem some way to us without our being aware of our perceiving that stimulus. HO theorists (e.g., Weisberg 2008: 173) themselves concur on this score. More generally, when we speak of seeing something *as* red (or *as* a turkey), and thus conceptualize that thing as red, that red thing must seem to us a certain a certain way—viz., the way red things typically seem. So, U1*, I venture, should be uncontested as well.

Before we press on, there is a final objection worth considering. It might be wondered whether we can frame the dispute between HO and FO theorists without TNC and 'what it is like'-talk. Perhaps once understood in an alternate manner, the dispute will no longer verbal. This objection stems from Chalmers's (2011) *method of elimination*, whereby one eliminates use of the contested term, and then attempts to determine whether any substantive dispute remains. But here's the thing: it's not clear whether there *is* some other way to frame the dispute. Suppose both sides could drop TNC pro tem, and instead adjudicate things in terms of what each side has to say about the scope of animal consciousness. That might seem promising since the truth of a HO theory is often taken to imply that there are fewer conscious animals than there would be if a FO theory were true.[27] Surely that issue is non-verbal.

This won't cut it. The reason is that FO theories and HO theories *don't* obviously make different predictions about which animals are conscious. It is true that some HO theorists (e.g., Carruthers 2000) are (in)famous for biting the bullet, saying that many, if not all, non-human animals are unconscious. But there is nothing about the HO-theoretic picture *as such* that says that any HO theory, simply by virtue of being an HO theory, will admit fewer conscious animals than any FO theory. This is because a HO theory is simply an implementation of TRANSITIVITY, and there is nothing obviously problematic about saying that (non-human) animals are aware of their conscious states. Where this does become a potential issue is when we make more specific claims about the nature of that HO awareness. For example, the HOT theory requires that non-human animals instantiate higher-order thoughts to be conscious, which in turn plausibly requires some measure of conceptual sophistication. Yet even here there is an open question about just how much conceptual sophistication is required. Maybe it's not much at all—so little, perhaps, that pretty much all of the animals we pre-theoretically believe to be conscious in fact are (Gennaro 2011). And other HO theories—like Hellie's (2007b) HO acquaintance theory, or Kriegel's (2009)

27. I thank an anonymous referee and Joel Velasco for pressing this specific way of adjudicating the dispute.

self-representational theory—don't clearly predict that there are fewer conscious creatures than the FO theorist would have us believe. My point: the issue at stake between HO theorists and FO theorists is, in the first instance, the truth of TRANSITIVITY, not the truth of this or that implementation of TRANSITIVITY. Only when we broach the latter question does a potential difference in the scope of animal consciousness arise.[28]

Another option is to frame the dispute in terms of the explanatory gap or experience. Kriegel seems to go in for the former, rigidifying 'phenomenal consciousness' as that property F, such that, in the actual world, F is (casually) 'responsible for the mystery of consciousness' (2009: 3). The problem here is that there is no clean split between a position on (say) the explanatory gap, and a position on the HO versus FO divide. For instance, Rosenthal (2010) rejects the intuitions behind spectrum inversion, Jackson's Mary, the explanatory gap, and the conceivability of zombies. But Kriegel (2009), another HO theorist, does not. On the latter, even the term 'experience' is ambiguous, and as Byrne (2004) points out, it's not clear that FO and HO theorists are using it in the same way—compare, e.g., Tye's usage (2000) with Carruther's (2000).

In sum, there are better and worse ways of framing the debate. I submit that the best way to do so is in terms of TNC. And that's so even if it turns out that the dispute is verbal.[29]

4. 'Conscious' and Metalinguistic Negotiation

I have provided a model of the dispute between HO and FO theorists according to which HO and FO theorists are best interpreted as speaking two languages:

28. What about adjudication in terms of neural realization? It is occasionally said that HO theories implicate the prefrontal cortex (PFC) as essential to consciousness (e.g., Lau & Rosenthal 2011), but this is controversial. For example, on Gennarro 'Wide Intrinsicality' HO theory (2011: 278), only *introspection* requires PFC activity. Gennarro claims that it is posterior areas (e.g., inferior parietal cortex) that are essential to consciousness proper, which is exactly what Prinz says (2012: 98) on his FO 'AIR' theory.

29. Take note of how the method of elimination works. After banning the contentious term T found in the originally disputed sentence *s*, you get an alternative sentence *s** in a newly restricted vocabulary (i.e. a vocabulary without T). The dispute over *s** must be *part of* the dispute over *s*. Chalmers defines 'part of' like this: "a dispute over *s** is part of a dispute over *s* when: (i) if the parties were to agree that *s** is true, they would (if reasonable) agree that *s* is true, and (ii) if they were to agree that *s** is false, they would (if reasonable) agree that *s* is false" (2011: 527). Now suppose we were to replace 'John's visual state is conscious' with 'John's visual state is the sort of state that seems like it cannot be reduced to a physical state,' such that 'conscious' is banned. The claim is this: there are participants in the HO-FO debate that would reasonably affirm (or deny) the former while denying (or affirming) the latter. That's evidence that disputes over the latter are not part of disputes over the former, and I think the same point can plausibly be made for most other ways of substituting out the term 'conscious.'

HO-English, where TRANSITIVITY is true, and FO-English, where TRANSITIVITY is false. What makes this model successful is its ability to provide the best rationalizing explanation for the ways in which the relevant parties contribute to the dispute. And what charity tells us, I've suggested, is that the parties' contributions are most reasonably explained as consistent responses to two different questions. In other words, the dispute between HO and FO theorists is verbal.

My thesis is radical. If true, it rewrites the default and predominant picture of a central debate in discussions about consciousness. This on its own will cause some unease. So, by way of closing, I want to quell at least one source of that apprehension.

When we say a dispute is verbal, we are saying that disputants are in a sense not really disagreeing, or that they are just disagreeing about language. Of course, this is far from the self-conception of HO and FO theorists. Even if claiming that HO and FO theorists are engaging in a verbal dispute enables us to see them as uttering truths, and so charitably understanding what they are *saying*, it doesn't give us a way to charitably understand what they are *doing* in partaking in these disputes. Thus we might ask, if it is true that HO and FO theorists are expressing compatible propositions at the level of literal content, what have HO and FO theorists been doing all of these years that makes their dispute seem so deep and important, i.e., a genuine, substantive disagreement? Or to put the question slightly differently, how could a collection of smart, careful philosophers end up in a verbal dispute without realizing it?[30]

This question is natural. It is liable to come up whenever a long-standing, seemingly deep dispute is said to be verbal. If we say, with Hirsch, that the debate over composition is verbal, we'll want to know how this could have escaped the notice of universalists and nihilists who are no doubt diligent and attentive. But it's also not clear how much force the question really carries. The charge on the table is that there is something fundamentally awry in the dispute between HO and FO theorists. As such, we shouldn't expect a perfectly tidy picture, or anything resembling the preservation of the participants' entire self-conception.[31]

All the same, I recognize that the current proposal would be strengthened if there was a story to tell. So here is a tentative one: HO and FO theorists are engaging in a *metalinguistic negotiation*. A metalinguistic usage of a term consists in a speaker using a term to express a view about the meaning of that term, or to express a normative view about how that term ought to be used (Plunkett 2015: 834). It is well-known that we don't always express our disagreements via literal

30. I thank an anonymous editor at *Ergo* and Christopher Hom for pressing me to clarify this issue.

31. See Thomasson (2017) for this point when it comes to disputes about physical object ontology.

semantic content. Genuine disagreements can also be expressed in a language via pragmatics. A metalinguistic dispute is a dispute that centers on pragmatically communicated information involving speakers who employ a metalinguistic usage of a term. A metalinguistic negotiation is just a normative metalinguistic dispute. It is a dispute wherein the speakers' metalinguistic use of a term does not, as Plunkett and Sundell put it, "simply involve exchanging factual information about language, but rather negotiating its appropriate use" (2013: 15).[32]

An example. Suppose two guests on a sports radio talk-show are debating whether Secretariat, the famous racehorse, is an athlete. The guests concur on Secretariat's prowess as a racehorse. They know he won the Triple Crown in 1973, and they know that he set speed records at all three races, and that his time at the Kentucky Derby still stands over forty years later. Still, one guest endorses the sentence 'Secretariat is an athlete,' the other guest endorses its negation. Now, upon hearing this, one might immediately shout 'Verbal Dispute!' Yet even if the dispute is verbal at the *object-level*, one can nonetheless preserve the intuition that there is a substantive disagreement by ascending to the metalinguistic level. The guests on the radio talk-show are arguing about how 'athlete' *ought* to be used.[33]

This is a dispute worth having. How we use words often matters (Chalmers 2011: 516; cf. Thomasson 2017). 'Athlete' is like 'marriage' or 'torture.' Thomasson (2017) observes that when we ask whether a non-human animal like Secretariat can be legitimately treated as an athlete, we are in an important sense negotiating how they are viewed and treated given the status *being an athlete* confers in our society. And whether non-human animals ought to be viewed and treated in this manner is well worth arguing about. As Thomasson (2017: 13) notes, it is for this same reason that metalinguistic disputes are not just 'about words'. Far from it: although speakers use these utterances (e.g., 'Secretariat is an athlete', 'Waterboarding is not torture') to reinforce or alter the norms for using the terms in question, what's at stake when we negotiate the use of a term like 'athlete' or 'torture' is something that is very much about the world and the people in it (cf. Thomasson 2017: 12; Plunkett & Sundell 2013: 19).

So, to say that HO and FO theorists are engaged in metalinguistic dispute is to say that they are advocating for what they think 'consciousness' *should* mean, or how it *should* be used. We have a verbal dispute at the object-level, but a genuine substantive disagreement at the metalinguistic level. And such substantive metalinguistic negotiations can be easily confused for substantive object-level

32. My discussion of metalinguistic disputes and metalinguistic negotiation here and in the rest of this section is an extension of, and thus indebted to, the work of Delia Belleri, David Plunkett, Timothy Sundell, and Amie Thomasson.

33. The Secretariat case originally comes from Ludlow (2008), but has since been discussed heavily.

disputes. Think about the Secretariat case; when we argue over what counts as an athlete, this dispute does not seem verbal. It seems deep. The reason it seems deep is because it is. It's just not deep at the level we think it's deep at; it's not deep at the object-level. Hence diligent and attentive participants to the debate don't realize that their (object-level) dispute is verbal.

But is this plausible? Is the debate about TRANSITIVITY relevantly similar to the debate about Secretariat? I'm not sure. An immediate challenge is that theorists on both the HO and FO sides might just deny that they see themselves as pragmatically communicating something normative about our linguistic or conceptual scheme. The strength of this worry will depend to some extent on how we model metalinguistic negotiation, and here is not the place to engage that issue.[34] This point notwithstanding, I do think there is a case to be made. Plunkett (2015: 847) identifies a set of criteria that provide defeasible evidence that a philosophical dispute is a metalinguistic negotiation. These criteria are:

(a) There being evidence that a dispute is occurring (i.e., an exchange that looks like a genuine disagreement);
(b) There being evidence that there indeed is a genuine disagreement, at least somewhere;
(c) There being evidence that the disputants are using the same key terms in different ways;
(d) There being evidence that the disagreement is normative (i.e., not descriptive).

On (a), there is no doubt, of course, that HO and FO theorists are at least having a dispute. The relevant parties enter into various linguistic exchanges wherein they appear to express a genuine disagreement over the truth-value of TRANSITIVITY. We'll grant (b) for sake of argument. With respect to (c), we saw that it is grossly uncharitable to treat HO and FO theorists as meaning the same thing by TNC. HO theorists either explicitly endorse the Operator View, or implicitly by way of the Emphatic Argument. They speak HO-English. But it uncharitable to interpret the FO theorist as also speaking HO-English. For if she did, she would be making a painfully obvious a priori error every time she endorsed ¬TRANSITIVITY.

Regarding (d), that the disagreement isn't about what the terms actually mean helps us make sense of Block and Weisberg's barbs (Section 2) that each other's use of 'consciousness' is technical. Sure, you are in pain when you are just in the first-order state, but you are not (*ex hypothesis*) aware of it. But—Weisberg

34. Conversational implicature is an obvious option for understanding what is done pragmatically in metalinguistic negotiation, but it isn't the only option, nor is it plainly the best option. For discussion, see Thomasson (2017).

might say—who cares? He can (and will) agree that the subject feels some when they are in that first-order state, but perhaps Weisberg's point is that 'conscious' should be used to describe states we *care about*—i.e., states we aware of. We don't care about pains that we are not aware of, at least directly. Block (2011: 426) invokes a similar move when he says that the 'what-it-is-likeness' that truly matters isn't present in cases of empty HOTs, since in such cases no first-order state is present. To appeal to what we care about, as both Weisberg and Block do, in judging whether a mental state is conscious, is to advocate on pragmatic grounds for a certain conceptual scheme. Indeed, even if Stoljar is right that FO-English is English, having realized this, I doubt Weisberg (or Rosenthal, or any HO theorist) would let matters rest. They would, I bet, continue to negotiate for how 'consciousness' ought to be used. That's a tell-tale sign of (d).

All of this suggests that there may be something to the idea that HO and FO theorists are engaged in a metalinguistic negotiation. Settling the matter, however, is a problem for another day.

Acknowledgements

For comments on an ancestor version of this paper, I thank audiences at the Australian National University and the University of Illinois at Chicago, and for comments on a more recent version I thank an audience at Texas Tech University. For especially helpful comments, written and otherwise, I thank Mahrad Almotahari, David Chalmers, Jonathan Farrell, Christopher Hom, Saja Parvizian, David Rosenthal, Jeff Speaks, Joel Velasco, and two anonymous referees. Extra special thanks go to Dave Hilbert and Daniel Stoljar: Dave, for wise council on this project from its infancy, and Daniel, for providing the semantic framework that made the project possible, and for encouragement throughout.

References

Armstrong, David (1968). *A Materialist Theory of Mind*. Routledge.

Belleri, Delia (2017). Verbalism and Metalinguistic Negotiation in Ontological Disputes. *Philosophical Studies*, *174*(9), 2211–2226. https://doi.org/10.1007/s11098-016-0795-z

Bennett, Karen (2009). Composition, Co-Location, and Metaontology. In David Chalmers, David Manley, and Ryan Wasserman (Eds.), *Metametaphysics: New Essays on the Foundations of Ontology* (38–76). Oxford University Press.

Block, Ned (1996). Mental Paint and Mental Latex. *Philosophical Issues*, *7*, 19–49. https://doi.org/10.2307/1522889

Block, Ned (2001). Paradox and Cross-Purposes in Recent Work on Consciousness. *Cog-*

nition, *79*(1–2), 197–219. https://doi.org/10.1016/S0010-0277(00)00129-3
Block, Ned (2002). Concepts of Consciousness. In David Chalmers (Ed.), *Philosophy of Mind: Classical and Contemporary Readings* (206-219). Oxford University Press.
Block, Ned (2007). Consciousness, Accessibility, and the Mesh between Psychology and Neuroscience. *Behavioral and Brain Sciences*, *30*(5), 481–548. https://doi.org/10.1017/S0140525X07002786
Block, Ned (2008). Consciousness and Cognitive Access. *Proceedings of the Aristotelian Society*, 108(1 pt. 3), 289–317. https://doi.org/10.1111/j.1467-9264.2008.00247.x
Block, Ned (2009). Comparing the Major Theories of Consciousness. In Michael Gazzaniga (Ed.), *The Cognitive Neurosciences* (4th ed., 1111–1123). MIT Press.
Block, Ned (2011). The Higher-Order Approach to Consciousness is Defunct. *Analysis*, *71*(3), 419–431. https://doi.org/10.1093/analys/anr037
Brogaard, Berit (2012). What Do We Say When We Say How or What We Feel? *Philosophers' Imprint*, *12*(11), 1–22.
Byrne, Alex (2004). What Phenomenal Consciousness Is Like. In Rocco J. Gennaro (Ed.), *Higher-Order Theories of Consciousness: An Anthology* (203–226). John Benjamins. https://doi.org/10.1075/aicr.56.12byr
Carruthers, Peter (2000). *Phenomenal Consciousness: A Naturalistic Theory*. Cambridge University Press. https://doi.org/10.1017/CBO9780511487491
Chalmers, David (2011). Verbal Disputes. *Philosophical Review*, *120*(4), 515–566. https://doi.org/10.1215/00318108-1334478
Chisholm, Roderick (1957). *Perceiving: A Philosophical Study*. Cornell University Press.
Dehaene, Stanislas, L. Naccache, G. Le Clec'H, E. Koechlin, M. Mueller, G. Dehaene-Lambertz, . . . and D. Le Bihan. (1998). Imaging Unconscious Semantic Priming. *Nature*, 395(6702), 597–600. https://doi.org/10.1038/26967
Dretske, Fred (1995). *Naturalizing the Mind*. MIT Press.
Dretske, Fred (1999). The Mind's Awareness of Itself. *Philosophical Studies*, *95*(1), 103–124. https://doi.org/10.1023/A:1004515508042
Dretske, Fred (2006). Perception without Awareness. In Tamar S. Gendler and John Hawthorne (Eds.), *Perceptual Experience* (147–180). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199289769.003.0005
Farrell, B. A. (1950). Experience. *Mind*, *59*(234), 170–198. https://doi.org/10.1093/mind/LIX.234.170
Farrell, Jonathan (2016). 'What It Is Like' Talk Is Not Technical Talk. *Journal of Consciousness Studies*, *23*(9–10), 50–65.
Gennaro, Rocco (2011). *The Consciousness Paradox: Consciousness, Concepts, and Higher-Order Thoughts*. MIT Press. https://doi.org/10.7551/mitpress/9780262016605.001.0001
Gottlieb, Joseph (2016). Transitivity and Transparency. *Analytic Philosophy*, *57*(4), 353–379. https://doi.org/10.1111/phib.12091
Hacker, Peter M. S. (2002). Is There Anything it Is Like to Be a Bat? *Philosophy*, *77*(2), 157–174. https://doi.org/10.1017/S0031819102000220
Hellie, Benj (2007a). 'There's Something It's Like' and the Structure of Consciousness. *Philosophical Review*, *116*(3), 441–463. https://doi.org/10.1215/00318108-2007-005
Hellie, Benj (2007b). Higher-Order Intentionalism and Higher-Order Acquaintance. *Philosophical Studies*, *134*(3), 289–324. https://doi.org/10.1007/s11098-005-0241-0
Hirsch, Eli (2011). *Quantifier Variance and Realism: Essays in Metaontology*. Oxford University Press. https://doi.org/10.1093/acprof:osobl/9780199732111.001.0001
Jackson, Brendan Balcerak (2013). Metaphysics, Verbal Disputes and the Limits of Char-

ity. *Philosophy and Phenomenological Research*, *86*(2), 412–434. https://doi.org/10.1111/j.1933-1592.2011.00569.x

Janzen, Greg (2011). In Defense of the What-it-Is-Likeness of Experience. *The Southern Journal of Philosophy*, *49*(3), 271–293. https://doi.org/10.1111/j.2041-6962.2011.00074.x

Kriegel, Uriah (2003a). Intransitive Self-Consciousness: Two Views and an Argument. *The Canadian Journal of Philosophy*, *33*(1), 103–132. https://doi.org/10.1080/00455091.2003.10716537

Kriegel, Uriah (2003b). Consciousness as Sensory Quality and as Implicit Self-Awareness. *Phenomenology and the Cognitive Sciences*, *2*(1), 1–26. https://doi.org/10.1023/A:1022912206810

Kriegel, Uriah (2006). Theories of Consciousness. *Philosophy Compass*, *1*(1), 58–64. https://doi.org/10.1111/j.1747-9991.2006.00008.x

Kriegel, Uriah (2009). *Subjective Consciousness: A Self-Representational Approach*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199570355.001.0001

Lau, Hakwan and David Rosenthal (2011). Empirical Support for Higher-Order Theories of Conscious Awareness. *Trends in Cognitive Sciences*, *15*(8), 365–373. https://doi.org/10.1016/j.tics.2011.05.009

Levine, Joseph (2006). Conscious Awareness and (Self)-Representation. In Kenneth Williford and Uriah Kriegel (Eds.), *Self-Representational Approaches to Consciousness* (173–198). MIT Press.

Levine, Joseph (2007). Two Kinds of Access. *Behavioral and Brain Sciences*, *30*(5–6), 514–515. https://doi.org/10.1017/S0140525X07002956

Lewis, David (1988). What Experience Teaches. *Proceedings of the Russellian Society*, *13*, 29–57.

Lewis, David (1995). Should a Materialist Believe in Qualia? *Australasian Journal of Philosophy*, *73*(1), 140–144. https://doi.org/10.1080/00048409512346451

Lormand, Eric (2004). The Explanatory Stopgap. *Philosophical Review*, *113*(3), 303–357. https://doi.org/10.1215/00318108-113-3-303

Ludlow, Peter (2008). Cheap Contextualism. *Philosophical Issues*, *18*(1), 104–129. https://doi.org/10.1111/j.1533-6077.2008.00140.x

Lycan, William (1996). *Conscious Experience*. Oxford University Press.

Lycan, William (2001). A Simple Argument for a Higher-Order Representation Theory of Consciousness. *Analysis*, *61*(269), 3–4. https://doi.org/10.1093/analys/61.1.3

Lyyra, Pessi (2009). Two Senses for 'Givenness of Consciousness.' *Phenomenology and Cognitive Science*, *8*(1), 67–87. https://doi.org/10.1007/s11097-008-9110-6

Nagel, Thomas (1974). What Is it Like to Be a bat? *Philosophical Review*, *83*(3), 435–450. https://doi.org/10.2307/2183914

Martin, Michael G.F. (2004). The Limits of Self-Awareness. *Philosophical Studies*, *120*(13), 37–89. https://doi.org/10.1023/B:PHIL.0000033751.66949.97

Plunkett, David (2015). Which Concepts Should We Use?: Metalinguistic Negotiations and The Methodology of Philosophy. *Inquiry*, *58*(7–8), 828–874. https://doi.org/10.1080/0020174X.2015.1080184

Plunkett, David and Tim Sundell (2013). Disagreement and the Semantics of Normative and Evaluative Terms. *Philosopher's Imprint*, *13*(23), 1–37.

Prinz, Jesse (2012). *The Consciousness Brain*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195314595.001.0001

Rosenthal, David (1997a). A Theory of Consciousness. In Ned Block, Owen J. Flanagan,

and Guven Guzeldere (Eds.), *The Nature of Consciousness* (729–754). MIT Press.
Rosenthal, David (1997b). Phenomenal Consciousness and What It's Like. *Behavioral and Brain Sciences*, *20*(1), 156–157.
Rosenthal, David (2000). Consciousness and Metacognition. In Dan Sperber (Ed.), *Metarepresentations: A Multidisciplinary Perspective* (265–295). Oxford University Press.
Rosenthal, David (2002a). How many kinds of consciousness? *Consciousness and Cognition*, 11 (4), 653–665. https://doi.org/10.1016/S1053-8100(02)00017-X
Rosenthal, David (2002b). Explaining Consciousness. In David J. Chalmers (Ed.), *Philosophy of Mind: Classical and Contemporary Readings* (109–131). Oxford University Press.
Rosenthal, David (2005). *Consciousness and Mind*. Clarendon Press.
Rosenthal, David (2010). How to Think about Mental Qualities. *Philosophical Issues*, *20*(1), 368–393. https://doi.org/10.1111/j.1533-6077.2010.00190.x
Rosenthal, David (2011). Exaggerated Reports: A Reply to Block. *Analysis*, *71*(3), 431–437. https://doi.org/10.1093/analys/anr039
Rosenthal, David (2015). Quality Spaces and Sensory Modalities. In Paul Coates and Sam Coleman (Eds.), *The Nature of Phenomenal Qualities: Sense, Perception, and Consciousness* (33–65). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198712718.003.0002
Shoemaker, Sydney (1994). Phenomenal Character. *Noûs*, *28*(1), 21–38. https://doi.org/10.2307/2215918
Snowdon, Paul (2010). The What-it-Is-Like-Ness of Experience. *Southern Journal of Philosophy*, *48*(1), 8–27. https://doi.org/10.1111/j.2041-6962.2010.01003.x
Sprigge, Timothy (1971). Final Causes. *Proceedings of the Aristotelian Society*, Supp. Vol. *45*, 149–170. https://doi.org/10.1093/aristoteliansupp/45.1.149
Stoljar, Daniel (2016). The Semantics of 'What it Is Like'-Sentences and The Nature of Consciousness. *Mind*, *125*(500), 1161–1198. https://doi.org/10.1093/mind/fzv179
Thomasson, Amie (2017). Metaphysical Disputes and Metalinguistic Negotiation. *Analytic Philosophy*, *57*(4), 1–28.
Tye, Michael (1997). The Problem of Simple Minds: Is there Anything it Is Like to Be a Honey Bee? *Philosophical Studies*, *88*(3), 289–317. https://doi.org/10.1023/A:1004267709793
Tye, Michael (2000). *Consciousness, Color, and Content*. MIT Press.
Tye, Michael (2009). *Consciousness Revisited: Materialism without Phenomenal Concepts*. MIT Press.
Weisberg, Josh (2008). Same Old, Same Old: The Same-Order Representational Theory of Consciousness and the Division of Phenomenal Labor. *Synthese*, *160*(2), 161–181. https://doi.org/10.1007/s11229-006-9106-0
Weisberg, Josh (2011). Abusing the Notion of What-it's-Like-Ness: A Response to Block. *Analysis*, *71*(3), 443–448. https://doi.org/10.1093/analys/anr040
Wittgenstein, Ludwig (1974). *Tractatus Logico-Philosophicus* D. F. Pears and B. F. McGuinness, Trans.). Routledge.