THE IDOLS OF INNER-AWARENESS: TOWARDS DISJUNCTIVE SELF-REPRESENTATIOANLISM

Chad Kidd, Auburn University
chad.kidd@Auburn.edu
(draft for 5th Online Consciousness Conference)

1. *INTRODUCTION*

My goal in this paper is to garner support for the superiority of the Self-Representational theory of experience (henceforth, SR) over its most formidable opposition – the Higher-Order Representational theory of experience (HO). I do this by fortifying and then building on two (perhaps bold) claims about the nature of phenomenal consciousness. (i) that the form of awareness that one has of one's own mental state in having an experience – what I call "inner-awareness" – is a kind of self-directed intentionality of experience that is constitutive of and essential to it. And (ii) it is essential to the inner-awareness of experience that it guarantees the truth of occurrent corresponding introspective judgments about the experience. I hope to motivate the claim that these two claims follow closely on common-sense intuition concerning the epistemically privileged access we have to our own conscious mental lives. And then I hope to demonstrate the superiority of SR by arguing that it well-positioned to provide a positive account of these two claims, but HO is not.

Here is how the paper will proceed: Section 2 specifies the varieties of HO and SR with which I am concerned here. The distinguishing mark of this family of views is their acceptance of (i) – also known as the "transitivity principle," which yields has come to be called an "*extrinsic*" concept of phenomenal consciousness. Sections 3 and 4 clear ground needed to set up the argument for SR. Section 3 clarifies and defends the thesis that inner-awareness is *fallible*, and it introduces the tactic that I believe the SR theorist must use in order to accommodate the fallibility of inner-awareness: to have recourse to certain tactics characteristic of *disjunctive* theories of cognition. Section 4 attempts to motivate the disjunctive reconfiguring of SR by starting from the "non-grippingness" of a global skepticism about the reliability of introspective belief and connecting this to a certain analysis of the notion of the privileged accessibility of conscious experience. Section 5 presents the core argument of the paper against the adequacy of HO *vis-à-vis* the epistemic privilege of experience. Finally, section 6 further explicates the disjunctive SR proposal and answers some important objections to it.

2. *HO, SR, AND THE EXTRINSIC CONCEPT OF PHENOMENAL CONSCIOUSNESS*

According to HO, for a subject to have a phenomenally conscious experience the subject must have an appropriate 'higher-order' representation of her first-order ('external-world-directed') mental life. More specifically,

> **HO** A subject *S* at time *t* has an experience with phenomenal character φ iff *S* has a mental representation M^* of herself as being in a mental state *M* that is φ.

SR advances a thesis that is closely related to HO, but importantly different. On SR, a phenomenally conscious experience is created when the subject of the state has an appropriate form of higher-order representation of her first-order mental life, *and* the higher-order representation and the first-order stream of consciousness are somehow constitutively related to one another. More specifically,

> **SR** A subject *S* at time *t* has an experience with phenomenal character φ iff *S* has a mental representation *M** of herself as being in a mental state *M* that is φ, *S* in fact is in mental state *M*, *M* is φ, and *M** and *M* (aside from standing in the appropriate intentional relation) are *constitutively* related to one another.

Here and following, I use the word "experience" as elliptical for "*phenomenally conscious* experience." So experiences are, by definition, mental states for which there is "something it's like for the subject" to have them; that is, they are by definition "phenomenal." I will call those properties of experiences that characterize the "*what* it's likeness" for the subject in the experience – e.g., the qualitative blue of the experience of the sky on a clear, sunny day – the "phenomenal character" of the experience. I will use the term "mental state" to pick out mental entities that are, at least contingently, *not* experiences. And, analogous to my use of "phenomenal quality" and "phenomenal character," I will use the term "*subjective* quality" to pick out individual features of mental states that play an important role in the creation of the phenomenal character of experiential states, and "subjective character" to pick out the entire set of subjective qualities belonging to a mental state.

It is useful to note that both HO and SR take phenomenal qualities to be identical to subjective qualities of mental states, which are also represented (in an appropriate way) by higher-order intentional states. Thus, both HO and the versions of SR with which I am concerned accept the claim that a subject's mental state is made *phenomenally* conscious – i.e., made into an experience – by virtue of the special intentional connections that the state has with other mental states of the subject.

This has come to be known as an "*extrinsic*" concept of phenomenal consciousness (cf., Weisberg 2010): "extrinsic" because the key to unlocking the mysteries of phenomenal consciousness is taken be in the unravelling of the *relations* or relation-like features that are characteristic of experience. It is important to notice that both HO and SR conceive of the relevant 'relation' to be *intentional* in character. Not all extrinsic conceptions think of the key extrinsic characteristic of phenomenal consciousness in this way. Some, for instance, conceive of the essential connective elements of conscious experience as special causal relations of some sort. For my purposes here, however, I set the non-intentional family of extrinsic views aside.

HO and SR typically stipulate the following division of explanatory labor in theories of phenomenal consciousness:

> (i) the aspect that determines *what* is phenomenally experienced by the subject (e.g., the subjectively experienced blue in visually experiencing a clear daytime sky);
> (ii) the aspect that determines *whether* a subjective has a phenomenal experience at all (i.e., whether the subject is 'phenomenally conscious' or not).

On both HO and SR, certain intrinsic properties of a first-order mental state (e.g., the sensation of blue, of pain, or of anxiety) fulfill function (i); and higher-order intentional states with a certain kind of intentional content fulfill function (ii).[1] In the following, I will call the intentional states that fulfill function (ii) either states of "inner-awareness," states of "pre-reflective higher-order awareness," or states of "higher-order awareness *in* experience" – emphasizing the "*in*" to help us keep in mind that this is an awareness of experience that one has simply in "*having*" the experience (as opposed to an introspective or reflective state, which is not built into the experience it represents).[2]

As such, both HO and SR are set over against a set of theories that advance an "*intrinsic*" concepts of phenomenal consciousness. On these views, phenomenal properties of experience are thoroughly intrinsic to experiential states. So an experience is such by virtue of a constitutive "glow from within" (cf., Weisberg 2010: 2) or "a mental light, without which sensory qualities simply cannot exist" (Rosenthal 2005b: 135). Contrary to the extrinsic view's division of phenomenal labor, an intrinsic view takes the determination of the "what" and the "whether" of experience to be fulfilled by the same aspect of conscious states: the intrinsic phenomenal qualities of experience.

There are many reasons to be disappointed with intrinsic conceptions; but I do not have space to state these in any satisfying detail. Yet, it will help clarify aspects of HO and SR that will be important in the following discussion to consider two points on which the intrinsic conception is typically faulted. First, if the view even acknowledges the peculiar pre-reflective form of awareness we necessarily have of our own experience in its occurrence in us, it would seem to make "inner-awareness" into a form of intentionality that is completely primitive and inexplicable; not a very satisfying place for an *explanatory* theory of experience to begin. Second, it also presses phenomenal consciousness into the mold of a "self-revelatory" phenomenon: a self-intimating feature of experience that is invariably accompanied by an infallible awareness of it (again, if that concept is even applicable here); both of which are (as I will argue in the next section) features best avoided by philosophical theories of animal cognition.

Turning now to the differences between HO and SR, which will be the focus of much of the following: where SR understands there to be a *constitutive* relation between the higher-order and first-order state in an experience, HO sees no need for any constitutive relation. Rather, the two mental states that are *typically* present in the mental life of an experiencing subject, to use Hume's phrase, are 'distinct existences'. So, on HO, it is possible to have an experience at a time *t* that is constituted merely by the realization at *t* of an 'empty' higher-order mental state, i.e., a higher-order state of inner-awareness in experience that has no corresponding first-order state.

I will have more to say about such 'empty' higher-order states in what follows. For now, I want to make clear that the variety of SR I will work with does not necessarily deny this possibility, but it does place certain important qualifications on it. Before stating these qualifications, I set out the general structure of the view: according to it the following thesis

[1] This division of labor in the production of experience is endorsed explicitly by the HO-Thought theorist David Rosenthal at (Rosenthal 2005c: 209) and by the SR theorist Uriah Kriegel in his (2009) distinction between "subjective character" (the "for-me-ness") and "qualitative character" (the "what" of phenomenal experience).

[2] For further reflection on this peculiar kind of awareness in experience and its role in the phenomenological tradition see (Zahavi 2006); and for a more 'analytic' perspective see (Rosenthal 2005a and 2005d).

concerning the *metaphysical* structure of conscious states should be added as a clarification of the constitutive relation posited in SR:

> ***M**-to-*M* One-Sided Dependence** The higher-order state (*M**) constitutive of a phenomenally conscious experience as of *S*'s being in *M* is "*one-sidedly* dependent" on the first-order state (*M*) it represents. That is:
>
> Necessarily, if a token of the state-type *M** exists in *w*, then a token of the state-type *M*, which is the intentional object of this token of *M**, exists; but, possibly, a token of *M* exists in *w* without a token of *M**.

This view of the constitutive relation of higher-order and first-order states in experience is not the only one on offer. But I take it to be preferable to an SR view on which the dependence runs in the other direction, i.e., from *M* (the first-order state) to *M** (the higher-order state). For, on the *M*-to-*M**, there are no contingently unconscious subjective qualities of experience. All qualities of an experience are necessarily either *phenomenal* qualities (i.e., they are represented by a higher-order state on which they are dependent) or "non-conscious" subjective qualities (that are not represented by, and so not dependent on, a higher-order state). This failure to recapture the intuitive idea that any particular case of experience is such only contingently is, I take it, enough for SR to renounce a *M*-to-*M** one-sided dependence view as well as a "*mutual* dependence" view, which combines both directions of one-sided dependence into one constitutive relation. Given that there are no other ways to spin the dependence here, *M**-to-*M* is the best one going for SR.

3. *The Fallibility of Inner-Awareness*

SR and *M**-to-*M* one-sided dependence, however, may seem an unpalatable combination because it seems to entail an unqualified commitment to the *infallibility* of inner-awareness. If a state of inner-awareness *M** is one-sidedly dependent on the object that satisfies its intentional content – i.e., on *M* – then, according to how I formulated this relation above, necessarily, if *M** exists, so does *M*. Thus, necessarily, all instances of inner-awareness are veridical.[3]

Now, the infallibility of inner-awareness is importantly different in kind from the infallibility of introspection. It is commonplace to understand persons as liable to error concerning their own attitudes, emotions, and beliefs. And – even though this fact is sometimes staunchly denied by philosophers[4] – it is also amenable to commonsense intuition to believe that our introspective judgments about *sensory* states sometimes go wrong (e.g., I introspectively judge myself to *feel* hungry when in fact I do not or to no longer be in pain when I still am – more on such cases soon). But these, by themselves, do not demonstrate the fallibility of inner-awareness, i.e., of the intentionality of higher-order states constitutive of states of experience. However, I believe that it is not very difficult to see why the fallibility of introspection must be interpreted in both HO and SR as implicating the fallibility of inner-awareness.

[3] Cf., the notion of infallibility described at (Armstrong 1963: 417).

[4] Especially after the influence of Kripke's (1980) arguments against identity theories. For criticism of these arguments from the perspective of an extrinsic theorist of phenomenal consciousness see (Rosenthal 2005a: 38-40, 2005b: 136-9).

To see how this goes, consider the following two dimensions of epistemic normativity:[5] One dimension of normativity – let's call it an "*externalist*" dimension – concerns the reliability of the connection that inner-awareness establishes between one's first-order mental life and introspective belief. The second dimension of normativity – an "*internalist*" dimension – concerns the status of the subject herself, e.g., the rational appropriateness (or non-arbitrariness) of the subject's responses to her mental life, as experienced from the inside. These two dimensions of evaluation are, I think, applied independently of, but alongside, one another. Consider a rationally arbitrary subject, who judges her experience to be an appearance as of a green surface (i.e., "qualitatively green") even though it actually pre-reflectively appears to her that her experience is qualitatively red, and she has no good reason to think that her experience is not qualitatively red (e.g., she has no reason to think that she is the victim of an Evil Demon that is subjecting her to systematically misleading states of higher-order awareness in experience). While we consider such a subject to be epistemically deficient in the internalist dimension, if we grant that the higher-order awareness in her experience still provides a reliable connection to the actual qualitative character of her first-order state, we should consider her to be epistemically better off than she would be otherwise. In other words, we would not consider her to be without some epistemic grounds for introspective knowledge in her pre-reflective awareness of herself as experiencing red.

The kind of epistemic failure that must be descried in commonsense intuition to motivate the fallibility of inner-awareness, then, is a failure along the externalist dimension: a failure that does not directly implicate subjective fault, but only a fault in the connection that inner-awareness establishes between first-order mental life and introspective belief about it. In other words, it must be a failure that does not consist in the use that a subject makes of the evidence at her disposal, but in the very nature of the evidence itself. For what HO and SR must posit to accommodate this sort of failure is a deficiency in intentional relatedness to our own mental lives that is established by the higher-order representational states constitutive of it. For instance, these theorists must posit a kind of failure wherein it pre-reflectively appears to the subject that she is having an experience that is qualitatively red, but the first-order state represented in such a state of awareness is *actually* qualitatively green. In such a case, a subject would be epistemically responsible in judging that her first-order mental life has the feature of qualitative redness; this is the rationally appropriate response. But her judgment would be epistemically awry along the externalist dimension: for her inner-awareness does not, at least in this kind of circumstance, yield reliable evidence.

One might think that this is all fine as an idle distinction, but wonder whether it actually has any application to us. However, it seems, examples of failure along the externalist dimension are not hard to find, when one approaches the phenomena from within the framework of the extrinsic conception of phenomenal consciousness set up by HO and SR.

Take, for instance, false pains due to "dental fear." In these cases, the patient has been administered anesthetics and it is indisputable that no pain is actually occurring; but, when the drilling starts, the subject is still experiences pain. In such cases, it is also not unusual for the pain sensations to 'disappear', as it were, after the patient has been reminded that she is anesthetized. The clearest cut interpretation of such cases is that the subject's introspective

[5] For further discussion of the distinction I invoke in the following, see (Sosa 2007: especially pp. 250-4).

judgment under the drill – "I am in pain" – is not epistemically deficient due to any fault of the subject, but rather to the nature of the nature of the pre-reflective higher-order awareness in her experience that purports to reveal that she has a first-order mental state with the quality of pain.

A similar interpretation is applicable to certain introspective judgments in the well-known cases of blindsight. In a blindsighter, certain part of the visual field is, from the subject's point of view, completely blind; and this is how she introspectively judges her experience in these fields to be. Nevertheless, there is good reason to maintain that she wrong about this. For, when forced to make a choice about what is there (e.g., a red ball or a green ball), she can give a correct answer (which she takes to be a mere guess) with a reliability well above chance. What this attests to is the actual presence of first-order mental states, produced by the visual system, that reliably represent the actual layout of the visible world; but the lack of a reliable process of producing appropriately correspondent higher-order representations of these first-order states. So, the epistemic failure in these cases is an failure along the externalist dimension: a failure in the misleading nature of the evidence provided by the inner-awareness the subject's have of the blindsighted portions of their visual fields.

One who rejects an extrinsic conception of consciousness would certainly not be convinced by these arguments for the fallibility of inner-awareness from particular cases. Indeed, there are many alternative interpretations to wrangle with here. But it is no matter. For, since my concern is with the acceptability of my preferred version of SR, there is also a line of argument against the infallibility of inner-awareness that starts from its basic commitments that is at once more decisive than the foregoing arguments from counterexample and indicative of a way for SR to avoid being caught in the net of infallibility without abandoning the commitment to one-sided dependence.

To state the argument, I must first clear up a confusion that commonly attends philosophical discussion of fallibility. Following the etymology of the word, "to be fallible" means "to be *able* to be fooled." This suggests that the concept of fallibility is to be understood as indicating primarily a feature of cognitive *capacities*, not (as is sometimes implied) a feature of cognitive *states*. This is not to say that he concept of fallibility cannot be applied to states; but it is to suggest that any such application is legitimated only by consideration of the cognitive capacities productive of mental states.[6]

Now for the argument against infallible inner-awareness I want to build on this clarification.[7] It starts from the observation that, according to the extrinsic concept of phenomenal consciousness employed by SR, the capacity of inner-awareness deployed in making states phenomenally conscious is best understood to be a *receptive* capacity of cognition. From there, it demonstrates the absurdity of the idea that inner-awareness is infallible by connecting it to the absurdity of the very idea of a infallible capacity of receptive cognition.

Let's understand a cognitive capacity to be *receptive* when its conditions of *successful* execution involve the delivery of something that is not provided merely in the actualization

[6] This can be understood as one of the central themes in McDowell's epistemology. See, for instance, (McDowell 1982 and 2011).

[7] This argument is an application of the more general argument against the infallibility of receptive cognitive capacities found at (Rödl 2007: 152-3).

of the capacity itself. In other words, a receptive capacity is such that its conditions of success require that something more than what is involved in bringing the capacity to actualization is '*received*' from beyond. To illustrate, consider a non-cognitive capacity that has a familiar receptive character, such as the capacity to catch a baseball. For the actualization of this capacity to be *successful*, for it to achieve the end for which it is typically actualized, more is required than the mere actualization of the capacity present in appropriately waving and gripping the glove; there must also be the flying of a ball. So also the success-conditions of an act of perception require more than the mere actualization of visual capacities; the environment must also provide appropriate conditions and objects. Finally, much the same analysis of our capacity for inner-awareness in experience follows from the extrinsic conception of phenomenal consciousness employed by SR (and HO): for a subject to have a 'successful' experience as of a red ball, more is required that the actualization of her capacity for inner-awareness by itself; the presence of a first-order visual state with the suitable subjective quality must also be provided.

We are now in a position to see why there is no good sense in which *infallible* cognitive capacities can be construed as receptive. In particular, there is no good sense in which an infallible capacity of cognition is dependent on anything external to what is provided for its actualization. For since successful execution of the capacity is logically guaranteed for each case of its actualization, there is no good sense in which such capacities depend on the reception of anything external in order to be successfully executed. Infallibility forces cognitive capacities into the divine mold of a *constitutive* capacity. And so, if SR were to construe inner-awareness as infallible in the proper sense, it would end up with the incompatible properties of being both receptive and, as a consequence of infallibility, constitutive. One or the other of these characterizations must be renounced.

The result of these arguments seem to indicate that SR is in dire straits. But, as mentioned, along with the fortification of the arguments against infallibility, the clarification of this concept also indicates a way forward for SR, a way that prizes the one-sided dependence of inner-awareness apart from infallibility.

Briefly, the tactic is to make SR compatible with the fallibility of inner-awareness by construing the doctrine of one-sided dependence as an assay *only* of those states of inner-awareness that are the products of *successful* execution of the underlying capacity. It is, in other words, to take the realization of one-sidedly dependent states of inner-awareness as the *standard* by which we judge the *aptness* of the capacity's actualization. And it is to conceive of those states of self-awareness that lack a corresponding mental state (e.g., experiences of false pain and blindsight) as products of the capacity for inner-awareness that fall short of the standard; and so are instances of an altogether different kind of state.

This is, in effect, a *disjunctivist* construal of the inner-awareness of experience. For experiences with veridical or "standard" states of inner-awareness are instances of a different *type* from experiences with non-veridical or "non-standard" inner-awareness. Yet, this difference in the nature of the two states does not prevent the possibility of pairs of standard/non-standard states of inner-awareness that are, from the subject's perspective within the experience, completely indistinguishable from one another. So that non-standard cases qualify as spurious or psuedo-states of inner-awareness

This position demands further explication. But, it would be better to begin again from another place, concerned with its motivation.

4. *The Non-Grippingness of Internal-World Skepticism*

It is not difficult to turn a group of perfectly normal undergraduate students into a company of external-world skeptics. Artful application of certain doubts concerning the reliability of the senses and our other means of coming to know the external world usually does the trick. We consider whether it is possible that all of my perceptions and beliefs are caused by a Cartesian Evil Deceiver, whether it is possible that I am constantly dreaming or plugged into a machine that generates the world presented through my lived experience. In light of the apparent lack of any subjectively accessible criteria on the basis of which to rule out such hypotheses, we easily become gripped by the possibility that all our beliefs about the external world are radically and systematically false. However, there are limits to this. As Descartes noticed, there is a striking asymmetry with the prospects of motivating global *internal* skepticism using similar methods, i.e., a skepticism that maintains that all all of our immediate, noninferential, introspective beliefs about our own current mental states are radically mistaken. Such considerations concerning the reliability of the self-awareness essential to our own subjective experience seem idle and powerless. *Local* skepticism, of course, is not as difficult to motivate; the reflections concerning the fallibility of inner-awareness in the previous section demonstrate that. But a global skepticism about the reliability of inner-awareness hardly gets any firm intuitive grip.[8]

The non-grippingness of global "internal-world" skepticism is often not discussed in prominent theories of phenomenal consciousness. This is so, perhaps, because it is not taken to be the subject-matter of a theory of the metaphysics of consciousness, but a topic in epistemology. Or, perhaps, because considerations of this phenomenon have too quickly motivated commitments to infallibilism or to a thoroughgoing fallibilism that resorts to (what Horgan, Tienson, and Graham (2006: 43) call) a "debunking" explanation of non-grippingness: to wit, an explanation that treats it as a pervasive error, due to a persistent cognitive illusion to which human subjects are prone.[9] Yet neither of these warrants the omission.

Concerning the claim that non-grippingness of internal-world skepticism is exclusively an epistemological matter, it is important to realize that this issue can be pursued from two angles (cf. Horgan and Kriegel 2007: 124). From one, the question is patently epistemological: what is it about *knowledge* that makes *self*-knowledge (of phenomenal consciousness) so special? From the other, however, the question is one about the metaphysics of phenomenal consciousness: to wit, what is it about *consciousness* that makes knowledge of it so special? These two approaches are complementary. The former provides insight into the nature of that variety of self-knowledge that enjoys the epistemic privilege of the first-persons perspective; the latter takes as its *explanadum* the source of this epistemic privilege in phenomenal consciousness itself. So, insofar as a theory of phenomenal consciousness wants to give a complete treatment of its subject matter, it must deal with the non-grippingness of internal-world skepticism and the (purported) privileged accessibility of phenomenal consciousness that underlies it.

---

[8] For further examples illustrating the non-grippingness of global internal skepticism and for an explanation of this phenomenon that is compatible with the explanation I advance below see (Horgan, Tienson, and Graham 2006).

[9] Descartes and, following in his footsteps, Kripke (1980) seem to take the path from such considerations to infallibility. And the considerations concerning the unreliability of introspection in (Schwitzgebel 2008) might indicate a need for a debunking explanation.

Concerning the reasoning of the "debunking" approach's advocate, note first that this theorist would claim that, even though we are not inclined to reject the privileged accessibility of experience, we rationally *ought* to. So, she does not deny the psychological fact of the non-grippingness of internal-world skepticism, only the import that we ought to take it at face value. So, as Horgan, Tienson, and Graham (2006: 43) observe, insofar as we take the precedence of common sense intuition seriously, and insofar as we have a viable positive explanation of it, *ceteris paribus*, the burden of proof rests on the side of the debunking theorist to motivate her dismissal of the theoretical import of non-grippingness.

Let's take up the patently epistemological angle on the issue of non-grippingness first. Here, the precedent question is what analysis of the epistemic import of the epistemic privilege of the first-person perspective befits the non-grippingness of internal-world skepticism?

As it is commonly understood, privileged access is the fact that one's self-referential judgments about one's present and conscious mental states are epistemically special in that they do not rely on ordinary "empirical" or "observational" evidence. Rather, each of us is so related to her *own* current and conscious mental states that we can make *warranted* judgments about them without resort to anything other than the occurrence of the experience itself. There are many contending analyses of the basic epistemic import of privileged access. But the analysis that I find most compelling – mostly due to its combination of modesty and fruitfulness – is what William Alston (1971) calls the "Truth-Sufficiency" account. As Ram Neta formulates the doctrine:

> *S* knows, in a privileged way, that *p* just if and just when: *S* knows that *p*, and *S* knows this because *S* believes that *p* on the basis of the following justification that she has for believing it: *p*. (Neta 2008: 20)[10]

In other words, a subject has privileged access to a fact *p* just when *p itself* serves as sufficient canonical justification for her knowledge that *p*.

The facts in question here are what I will call "phenomenal facts." These are facts concerning the phenomenal '*feel*' or 'face-value' appearance that an experience has *for* a phenomenally conscious subject. It is, as one might also put it, the *intentional* object of the inner-awareness in an experience (the inner-awareness that is partially constitutive of experience). As such, just as an 'externally' directed intentional state may have an intentional object without also having a "material" object (e.g., my aural hallucination as of my infant daughter crying in the room next door has an intentional object – the sound of my daughter crying – but no material object – a sound produced in my daughter's expression of grief), so my 'internally' directed state of inner-awareness may have an intentional object without a corresponding material object. This is why, on an extrinsic conception of phenomenal consciousness, my *feeling* like I have a headache does not also justify belief in the fact that I actually *have* a headache. In other words, because the capacity for inner-awareness is *receptive*, it is possible, in principle, for the capacity to be actualized without having been provided the external conditions responsible for its *successful* actualization. So, on an extrinsic conception, there is an essential distinction to be drawn between the actual qualitative structure of my first-order mental life (the conditions of successful inner-

[10] (Neta 2008) also contains a clear and careful defense of the Truth-Sufficiency account against the objections which Alston (1971) raises against it.

awareness) and the way that my higher-order states of inner-awareness represent this first-order structure (the product of the actualization of the capacity for inner-awareness). And, on the basis of this, there is an essential distinction to be drawn between the kinds of beliefs that are immediately justified on the basis of phenomenally conscious experience, *viz*., beliefs about *phenomenal* facts about the intentional object of inner-awareness, and *not* beliefs about material facts about the actual structure of our first-order mental lives. This point, I think, is the pivot of the Truth-Sufficiency view's *modesty*.

This analysis also has strong intuitive support. Intuitively, I have privileged access to the fact that I *feel like* I have a headache. And it seems that this fact itself – the fact that I feel like I have a headache – in normal circumstances, is all I need to be justified in believing that I feel like I have a headache. I do not need a separate belief to be justified in believing this fact. Nor do I need any other experience.

The Truth-Sufficiency view thus helps illuminate the non-grippingness of internal-world skepticism by showing that it is part of the very epistemic import of privileged access that the very inner-awareness of experience itself is sufficient warrant for introspective belief. And so, it seems, there is no room to sow doubts that require a subjective criterion to rule out misleading cases of inner-awareness. For the experience itself serves as the canonical form of justification for introspective beliefs about it.

However, at this point, the significance of the angle on privileged access concerned with the nature of phenomenal consciousness comes into view. For what is it about a phenomenally conscious experience that makes it such that it can serve as a canonical form of justification for introspective beliefs about it?

The answer, in short, is that one must resort to a constitutive dependence relation, as SR urges, in order to illuminate this aspect of the issue concerning privileged access and the non-grippingness of internal-world skepticism. And it is precisely this omission in HO that renders it incapable of providing an acceptable positive account.

5. *SR's Explanatory Edge*

To make it sufficiently clear that HO does not provide a hospitable ontological framework for privileged access and the non-grippingness of internal-world skepticism, it will be helpful to note what the Truth-Sufficiency analysis of privileged access entails *by itself* concerning the defeasibility of introspective belief *states* about current and conscious experience. (In the following, I take "defeasibility" and the correlative "indefeasibility" to apply directly to cognitive states, unlike the notion of infallibility which applies directly to cognitive capacities, and only derivatively to cognitive states.)[11]

First, I distinguish two notions of the indefeasibility of introspective belief. One I call indefeasibility *simpliciter*. On this doctrine, *all* introspective belief concerning current conscious experience is such that the warrant or evidence which justifies the belief cannot be undermined or overturned in light of further evidence. For instance, an introspective belief state that is indefeasible *simpliciter* is such that no further considerations can show the belief state to be unjustified (in either the 'internalist' or 'externalist' dimensions of epistemic evaluation distinguished earlier – more on this soon). The other concept of indefeasibility derives from Scott Sturgeon's work; it is a variety of indefeasibility I call "conservative indefeasibility." A conserva-

[11] This seems to be the use that McDowell (2011) – from which work I take the point about the proper use of the concept of infallibility in section 3 – makes of the concept of indefeasibility.

tively indefeasible belief state is one that is not subject to a certain kind of epistemic defeater – what Sturgeon calls a "conservative defeater" – that, as Sturgeon expresses it,

> break[s] the link between a concept and its canonical evidence while leaving that evidence intact. Conservative defeaters place one in an epistemic situation in which one may fully appreciate the possession of canonical evidence for the application of a concept, but remain unable to apply that concept with warrant. (Sturgeon 1994: 229)

In other words, a conservatively indefeasible belief is one whose justification is not possible to undermine without also modifying the evidence or warrant that epistemically grounds the belief.

So, applying this to the concept of privileged access articulated in the Truth-Sufficiency analysis, a conservatively indefeasible introspective belief about the phenomenal character of a current, conscious experience is one that is not subject to defeat without also modifying the lived qualitative character of the experience which serves to justify the introspective belief. In other words, so long as the phenomenal character of the experience remains unchanged, the introspective belief grounded on it maintains its justification.

Beliefs about external facts, on the other hand, whose justification is based on perceptual experience (e.g.), are typically understood to be subject to conservative defeat.[12] For instance, it is standard to acknowledge that it can *visually* appear to me as if there is a red ball before me (this appearance being a fact about my experience) while I consider myself to be unjustified in believing that there actually is a red ball before me (this being an external fact) because, say, I have discovered that the environment is subtly flooded with red light. Because the experience serves here as a sort of evidential "intermediary" – i.e., since it serves to indicate that a separate, independently existing fact obtains – it is possible for the experience to remain as it is while the warrant it provides for the belief about the external state of affairs is undermined (cf., Sturgeon 1994: 228).

In light of this, it might seem that the Truth-Sufficiency analysis of the privileged access entails the conservative indefeasibility of introspective beliefs that enjoy privileged access. But this is not correct. This implication depends crucially on the metaphysics of phenomenal consciousness that complements the epistemological analysis. Recall the two dimensions of epistemic evaluation distinguished earlier: the 'externalist' dimension, concerned with the reliability of the connection between the object and the belief state; and the 'internalist' dimension, concerned with the rational appropriateness of a subject's cognitive responses to her experience. The dimension of epistemic evaluation that is at issue for conservative defeat is the externalist dimension. For what is at issue in the metaphysical supplement to the theory is a picture that either affirms or undermines the claim that the evidential ground of introspective belief about current, conscious experience guarantees the truth of the belief, i.e., whether one's consciously experiencing herself to be in a certain mental state can serve as conclusive externalist warrant for the corresponding introspective belief.

Now we are in a position to see the deficiency in the HO metaphysics of consciousness relative to the metaphysics of dependence SR offers. In particular, the HO view cannot accommodate a *positive* account of the conservative indefeasibility of introspective belief. Rather, it must

[12] Of course, a disjunctivist about perceptual experience will dispute this point about perceptual belief, because they deny the non-factivist view of perceptual states (cf., Martin 2006 and McDowell 1982, 2011). But since I here only use the non-factivist view of perceptual states as a foil for the factivist view of states of inner-awareness, I need not also endorse non-factivism about perception; and so, I need not address this debate in the theory of perception.

opt for a "debunking" approach: an approach that maintains that we rationally ought not take our experience to guarantee the truth of occurrent corresponding introspective beliefs.

Consider the following line of thought:

> (1) According to HO, it is possible for a subject S to be in a state of inner-awareness M*, which represents S as being in a state M, without S's actually being in M.
> (2) Also, M* and M are distinct existences, i.e., the same most specific kind of state of inner-awareness M* can exist in both standard and non-standard cases of experience.
> (3) Thus, since M* is a common factor of both standard and non-standard cases of inner-awareness, it is also possible for S's introspective belief that S (herself) is in M to be defeated while S is still experientially aware of herself as being in M.
> (4) Therefore, by definition, introspective belief about current conscious experience is not conservatively indefeasible.

Premise (1) is a restatement of the key explanandum: that it is possible to have subjectively indistinguishable pairs of standard and non-standard cases of inner-awareness in experience. Premise (2) articulates the consequence of the HO stipulation that the higher-order state of inner-awareness and the first-order state normally constitutive of conscious experience are ontologically independent. So it is possible for the very same higher-order state of inner-awareness *M** to be either 'standard' or 'empty'. The characterization of the inner-awareness in experience as one or the other accrues to it on the basis of factors that are completely independent from the state of inner-awareness itself. From this (3) and (4) follow. For now it is possible for a state of inner-awareness *M** that provides justification for phenomenal belief on some occasions – i.e., on those occasions where a corresponding first-order state is present – to be present in and constitutive of the experience of a subject while not justifying her introspective belief that her phenomenally conscious experience actually is the way it seems, since the corresponding first-order state is actually absent. And so in any case the subject formulates her introspective beliefs about current and conscious experience on an evidential basis that does not guarantee truth. So, by definition, these beliefs are subject to conservative defeat.

SR, on the other hand, does not face this consequence. For SR denies premise (2), maintaining instead that:

> (5) It is impossible for *S* to be in a state of inner-awareness *M**, which represents *S* as being in a state *M*, without *S*'s actually being in *M*.

This modification blocks the way to the common-factor claim in (3), which, in turn, leads us away from the conservative defeasibility introspective phenomenal belief.

6. *Disjunctive Self-Representationalism*

Commitment to (5), however, seems only to lead us out of the frying pan and back into the fire. For now, how should SR accommodate the introspective indiscernibility of standard and non-standard (or 'empty') cases of inner-awareness? And, along with this, the fallibility of inner-awareness?

To see our way forward with this question, it is important to realize the ontological assumption that motivates it:

Introspective Difference Principle (IDP) Two states of inner-awareness have different natures (they are of a different most specific type) iff the experience constituted by these states are introspectively distinguishable.

With IDP in place, there is no way to explain the indiscernibility of standard and non-standard cases of inner-awareness except by positing a mental state type that is *common* to both cases.

But the disjunctivist urges that there is no conclusive reason to accept this principle. For one can also opt for another principle in explaining the indiscernibility of standard and non-standard cases of inner-awareness:

Disjunctivism (DISJ) An experience that is, from *S*'s point of view, as of *S* (*herself*) being in *M* is either a case of *S*'s actually being in *M* or merely *seeming* to be in *M*.

On the conjunctivist view that follows from IDP, *M** is all that is needed, on any given occasion, for the occurrence of an experience wherein *S* is aware of herself as being in *M*. On this view, as mentioned above, "aware of" introduces and intensional context. And, as a result, the HO view allows room for the conservative defeat of phenomenal belief.

On the disjunctivist SR view that follows from DISJ, the essence of *S*'s awareness of herself as being in *M* involves a "*factive*" state of inner-awareness, which is such because it is one-sidedly dependent on its intentional object, *M*. And it follows that any introspective phenomenal beliefs grounded on a standard case of inner-awareness is immune to conservative defeat. For this mental state itself guarantees its own justification when its phenomenal nature is regarded introspectively.

Now, one might object at this point that, regarded in light of familiar skeptical considerations, the ontological distinction I am drawing between HO and disjunctive SR really makes no epistemological difference. For on both the HO and SR views, the experience – i.e., whatever the subject of an experience 'lives through' or has access to 'from the inside – does not by itself, issue the subject the assurance that it is actually a standard case of experience, instead of a non-standard case. This is a hazard that issues forth from the very fact of indiscernibility. Of course, the obstacle to knowledge that the skeptic wants to reveal here takes a different shape in each theory. On the HO view, to overcome the obstacle is to find a way to discern from the inside when one's first-order states don't correspond with the intentional object of inner-awareness, given that our capacity of inner-awareness is, as it were, not attached to the actual fabric of mental life. And on the SR view, the challenge is to discern from the inside when one is in a standard state of experience as opposed to a non-standard state. So that, it seems, introspection by itself falls short of being able to deliver a subjectively accessible guarantee that one's mental life is as it appears to be.

However, HO and SR are not, in fact, on equal epistemological footing with regard to this "internal-world" analogue of Cartesian skepticism.[13] To see how, consider a distinction Quassim Cassam (2007: 2) draws between two ways one might go about showing the possibility of knowledge in light of (purported) obstacles to it such as the indiscernibility of standard and non-standard cases of experience. One way employs an "obstacle-*overcoming*" strategy, which acknowledges that there is an obstacle to knowledge in place, and attempts to show how our

[13] See (Horgan, Tienson, and Graham 2006) for further discussion of this application of skeptical doubt.

cognitive capacities can overcome these obstacles. The other way employs what Cassam calls an "obstacle-*dissipating*" strategy, which attempts to start from a view of our cognitive capacities that entails the non-existence of the very obstacle on which the skeptical challenge depends.

HO, it seems, must employ a obstacle-overcoming strategy. For the skeptical challenge here stems from the metaphysical picture of phenomenally conscious states that the theory employs: a picture that stipulates the complete independence of first-order states and higher-order states constitutive of experience. The HO theorist, then, must specify what *else* is to be added to the nature of subjectively manifest data that can overcome the obstacle to knowledge the theory entails. The disjunctive SR theorist, on the other hand, is free to abandon this project and the associated skeptical worries by adopting the obstacle-dissipating strategy. For since there is, by the very picture of the ontology of standard cases of inner-awareness it employs, no 'gap' between states of inner-awareness and corresponding first-order states, there is no place for the internal-world Cartesian skeptical worries to take root. This is not, of course, to say that one is *completely* free from all doubt about the epistemic adequacy of one's introspective beliefs. For indistinguishable pairs of standard and non-standard states are not banished. But one is free from the doubt that, even in a standard case of inner-awareness in experience, the experience itself still falls short of delivering grounds for knowledge.

Works Cited

Alston, William P. 1971. "Varieties of Privileged Access." *American Philosophical Quarterly* 8 (3): 223-41.

Armstrong, David M. 1963. "Is Introspective Knowledge Incorrigible?" *Philosophical Review* 62: 417–432.

Bernecker, Sven. 2008. "Representationalism, First-Person Authority, and Second-Order Knowledge." In *Self-Knowledge*, ed. Anthony E Hatzimoysis, 33–52. Oxford University Press.

Block, Ned. 2011. "The Higher Order Approach to Consciousness Is Defunct." *Analysis* 71 (3): 419–431.

Byrne, Alex. 1997. "Some Like It Hot: Consciousness and Higher-Order Thoughts." *Philosophical Studies* 86 (2): 103–129.

Dretske, Fred. 1995. *Naturalizing the Mind*. MIT Press.

Horgan, Terry, and Uriah Kriegel. 2007. "Phenomenal Epistemology: What Is Consciousness That We May Know It So Well?" *Philosophical Issues* 17: 123-144.

Horgan, Terry, John Tienson, and George Graham. 2006. "Internal-World Skepticism and the Self-Presentational Nature of Phenomenal Consciousness." In *Self-Representational Approaches to Consciousness*, eds. Uriah Kriegel and Kenneth Williford. MIT Press.

Kriegel, Uriah. 2009. *Subjective Consciousness: a Self-Representational Theory*. Oxford University Press.

Kripke, Saul A. 1980. *Naming and Necessity*. Harvard University Press.

Martin, Michael G. F. 2006. "On Being Alienated." In *Perceptual Experience*, eds. Tamar Szabó Gendler and John Hawthorne. Oxford University Press.

McDowell, John. 1982. "Criteria, Defeasibility, and Knowledge." *Proceedings of the British Academy* 68: 455–479.

. 2011. *Perception as a Capacity for Knowledge*. Marquette University Press.

Neander, Karen. 1998. "The Division of Phenomenal Labor: a Problem for Representational Theories of Consciousness." *Noûs* 32: 411-434.

Neta, Ram. 2008. "The Nature and Reach of Privileged Access." In *Self-Knowledge*, ed. Anthony E. Hatzimoysis, 10–32. Oxford University Press.
Rödl, Sebastian. 2007. *Self-Consciousness*. Harvard University Press.
Rosenthal, David M. 2005a. "Two Concepts of Consciousness." In *Consciousness and Mind*. Oxford University Press.
Rosenthal, David M. 2005b. "The Independence of Consciousness and Sensory Quality." In *Consciousness and Mind*. Oxford University Press.
Rosenthal, David M. 2005c. "Sensory Qualities, Consciousness, and Perception." In *Consciousness and Mind*s. Oxford University Press.
Rosenthal, David M. 2005d. "Introduction." In *Consciousness and Mind*. Oxford University Press Press.
Rosenthal, David M. 2011. "Exaggerated Reports: Reply to Block." *Analysis* 71 (3): 431–437.
Schwitzgebel, Eric. 2008. "The Unreliability of Naive Introspection." *Philosophical Review* 117 (2): 245–273.
Sosa, David. 2007. "Perceptual Friction." Philosophical Issues 17 (1): 1–17.
Sturgeon, Scott. 1994. "The Epistemic View of Subjectivity." *Journal of Philosophy* 91 (5): 1–16.
Weisberg, Josh. published online 2010. "Misrepresenting Consciousness." *Philosophical Studies*.
Zahavi, Dan. 2006. "Thinking About (Self-)Consciousness: Phenomenological Perspectives." In *Self-Representational Approaches to Consciousness*, eds. Uriah Kriegel and Kenneth Williford. MIT Press.