Philos Stud (2015) 172:1759–1782
DOI 10.1007/s11098-014-0389-6

The idols of inner-sense

Chad Kidd

Published online: 28 October 2014

Abstract Many philosophers hold one of two extreme views about our capacity to have phenomenally conscious experience ("inner-sense"): either (i) that inner-sense enables us to know our experience and its properties *infallibly* or (ii) the contrary conviction that inner-sense is utterly *fallible* and the evidence it provides completely defeasible. Both of these are in error. This paper presents an alternative conception of inner-sense, modeled on *disjunctive* conceptions of perceptual awareness, that avoids both erroneous extremes, but that builds on the commonsense intuitions that motivate them.

1 Introduction

The title of this paper comes from a classic but uncelebrated paper on the nature of self-knowledge by Max Scheler. In it, Scheler sets out "to purify the 'turbid mirror' of our understanding" of self-knowledge by articulating a "critical theory of Idols." By "idols" he means the "natural inclinations to illusion and error" that have a pervasive and unnoticed influence on empirical and philosophical theorizing (Scheler 1992, p. 3). Part of Scheler's procedure was to list these idols and to itemize their effects on philosophical theorizing. Another part was diagnostic. This Scheler carried out by tracing the origin of these idols to our otherwise innocuous body of commonsense intuitions about the mind. My procedure in this paper is like Scheler's. I seek first to list two idols that distort contemporary thinking about phenomenal consciousness. I then seek to disclose their effects and origins. Both idols primarily distort our thinking about the nature of "inner-sense" (to use Russell's 1912 terminology): a power or faculty that is responsible for the self-awareness that is constitutive of phenomenally conscious experiences. The two

C. Kidd (✉)
Auburn University, Auburn, AL, USA
e-mail: chad.kidd@auburn.edu

erroneous beliefs I focus on are: (i) inner-sense enables subjects of phenomenally conscious experience to know these experiences infallibly and (ii) the contrary conviction that inner-sense is just as fallible and limited as our capacity for perceptual awareness. Both of these are in error. In the following, I present an alternative conception of inner-sense, modeled on *disjunctive* conceptions of perceptual awareness, that avoids both erroneous extremes, but that builds on the commonsense intuitions that motivate them.

Here is how the paper proceeds: In Sect. 2, I present an overview of the prominent theories of phenomenal consciousness by isolating their common *explanandum*—phenomenally conscious experience—and by drawing a distinction between two conceptions of phenomenal consciousness—the *intrinsic* and *extrinsic* conceptions—that are often confused in the literature. I then indicate the family of theories that I am concerned with in this paper: the family of extrinsic intentionalist theories of phenomenal consciousness, which take conscious experience to consist in a subject's becoming aware of herself as having a certain mental state, and which is facilitated by one of the subject's mental states becoming *represented by* another (higher-order) mental state of the same subject. After that, in Sect. 3, I present the two idols that provoke the dispute between these two prominent members of the extrinsic intentionalist family of theories—the higher-order representational (HO) and the self-representational (SR) theories of phenomenal consciousness. In Sect. 4, I present an overview of HO and SR, of their motivation, and then I give an initial presentation of my alternative SR view—*disjunctive* SR—that avoids the idols associated with HO and traditional SR. Section 5 criticizes the traditional SR commitment to the infallibility of phenomenal self-awareness. Here I argue that such a commitment is itself incoherent and I consider cases (often used to motivate HO) that show the possibility of illusory higher-order awareness in experience (or, as I also call it, *inner-illusion*). In Sect. 6, I outline the structure of the disjunctivist SR view, and show that this is the *only* variety of SR that can accommodate the possibility of illusory inner-awareness. The rest of the paper attempts to motivate the disjunctive SR over HO by combating the inclination to reject the legitimacy of first-person authority and the conceptual connection between phenomenal consciousness and first-person authority. More specifically, Sect. 7 attempts to vindicate first-person authority by analyzing the fact that Cartesian-style skeptical thought experiments seem to have no skeptical grip at all when applied to introspective judgments about occurrent conscious experience. And it shows that disjunctive SR can recover this intuition without postulating an infallible inner-sense, as traditional SR does. Section 8 presents the core argument of the paper against HO: in principle, it *cannot* vindicate the non-grippingness of skepticism about introspective judgments concerning occurrent conscious experience. And I close in Sect. 9 by answering two important objections to the superiority of the disjunctive SR theory over the HO theory of phenomenal consciousness.

2 Intrinsic and extrinsic conceptions of phenomenal consciousness

Given the wide variety of theories of consciousness available today, it is helpful to get a handle on them by first isolating the relevant meaning of the multiply

ambiguous concept of *phenomenal consciousness* or *experience* itself. Ever since Thomas Nagel's influential article "What is it like to be a bat?" (Nagel 1974), it has become customary for philosophers to distinguish between "*unconscious* awareness" and "*conscious*" or "*phenomenally conscious* awareness" of the external world in the following way. In phenomenally conscious awareness there is "something it is like *for the subject*" to be the subject of awareness, whereas there is no such "what it's likeness" for the subject of a state of unconscious awareness. To illustrate, consider an extreme case of unconscious awareness: the kind of awareness that a blindsighted person has of objects in the blind-sighted portion of her visual field. A subject of blindsight can visually detect features of objects in certain portions of her visual field, even though she does not take herself to be visually aware of anything. So, going merely on the evidence provided by her own experience, she has sufficient reason to think that she is blind. However, psychologists nevertheless maintain that she still detects visual stimuli. For when given a few options about what is before her eyes and *forced* to guess, the subject guesses correctly at a remarkably high rate, well above chance. So—here again, despite the subject's initial understanding of her situation—the psychologist concludes that the subject must not *really* be guessing, but must be "unconsciously seeing" what is there before her. This interpretation becomes even harder to avoid in those cases of blindsight where the subject can handle and manipulate objects with the facility of a normally sighted individual.[1] In such cases, it seems as if the malady of blindsight has little to no effect on a subject's visual connections to the world; but that it only somehow obstructs the connection visual awareness normally has to the subject's awareness of her own mental states.

It is this "consciousness of consciousness" or "awareness of awareness" that some philosophers have identified as the conceptual core of phenomenal consciousness. These philosophers take seriously Nagel's phrase that a subject is phenomenally consciousness when there is "something it is like *for the subject.*" And they emphasize the fact that this "something" the conscious subject is aware of is other than the objects of visual awareness—"external" objects such as tables, cars, houses. Instead, it is an awareness of one's own perceptual awareness itself, i.e., of the subjective state (or event) of perceiving itself.

These philosophers also typically take blindsight to be a case of *unconscious awareness*. It is important to distinguish unconscious awareness from the kind of "unconsciousness" that we experience with regard to objects in the periphery of our visual field or to bodily sensations at the fringe of attention. Here too, there is good evidence to support the fact that we perceptually detect objects even though, prior to shifting our attention, we do not take ourselves to be aware of them.[2] The key difference from blindsight, however, is that, with peripheral consciousness, we *can* spontaneously become consciously aware of these stimuli by shifting our attention. And many philosophers hold that this shows we were already consciously aware of

[1] For further discussion of the empirical evidence from the psychologist that discovered this phenomenon see (Weiskrantz 2009).

[2] See, for instance, the strong evidence for this claim in (Norman et al. 2013).

these peripheral stimuli, but in a dim way. Merleau-Ponty (2012), for example, observes that for the subject of normal visual consciousness, the field of consciousness has a perspectival "object-horizon" structure wherein the object of attention "is the mirror of all the others" that I do not attend to at the moment. Therefore, normal conscious subjects "can see one object insofar as objects form a system or a world, and insofar as each of them arranges the others around itself like spectators of its hidden aspects and as the guarantee of their permanence" (Merleau-Ponty 2012, pp. 70–71). The solicitation of attention by objects in the periphery is never completely silenced in normal visual experience. Unattended objects in visual experience *are* present to mind, albeit *as* more or less determinately articulated potential abodes for visual attention.[3] For the blindsighter, however, *all* of this subtle phenomenology is absent in the blindsighted portion of her visual field: all solicitation of objects is silenced, cut off from the "world" of her conscious awareness. And so, the "what it's like" for the blindsighter in this regions is, in a word, *nothing*.

The theories of phenomenal consciousness with which I am concerned here are attempts to account for the *categorical difference* between "conscious" and "unconscious" awareness exhibited in the contrast between the normal sight and blindsight. Thus, this distinction is not that between two extremes on a continuum, like that which distinguishes the center from the periphery of conscious awareness for the normal subject. The center/periphery distinction is a distinction *within* the rich and varied experience of the normal subject, while the phenomenally conscious/unconscious distinction is one that limns the very boundaries of experience.[4]

Now, there are a wide variety of theories of phenomenal consciousness and it is difficult to articulate a system by which one may categorize them all. However, one helpful benchmark is the role of inner-sense in the explanation of phenomenal consciousness that each provides. The default position for theories of consciousness that take their lead from Nagel's observations is to construe phenomenal consciousness as an intrinsic and *sui generis* property of conscious mental states.[5] For a state of intentional awareness to be phenomenally conscious, then, is for it to have this property; and to have this property the state need not stand in any type of relation to anything else. So, phenomenal consciousness, on this view, is like an "inner-light" that illuminates intentional awareness for the subject. I will (following Weisberg 2010) call this an *intrinsic* conception of phenomenal consciousness. Now, given that phenomenal consciousness is taken care of by the internal structure of the experience itself, the intrinsic conception of phenomenal consciousness typically repudiates any explanatory role for a special mechanism of inner-sense in the production of phenomenal consciousness.

[3] An even more extreme view of the holistic structure of the field of consciousness can, on certain interpretations, be found in the work of Gurwitsch. See Chudnoff (2012).

[4] I might have tried to indicate the phenomenon of phenomenal consciousness by contrast with a deep dreamless sleep, being knocked out, or some other familiar form of *un*consciousness. But what is helpful about blindsight for my purposes is that it helps one get a grip on the distinction between conscious and unconscious *awareness*, instead of just consciousness and unconsciousness *simpliciter*.

[5] Cf., e.g., Block (2002), Chalmers (1996), Levine (2001), and McGinn (1996, chap 3).

However, there is another set of prominent views. These advance what I will call (again, following Weisberg) an *extrinsic* conception of phenomenal consciousness. They maintain that phenomenal consciousness is not an intrinsic property of individual states, but is instead a *relational* or *extrinsic* property of mental states, i.e., a property a mental state has only by virtue of a connection to something other than itself. There are two main varieties of extrinsic theory: one takes the structure to consist of intentional connections, the other to consist of non-intentional, "causal" or "functional" connections.[6] In the following, I will be concerned only with the former variety of extrinsic views, which take phenomenal consciousness to reduce to *intentional* facts about mental states.

Unlike the theories built on the intrinsic conception, extrinsic theories make ample room for inner-sense as an explanatory posit. Inner-sense is typically taken to be the capacity of mind that establishes the appropriate, "conscious-making" intentional connection. For example, on David Rosenthal's view, phenomenally conscious states are states that "we are conscious of being in" (Rosenthal 2005d, p. 26). Phenomenal consciousness is, thus, reducible to an intentional fact: the fact of our consciousness *of* consciousness. Inner-sense, then, is postulated as a mechanism of the mind, which monitors first-order states of awareness.[7] The extrinsic theorist thus postulates two cognitive capacities at work in every experience: one a capacity of "outer-sense" that yields externally directed intentional states; the other a capacity of "inner-sense" or, in the words of William Lycan, an "internal scanner or monitor that outputs second-order representations of first-order psychological states" (Lycan 1996, p. 31). The outer-sense is responsible for *what* we are consciously aware of (when we are consciously aware), the inner-sense is responsible for *whether* we are consciously aware of it.[8]

[6] For examples of non-intentional extrinsic views see Dennett (1991) and Baars (1997). Key examples of intentional extrinsic views are Kriegel (2009), Lycan (1996), and Rosenthal (2005b).

[7] It would be wrong, then, to read this as suggesting a "sixth sense," alongside the "outer-senses" of vision, touch, taste, etc, which is just like the outer-senses except that it is directed at "internal" objects. As I am using this term here, "inner-sense" is not an organ—an inner-eye—or even a separate "module" of mind, in the sense of (Fodor 1983). Rather, the term "inner-sense" is chosen because it is suggestive of something to which all intentional extrinsic views are committed: that phenomenal consciousness is due to the proper functioning of a special cognitive *capacity* that produces (higher-order) intentional awareness of our individual (first-order) mental lives. I abstain from any further presumption about what this capacity is. I might just as well have used the metaphorical terminology of there being an 'internal scanner', an 'internal monitor, or an 'internal detection mechanism'. In any case, on my own view, what this mechanism is must be revealed by further philosophical and empirical research. The goal of a philosophical analysis of inner-sense, which I take up here, is to get an initial idea of what this capacity does—what its function is—so that cognitive scientists can both help clarify our conception of the capacity further and, ultimately, discover the mechanisms that are responsible for it.

[8] It is important to note that on the extrinsic intentional conception, it is *not* that higher-order monitoring *makes the mental state that it monitors* a conscious state. Higher-order monitoring is intentionality like any other. And neither intentionality nor phenomenal consciousness are a communicable properties. We don't give a rock the property of conscious vision by looking at it. So also higher-order intentionality does not give its objects—first-order intentional states—the property of being "phenomenally conscious," just by monitoring them. Rather, phenomenal consciousness just *is* the awareness that a subject has of her own mental life, made possible by the joint contributions of higher-order monitoring and the mental states that are monitored.

3 The two idols of inner-sense

We are now in a position to understand the two idols—or fundamental errors and the inclinations to error that motivate them—that I address in this paper. These fundamental confusions affect the debate surrounding two competing *extrinsic intentional views* of phenomenal consciousness: the HO and SR theories of phenomenal consciousness. The divide between these two theories concerns the structure of the conscious-making higher-order awareness and the nature of the capacity of inner-sense that produces it. SR theorists take the relevant form of higher-order awareness to have a token-reflexive structure; it is, in other words, an awareness that the intentional state has *of itself*. And, given this token-reflexive structure, SR theorists often construe the capacity of inner-sense productive of higher-order awareness to be *infallible* and the evidence it provides to be *indefeasible* (further clarification of these terms comes below). HO theorists, on the other hand, take the structure of higher-order awareness to be non-reflexive; it is an awareness grounded in a numerically distinct mental state, whose sole function is to monitor the goings-on of other ("first- or, at least, lower-order") mental states. Given this non-reflexive structure, HO theorists also typically understand the monitoring function of higher-order states—our capacity of inner-sense—to be *fallible* and the evidence that they provide to be completely *defeasible*.

The first idol I address in this paper is the commitment of SR theorists concerning the absolute infallibility of inner-sense and the indefeasibility of its evidence. The second idol is the contrary commitment of HO theorists concerning the utter fallibility of inner-sense and the defeasibility of its evidence.[9] The former is in error because it goes against the intuition (which partially motivates the HO view) that the (*pre-reflective*) awareness we have of our own experience—the awareness which is, by the extrinsic intentionalist hypothesis, *constitutive of* experience—can be illusory. And the latter is in error because it conflicts with the intuition (which partially motives the SR view) that phenomenal consciousness affords the subject a special epistemic authority over her own mental states.

The goal of this paper is to motivate an alternative SR theory of phenomenal consciousness, which accommodates the just mentioned intuitions that motivate *both* HO and traditional SR—the intuition that inner-sense is fallible (HO) and the intuition that inner-sense (by the production of conscious experience) is a source of epistemically privileged knowledge of our own minds—while avoiding the tendency for one of these intuitions to crowd out the other. I will call the new SR view the *disjunctive* SR theory, since it distinguishes itself from the traditional SR view by incorporating key aspects of a *disjunctive* theory of cognition. In particular, disjunctive SR accommodates the fallibility of phenomenal

[9] I should note that I do not here assume that the "evidence" provided by perception is inferential. It is not that my visual awareness of the coffee in the cup or the tactile awareness of heat are premises from which I then infer the belief that there is fresh coffee in the cup. Rather, I follow Husserl in holding the evidential relation between perception and belief (or intuition and judgment) to be much more intimate than that. The perceptual awareness is a presentation of the "thing itself"—the individual object or state of affairs—which is the object of my belief. For more on this view of perceptual evidence see (Kidd 2014).

consciousness by construing fallibility as a feature of the *mechanism* of inner-sense. But it accommodates first-person authority by construing the higher-order, token-reflexive states produced by the *successful* actualization of inner-sense as being (*per* their token-reflexivity) *metaphysically dependent* on the first-order states they represent. The dependence of higher-order states, produced by the successful actualization of inner-sense, on their lower-order objects secures the indefeasibility of the evidence provided by phenomenally conscious experience. For, so construed, these cases of experience actually *guarantee* their own veridicality.

4 HO, traditional SR, and disjunctive SR

As mentioned above, there are a variety of theories of phenomenal consciousness that differ in terms of their explanations of the categorical difference between "conscious" and "unconscious" awareness. Both HO and traditional SR advance the *extrinsic intentional* conception of phenomenal consciousness. Both therefore maintain that the key structure underlying phenomenal consciousness is an *intentional* fact about mental states: that a conscious subject has a mental state that represents herself as being in (or having) a certain first-order mental state.

In order to understand the differences between the two, it is helpful to begin with a precise statement of what each endorses. Put in a formula HO asserts:

> **HO** A subject *S* at time *t* is a phenomenally conscious subject if and only if *S* has a mental representation M* of herself as being in a first-order mental state M.

According to HO conscious awareness *normally* involves the tokening of two mental states in a subject's mind: (1) a first-order, world-directed mental state and (2) a higher-order mental state that represents the fact that the subject has the relevant first-order state. However, according to HO, the only state that is *necessary* and *sufficient* for conscious awareness is the higher-order intentional state that represents the subject as having a certain first-order mental state (cf., Rosenthal 2005a, p. 209).

Traditional SR, on the other hand, advances a more complex view of the metaphysical underpinnings of phenomenal consciousness. This view takes phenomenal consciousness to be grounded in a token-reflexive higher-order awareness, an awareness that a state has *of itself*, alongside whatever else the state may represent.[10] And so, at the very least, traditional SR requires the existence of a first-order state that the subject is (rightly or wrongly) aware of herself as having. Again, in a formula:

[10] Aron Gurwitsch (1985, p. 3) expresses the idea as follows: "When an object is given in experience, the experiencing subject is conscious of the object and has an awareness of this very consciousness of the object. Perceiving a material thing, listening to a musical note, thinking of a mathematical theorem, etc., we are not only conscious of the thing, the note, the theorem, etc., but are also aware of our perceiving, listening, thinking, etc. […] When we experience an act which presents us with an object other than itself […] we are aware in being confronted with the object of our being so confronted, we are aware of our experiencing the act through which the object in question appears to consciousness."

> **Traditional SR** A subject *S* at time *t* is a phenomenally conscious subject if and only if (a) *S* has a mental representation M* of herself as being in an intentional state M, (b) *S* in fact has M, and (c) M* is *dependent* on M.

With this, we can see (condition (a) in the formulation) that both HO and traditional SR agree that a higher-order awareness of oneself as being in a first-order mental state is a *necessary* condition for phenomenal consciousness. However, traditional SR denies that this alone is sufficient. For, by hypothesis, (condition (b)) the first-order state that the subject is aware of herself as having must also *exist*, and (condition (c)) the state of higher-order awareness must be dependent on the first-order state it represents. Only when these three conditions obtain, according to the most straightforward formulation of SR, is a subject the subject of conscious awareness.

In the next section, I will present one of the most forceful reasons for preferring HO over SR: that only HO can avoid the pressure to construe inner-sense as an *infallible* representational mechanism. However, before doing so, in order to avoid perplexities that inevitably arise when one considers objections to the infallibility of inner-sense, I should clarify two points. The first is that HO and SR are meant to explain what it is that differentiates *phenomenally conscious* from *unconscious* awareness. The capacity of inner-sense is the mechanism that produces the differentiating features of these two. The second is that the dispute over infallibility concerns whether this "phenomenally conscious-making" mechanism is itself fallible or infallible.

Now, the perplexity to be cleared away concerns the way a problem of misrepresentation of experience itself can arise in the domain of *non-reflective* conscious experience. There is little difficulty in understanding how misrepresentation can arise in the domain of introspective judgment. But *judging* that I am in pain or that I am jealous of my brother is a different matter from actually *being (consciously) in pain* or *being (consciously) jealous*. For I may be *conscious of* my jealousy in the mode of introspective judgment without being *consciously* jealous (say, by this being revealed to me through psychiatric analysis). And I may be consciously jealous without also reflecting on this fact. Given that the reflective state can exist without the state reflection represents, it leaves room for the possibility of reflective *misrepresentation* of experience. But this says nothing about the possibility of non-reflective misrepresentation of experience—a kind of misrepresentation that is somehow built into the experience itself. Is such a notion even intelligible?[11]

According to all extrinsic conceptions of phenomenal consciousness *it is*. For constitutive of (phenomenally conscious) experience itself is a relation, be it a relation between two things or a relation with a reflexive structure. Given this relationality,

[11] This sort of observation—that it is possible for it to seem to one that she is consciously seeing or feeling, when she is in fact not—is sometimes used as basis for an influential objection to extrinsic intentional theories. These objections typically assume that these sorts of inner-illusions are impossible *a priori*. For they assume that the reduction of phenomenal consciousness to a kind of "consciousness of" is either unintelligible or, at least, completely unmotivated (cf., Finkelstein 2003, pp. 22–23 and Siewert 1998, chap 6.3).

the possibility of misrepresentation arises. Henceforth, in order to distinguish this possibility of a misrepresentation that is built into experience itself from misrepresentation that arises out of introspection, I will call the former *inner-illusion* and the latter *reflective misrepresentation*.[12, 13]

Now that we can see how possibility of inner-illusion arises, we can more easily understand the difference between traditional SR and disjunctive SR. Again, in a formula:

> **Disjunctive SR** A subject *S* at time *t* is a phenomenally conscious subject if and only if (a) *S* has a mental representation M* of herself as being in an intentional state M and: *either* it is the case that (b) *S* in fact has M and (c) M* is *dependent* on M; *or* it is the case that (d) *S* does not have M, but it merely seems to her as if she does.

With this formula, we can see that Disjunctive SR splits the difference between HO and traditional SR. For where traditional SR denies the that inner-illusion can occur at all, disjunctive SR (by condition (d)) agrees with HO in allowing its possibility. But where HO abandons the idea that M* is dependent on M, SR (by conditions (b) and (c)) retains it. For the disjunctive SR view postulates that the capacity of inner-sense, when successfully actualized, produces higher-order representational states which are dependent on the first-order mental states they represent. But disjunctive SR also maintains that certain non-successful actualizations can produce cases of inner-illusion, which HO takes to be the core of phenomenal consciousness.

[12] It is helpful to note that the intentional extrinsic conception of consciousness has impressive historical precedence. It is obviously present, for instance, in the Phenomenological tradition coming out of Brentano and Husserl, where there is a commonplace appeal to the notion of *pre-reflective* self-consciousness in order to articulate the categorical distinction between conscious and unconscious mental phenomena, as well as to distinguish the "consciousness of consciousness" constitutive of experience from the "consciousness of consciousness" constitutive of introspective judgment. See, e.g., the discussion see the discussion in (Kriegel and Williford (ed) 2006) and in (Gallagher and Zahavi 2010).

[13] David Finkelstein (2003, p. 23) objects that this response doesn't answer anything. For it only replaces the claim that phenomenal consciousness is a "consciousness of" one's own mental states with the claim that it is a "particular kind of consciousness of" one's own mental states. Moreover, Finkelstein argues, the difference between conscious and unconscious experience "cannot be understood as the difference between learning a set of facts by one mode of perception rather than another" (p. 24). For then the extrinsic theorist would be committed to the absurd claim that what is lacking in blindsight is the "phenomenology" that inner-sense itself provides us. I respond: Formulated that way, I concede that the extrinsic intentional theory does not explain anything. However, this also betrays Finklestein's misunderstanding of the extrinsic view. For it does not, as Finkelstein claims, hold that "Each mode of perception provides us with phenomenology as well as information" (p. 24.) It is, rather, that each mode of perception—vision, touch, taste, smell, hearing—provides us *only* information without "phenomenology." The "phenomenology" is added to the information by the actualization of another capacity, the capacity of inner-sense. That is the whole point of the extrinsic conception of phenomenal consciousness. The "phenomenal character" of an intentional state is an extrinsic property of the state: it is something that the state has by virtue of its co-instantiation and connection with something else. In other words, the key theoretical commitment of the extrinsic theorist is that phenomenal consciousness is *not*—as is often assumed—fully dissociable from the rest of the mind and the world. For other formulations of this point in response to objections similar in character to Finklestein's see Brown (2010) and Weisberg (2010).

5 The possibility of inner-illusion

How do traditional SR theorists motivate their view? The most common motivational factor—which traditional SR shares with intrinsic conceptions of phenomenal consciousness—is the intuition that the distinction between appearance and reality simply gets no grip on phenomenal consciousness, and any attempt to drive the distinction into this domain saddles the view with incoherence (cf., Block 2011; Levine 2001, pp. 108, 168). However, where the intrinsic theorist takes this point as an indication that there "really is something about our conception of the property itself, the pain itself, that makes it essentially a mode or kind of experience" (Levine 2001, p. 108), the extrinsic SR theorist simply takes it as an indication that there is a relation of dependence between the higher-order awareness and the qualities of the first-order mental state that it represents, i.e., a relation wherein, necessarily, if the higher-order state exists, the first-order state it represents exists (cf., Kidd 2011, p. 364). But, either way, the result is that inner-illusion is impossible (cf., Gennaro 2006, p. 242; Kidd 2011, Sects. 4 and 5).[14]

Now, it seems to me that if we can keep certain distinctions in view, commonsense intuition can be interpreted as committed to the idea that the appearance-reality distinction is applicable to conscious experience. It is unexceptionable to claim that we reflectively misrepresent what our own beliefs and desires actually are.[15] However, the majority of philosophers (but fewer psychologists) balk at the idea that we can also sometimes be wrong about the qualitative character of our own occurrent sensuous experience.[16] Why is this? There are probably as many answers to this question as there are philosophers. Nevertheless, many of them probably can be classed under one of two headings: either one is assuming an intrinsic conception of phenomenal consciousness or one is an extrinsic theorist that assumes that inner-sense is infallible. Given my project, we can set aside arguments against the intrinsic conception of consciousness and focus only on arguments against the infallibility of inner-sense.

[14] While I still believe that the token reflexive structure articulated in the cited paper can help SR intelligibly deny inner-illusion, I no longer believe that a theory of phenomenal consciousness *must* deny such cases in order to account for the epistemic authority of the first-person perspective. The relation of this paper to my earlier work can be seen as follows: in the earlier paper, I attempted to show that SR can both maintain the epistemic privilege of phenomenal consciousness by denying the possibility of inner-illusion, while maintaining its neutrality between naturalistic and non-naturalistic accounts of phenomenal consciousness. In this paper, however, I attempt to show that SR can also save the epistemic authority of the first-person without denying the possibility of inner-illusion. Therefore, if the conclusions of this paper are correct, SR maintains key explanatory advantages over both the extrinsic HO and the intrinsic views. For, unlike these, the SR theorist has available two views of the metaphysical structure of phenomenal consciousness that are compatible with the epistemic privilege of phenomenal consciousness: the earlier model, which denies the possibility of inner-illusion, or the model put forward in this paper, which acknowledges the possibility of inner-illusion. Whereas the extrinsic HO and intrinsic models are limited to one option each. Given our rather paltry knowledge of the nature of phenomenal consciousness at this stage of philosophical and empirical research, I take this neutrality to be a virtue.

[15] Cf., Nisbett and Wilson (1977), which documents the surprisingly commonplace tendency we have to make *ex post facto* sense of our own behavior by self-attributing beliefs and desires that allow us to appear in a favorable light to ourselves and to others, even if these beliefs and desires are deeply delusory.

[16] Cf., Kripke (1980, p. 151).

The first argument against the infallibility of inner-sense is to point out the incompatibility between the postulation of infallible cognitive mechanisms—mechanisms that necessarily function in a successful manner—and the view that cognition is normative, i.e., that it is intelligible to hold cognitive mechanisms to standards of right functioning that they may violate. If a cognitive mechanism is infallible, then there is no room for mistake. But where there is no room for mistake, as Wittgenstein famously pointed out, there is no room for being *right*. For there is no place for normative notions to get a grip. So, if inner-sense is to be a source of evidence for introspective judgment, this cognitive mechanism must be somehow conceived as subject to normative/cognitive evaluation. And this requires that either we follow the expressivists (who follow Wittgenstein) in taking "inner-sense" not really to be a "sense" at all, but rather a capacity to (somehow) spontaneously determine the phenomenal quality of our own experience, or we take inner-sense as a fallible receptive capacity.

Now, I take the idea that we somehow spontaneously constitute the phenomenal quality of our sensuous experience to be objectionable on phenomenological grounds. While it seems acceptable to claim (following Moran 2001) that I know *what I believe* in a privileged manner simply because my act of making up my own mind, i.e., of *committing myself* to some idea, is a constitutive moment of all belief, it is not acceptable to claim that I also make up the sensuous character of my perceptual experience in the same way. No matter how hard I try, I simply cannot make myself actually *feel* pain by committing myself to the idea that I feel pain. Rather, it seems to be a part of the very concept of phenomenally conscious perceptual experience that its being *receptive* entails that I, as the subject, am *passive* with regard to its content.

This much proves that inner-sense cannot be infallible in the sense of being a mechanism for which it is metaphysically (or logically) impossible that it function unsuccessfully. But it does not answer the claim that it may just be, not metaphysically, but *medically* impossible for inner-sense to function unsuccessfully. That is, the foregoing argument does not show that it is possible for beings *like us*, in the actual physical universe, to suffer inner-illusion. To answer this, consider two cases, often used by HO theorists to motivate the possibility of inner-illusion.

> **Case 1: The Dull Headache** Suppose you are suffering a dull all-day headache, but experience a brief period of respite, where you do not "*feel* the pain", due to some distraction. Yet, after the excitement of the distraction subsides, the feeling of the headache returns. It would be strange to consider the headache to have disappeared altogether from your mental life in the intervening moments. Rather, common sense would claim that in the meantime you simply did not *feel* the pain that was still there; for it was crowded out by the intense excitement of the distraction.
>
> **Case 2: Dental Fear** Imagine that you are one of the unfortunate dental patients who, even when sufficiently well anesthetized, nevertheless (seem to) feel pain in their tooth when the drilling begins. Fortunately, after you are reassured that the anesthesia is in full effect, you no longer feel pain

> under the drill. However, you still remember that, a moment ago, it certainly seemed to you that you felt pain. Here again, it would be strange to claim that you actually *did* feel pain under the drill the first time (say, simply because it seemed to you that you did), but that it went away after you were reminded that you were anesthetized. For whether you believe that you are anesthetized or not, you *are* anesthetized. And so your capacity to feel pain at all has been undermined, regardless of what you believe about what you feel. Therefore, whatever pain you "feel" will not be real, but either *illusory* (in the case where a change in your beliefs or in what you recall that you believe won't make a difference) or *delusory* (in the case, imagined here, where a change in your beliefs will make a change to how things phenomenally seem to you).

If these cases are properly interpreted, then inner-illusion is possible. But the deniers of inner-illusion are not at the end of their rope yet. For instance, the dull headache case can be considered as a case where the phenomenon of pain recedes so far into the periphery of attention that the subject simply "forgets" it is there. It is still there, still experienced by me; I simply do not notice it, just as I still hear but do not notice the hum of the refrigerator when absorbed in a good book. And the false pain under the drill can be explained away simply as a case where the subject actually does not feel pain or "seem to feel pain," but simply introspectively misidentifies the phenomenal quality of her experience under the drill—say, by mistaking the feelings of vibration and pressure for the feeling of searing pain. What the doctor's reassurance does, then, is help the subject to introspect more carefully the second time around.

Now, of these two responses, the interpretation of the brief respite from the all-day headache seems the more plausible; and it is probably the right interpretation for most cases of this phenomenon. But this does not mean that this is what always happens when headache pain subsides. In fact, neuroscientists have demonstrated that certain general anesthetics have no effect on the activities associated with pain in the lower-level brain regions, but that, owing to their effect on higher-level brain regions, the subject is not aware of being in pain at all (cf., Flohr 2000). If this is happening with general anesthetics, why could it not happen in the case of the all-day headache? Concerning the no-inner-illusion interpretation of dental fear: this seems to me very implausible. For, as Rosenthal (2005e, pp. 209–212) argues, such cases need not always be categorized as cases of introspective misattribution, and many cases are most plausibly not so categorized.[17] When the subject is under the drill, she is not already reflecting on the experience, puzzling over which concepts to apply in describing how it actually feels. Rather, there is just a searing burst of pain. Now, certainly this searing burst is not an experience of actual pain. For, by medical decree, she can have no such experience. But it also not a reflective judgment about a kind of experience that the subject finds antecedently puzzling. So it is most plausibly construed as one of those rare cases where the experience itself is a misrepresentation of what actually passes through the subject's mind—as

[17] See also (Rosenthal 2005a, pp. 138–139, 2005c, pp. 38–39).

Rosenthal, an HO theorist, puts it, it is one of those case where there is "something it's like to be in a state that one is not actually in" (Rosenthal 2005d, p. 209).

6 Disjunctive SR and the possibility of inner-illusion

If the arguments of the foregoing section are correct, then intentional extrinsic theorists must acknowledge the possibility of inner-illusion—a kind of illusion that is due to a misrepresentation on the part of the higher-order mental state, which is constitutive of phenomenally conscious experience itself. However, the HO theory, which accommodates inner-illusion by construing higher-order mental states as having *no* constitutive connection to the lower-order states they represent, cannot vindicate the first-person authority attendant to phenomenally conscious experience. And this is a burden for HO, since first-person authority is typically taken as a key characteristic of phenomenally conscious experience (cf., Siewert 1998, chap 1). Thus, I will present and begin to motivate the disjunctive SR view by answering the following question: is there a view that can acknowledge the possibility of inner-illusion while maintaining the epistemic authority accorded to the subject of phenomenally conscious experience? Since one way to recover epistemic authority in experience is to attribute a token-reflexive structure to higher-order representation, which entails the dependence of the higher-order representation on its lower-order object, i.e., to accept the key commitments of the traditional SR view (see Sect. 4), the question becomes: is it possible to formulate an SR view that is compatible with the possibility of inner-illusion? Again, I argue that disjunctive SR fits the bill.

One important point concerning the formulation of an SR theory is that if SR is to be different from HO, the *dependence* of higher-order awareness on first-order mental life should bear the force of metaphysical necessity. Thus, if M* is a higher-order state that is *dependent* on M, its first-order object, then, in any possible world where M* exists, M must exist as well.[18] HO rejects the dependence of M* on M as incompatible with the possibility of inner-illusion. Instead, HO theorists assert that there is merely a contingent connection between the two mental states wherein M is, or is a part of, the efficient cause of M*, which represents M; and that M* represents M in part by virtue of M's being (part of) its efficient cause (cf., Rosenthal 2005e, p. 29). To differentiate itself from HO, SR must postulate a connection between M* and M that prevents M*'s being a "distinct existence" from M.[19] I will call this M*-to-M dependence.

[18] For further discussion of this kind of SR view, and an argument that it is compatible with a naturalism, see Kidd (2011, Sects. 5–7). See also Thomasson (1999) and Simons (1982) for general discussion of the concept of dependence in ontology.

[19] This is not to say that SR theory must conceive M* as numerically distinct from M. It is still open to conceive of the representational relation as token reflexive, so that M is a state whose intentional content has two aspects: one which represents something other than M, the other which represents the state M itself (Smith 1986). However, since this strict self-representational view makes it logically impossible for the token-reflexive higher-order representational content to come apart from the existence of its object

However, this is only the first step. For M*-to-M dependence by itself entails the infallibility of inner-sense. By hypothesis, generally, if M* (a higher-order representation) exists, so does M (its lower-order object). As a consequence, it is necessarily true that all instances of higher-order awareness in the unity of phenomenally conscious states are veridical. Therefore, it must be the case that the mechanism productive of these states—the mechanism of inner-sense—is infallible, since it is the definition of "infallibility" that it is metaphysically impossible for any infallible cognitive capacity to represent falsely or non-veridically (cf., Armstrong 1963, p. 417).

What we need, then, is a view of the nature of higher-order awareness (M*) that somehow cleaves M*-to-M dependence from an attribution of infallibility to the mechanism that produces higher-order states (inner-sense). A way to do this is suggested by recent disjunctivist theories about perceptual experience. The fundamental idea of disjunctive views of perception is that it is possible for two perceptual experiences to be instances of different *most specific kinds* of mental states without there being *any* qualitative (or subjectively discernible) difference between the two (cf. Martin 2006, p. 357ff and McDowell 1998, p. 240ff). For instance, suppose that someone is veridically seeing a lemon one moment (Case 1) and a piece of soap, carefully carved to look exactly like a lemon, the next (Case 2) without any discernible difference *for the subject* in each. Case 1 and Case 2 are, as one might put it, *phenomenologically* equivalent, but *metaphysically* inequivalent. Disjunctivists maintain that such cases are possible. And so they deny that the "*phenomenology*" of perception—understood in the sense of a *complete* description of the "what-it-is-like for the subject" of the perceptual state—fully determines the *essence* or *nature* of a perceptual state. Instead, the phenomenological description of a perceptual experience is taken to underwrite a disjunctive ontological assay, such as,

> **Disjunctive Metaphysics of Perception** This experiential state as of X (for me) is either a genuine (veridical) perceptual awareness of X or an illusory (non-veridical) awareness of X.

This, in effect, makes phenomenological description neutral between two possible construals of what the state with the given phenomenal character actually is: either a state that is what it phenomenally seems to be (a "genuine" or "veridical" perceptual awareness) or not (an "illusory" or "non-veridical" awareness).

SR can incorporate the disjunctive paradigm by the following two steps. First, generalize the disjunctive theory of perceptual experience to *all* experiential states. Second, explicate the possibility that metaphysically different states can bear equivalent phenomenologies by reference to differences in both the first-order state one is aware of oneself as being in—a genuine (veridical) or illusory perceptual state, for instance, or in the content of the higher-order awareness that makes one

Footnote 19 continued
(since whatever represents *itself* must exist), I think its better to avoid this way of construing the relation. For discussion of other problems with the strict self-representational view, see Kriegel (2006).

phenomenally conscious of first-order mental states. *Every phenomenological description of experience*, then, is to be ontologically assayed in the following terms:

> **Disjunctive Metaphysics of Phenomenal Consciousness** This experiential state is *either*:
>
> (i) a *standard* case of higher-order awareness that is dependent on the first-order (perceptual, emotional, judgmental) state it represents (a case of M*-to-M dependence), *or*
> (ii) a *non-standard* case of higher-order awareness that is *not* dependent of the first-order state it represents.

Thus, if the experience is due to a standard case of higher-order awareness, then, necessarily, the first-order state it represents exists, and the state of higher-order awareness is veridical. However, if the experience is due to a non-standard case of higher-order awareness, then the first-order state that the higher-order state represents might not exist, and the state of higher-order awareness might be radically non-veridical. So this model delivers what we seek: a way to retain the dependence of higher-order awareness on the first-order state it represents while denying the infallibility of the mechanism that is productive of the relevant sort of higher-order states. The mechanism is fallible because it is possible that it might not produce a standard state of higher-order awareness. But, in those cases where it does produce standard states of higher-order awareness, the states are necessarily veridical. The kind of "infallibility" recovered here, if we are inclined to speak in this manner,[20] is at best a *contingent* infallibility.

A hint in ordinary language that leads to this disjunctive SR analysis of phenomenal consciousness is provided by consideration of the word "fallible" itself.[21] "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* or *mechanisms*, not (as is sometimes implied) a feature of cognitive states themselves, which are the products of cognitive capacities or mechanisms. Given this understanding of the concept of fallibility, one can win the priority of dependent cases of higher-order awareness over non-dependent cases by construing the dependent states as the products of a *successful* execution of the capacity for higher-order awareness, and the non-dependent states as products of an (in some way) unsuccessful execution of the capacity. In other words, this takes dependent states of higher-order awareness as logically more fundamental than non-standard cases. For dependent states are the "standard" by which one judges the *aptness* of the actualization of inner-sense; and this, in effect, forces the construal of all inner-illusion as the product of actualizations of inner-sense that fall short of the standard. In this picture, the postulation of (at least) two different kinds of states of higher-order awareness is motivated by the analysis of non-standard cases as

[20] As I was at Kidd (2011, p. 373).

[21] This tactic is adapted from the work of John McDowell, who applies similar considerations in motivating the priority of veridical perception over their non-veridical counterparts. See especially McDowell (1982, 2011).

failures of the mechanism of inner-sense. These two kinds of states of higher-order awareness can be alike in that they both underwrite phenomenal consciousness, even a phenomenologically indistinguishable phenomenal consciousness. But the two states differ in that one is the product of a completely successful exercise of inner-sense, and so is dependent on the state it represents, but the other is not dependent on the first-order state it represents.

HO, on the other hand, postulates no priority for dependent states, and so it takes the standards for judging the aptness of inner-sense to be lower, for it does not require dependence or even representational accuracy. Rather, in the HO view, inner-illusion is just another product of a normally functioning capacity of inner-sense.[22]

7 First-person authority and the non-grippingness of internal-world skepticism

What reason is there to prefer a view that takes the production of dependent higher-order states as the standard for the successful functioning of inner-sense? The argument I explore in the rest of this paper is: without taking dependent higher-order states as the standard for the successful actualization of inner-sense, there is no way to *vindicate* the intuition that our introspective judgments about our own occurrent sensuous states have an epistemic authority that no third-person judgment about sensuous experience can have. In particular, introspective judgment about my own *conscious and occurrent mental states* (henceforth: COMS) are not susceptible to global skeptical doubt in the way that judgments about the "external world" are.[23] I will call this insusceptibility of COMS to global skeptical doubt the *non-grippingness* of *internal-world skepticism*. I will argue for this by first showing that the non-grippingness of internal-world skepticism is the key commitment behind our concept of first-person authority, and that this commitment *is* consistent with the idea that inner-sense is fallible, given the disjunctive SR conception of phenomenal consciousness. In the next section (Sect. 8), I'll complete the argument against HO by reference to a version of the argument from illusion that attempts to motivate internal-world skepticism. I'll show that HO, given its insistence on the independence of higher-order and first-order awareness, cannot rebut this argument. If this is correct, then disjunctive SR has a distinct explanatory advantage over HO.

It is not difficult to turn a group of perfectly normal undergraduate students into a company of external-world skeptics. As Descartes shows, artful application of certain doubts concerning the reliability of the senses, dream scenarios, and all-powerful deceivers usually does the trick. We consider whether it is possible that all of our beliefs about the external world are caused by an all-powerful Deceiver and, in light of the apparent lack of any subjectively accessible criteria on the basis of which to rule out such hypotheses, it seems to be a possibility (however remote and

[22] Rosenthal offers reasons in support of this idea at Rosenthal (2005d, p. 29) and the other works he references there.

[23] This argument is inspired by Horgan et al. (2006).

idle) that all our beliefs about the external world are radically and systematically false.

However, as Descartes noticed, there is a striking asymmetry with the prospects of motivating a global *internal* skepticism using similar methods of doubt. In other words, the possibility that *all* of our immediate, non-inferential, introspective beliefs about our own current experiential states are radically mistaken is not a possibility that is awakened in our imaginations with ease. Rather, these considerations, in light of the fact that our awareness of ourselves as being in a certain mental state, if anything, *seems* certain, do not move us to a global skepticism. A *local* skepticism, of course, is not as difficult to motivate, especially if one is open to the possibility of inner-illusions motivated in Sect. 5. But what seems to verge on unintelligibility is the claim that, in every case of phenomenally conscious experience, we are radically and systematically mistaken about the quality of our experience.

I do not take this fact about folk psychological intuition to be a brute fact. Nor do I take it to be a self-evident truth. Instead, I take it to be a pre-theoretical commitment in our everyday practice of evaluating introspective beliefs about COMS. More specifically, this commitment seems to be a kind of ontological commitment about the *objects* of introspective COMS beliefs, viz., conscious occurrent mental states themselves. If this is so, then the question of what it is about introspective COMS beliefs that makes them so special becomes: what is it about phenomenally conscious mental states that makes occurrent introspective belief about them so special, such that a global skepticism about these kinds of beliefs does not get a grip on us as easily as (non-introspective) beliefs about the external world? What explains this fact?

The answer I attempt to motivate here is: the best explanation comes in the form of the disjunctive SR view, i.e., because conscious human subjects are endowed with a capacity of inner-sense that, when functioning successfully, produces higher-order mental states that are dependent on their first-order objects. Now, for this argument to the best explanation to be clear, we must first discern what the phenomenon of "first-person authority" is. The traditional answer to this is that our knowledge of our own COMS is authoritative insofar as the very fact that I have a certain COMS is all that I need to justify my introspective belief about it and its qualities. There is, in other words, no other justification for a belief about a particular COMS needed aside from citing the COMS itself. Not all beliefs are like this. Most of our beliefs about the "external world" require that we cite something *other* than the content of the very belief we are attempting to justify. For instance, suppose I assert that I know the President is currently on vacation or that the Rangers will win the American West division title. If I am asked to justify these beliefs, then to respond simply with the contents of these beliefs—"the President is on vacation" or "the Rangers will win the title"—could only be taken as a smug indication that I *don't* in fact know what I claim to know. However, in those rare cases where justification for introspective belief about my conscious occurrent mental state is requested—e.g., "How do you know that you have a headache?" or "How do you know that you *believe* the Rangers will win the West?"—it seems that all I can do in response is cite the content of my belief—"Because I have a headache" or "Because I believe the Rangers will win the West." Again, what

seems to underlie the authority of introspective belief about occurrent conscious mental states is that the subject of these states needs only cite the content of their belief in order to justify holding it; indeed, this seems all that they can do (cf., Siewert 1998, p. 14 and Neta 2008, Sect. II).

Now, the self-citing nature of the warrant for introspective COMS belief indicates why internal-world skepticism does not get a grip on us in the way that external-world skepticism does. For to motivate the conviction that all of one's beliefs about occurrent and conscious experience are radically mistaken, one must somehow undermine the self-citing evidentiary support of introspective COMS belief. However, this is difficult to do. For, as Descartes realized, a defeating consideration from outside of experience—such as raising the hypothesis that an all-powerful deceiver is undetectably manipulating your experience in a way that radically misleads you about your first-order mental life—would not get a grip on introspective COMS beliefs, so long as one still has the occurrent experiences that the introspective COMS belief is about. The connection between introspective COMS beliefs and their evidentiary basis is such that the warrant for COMS beliefs cannot vary while the content of the belief remains the same, as can happen for belief about the external world. For there is no distinction in introspection between the content of the belief and what warrants its affirmation.

At this point, however, the question becomes: why should this self-citing nature of introspective warrant be considered *legitimate*? Why should we not consider the dearth of separate evidence for introspective belief a mark of epistemic *deficiency* rather than a mark of epistemic *privilege*? With this, the question turns from an epistemological question concerning the kind of evidence or warrant for introspective COMS belief—i.e., the task of articulating what makes introspective COMS belief descriptively and normatively different from other kinds of belief—to a metaphysical question concerning the nature of conscious experience, i.e., *explaining* what conscious experience is such that we can know it in this self-citing way. The question, in other words, concerns what it is that *vindicates* the difference between our treatment of the warrant of introspective COMS beliefs and beliefs about the external world.

One metaphysical position that vindicates this difference is the rejection of the appearance-reality distinction in phenomenal consciousness. On this view, necessarily, every aspect of the conscious subject's experience appears to her exactly as it is, in *every* case of conscious experience. However, this is just a return to the no inner-illusion views that we have found good reason to reject above.

The other metaphysical conception of conscious experience is offered by the disjunctive SR theory. In this view, it is *only* in the context of a *standard case* of higher-order awareness, which is the product of the *successful execution* of the mechanism of inner-sense, that our first-order mental life appears exactly as it is. And since this "appearance" of consciousness to the subject is grounded in a higher-order *representation* that is dependent on its lower-order object in a way that guarantees its veridicality (Sects. 4 and 6 above), all the subject needs to do is *replicate* or literally *take up* this content into the content of an introspective

judgment.[24] This picture preserves everything needed to vindicate the self-citing nature of the evidence for introspective COMS beliefs. For it preserves the idea that, in cases where inner-sense functions successfully, our experience appears to us just as it is. And so any claim that aims only to report this appearance is guaranteed, in that circumstance, to be true. Now, as mentioned, this still leaves room for introspective error and inner-illusion. For the mechanism for inner-sense *is* fallible, and so there may be cases where it delivers an appearance that does not manifest the actual layout of our first-order mental life (e.g., a respite from pain where pain is still present, or the manifestation of pain where there really is none). But, unlike HO, it retains the idea that in normal experiential circumstances our own phenomenally conscious experience provides indefeasible warrant for introspective beliefs about it. All that needs to be done is to cite the experience that is already cited in the formulation of the introspective belief.

8 The argument from inner-illusion

If the reflections of the foregoing section are correct, then a complete theory of phenomenal consciousness must provide an explanation for the self-citing nature of the warrant for introspective COMS beliefs which vindicates the legitimacy of this kind of warrant. I then argued that disjunctive SR can provide a vindicating explanation in the form of a theory of the relation between our lower-order mental states and the higher-order mental states that provide awareness of them and their qualities. I will now present an argument that motivates disjunctive SR over HO: that HO's account of this relation cannot possibly provide an explanation of self-citing warrant that also vindicates its legitimacy. HO, in other words, would force a subject to have to look for further evidence to supplement the warrant provided by phenomenally conscious experience itself.

The crux of my argument is the fact that it follows from the HO account that a case of inner-awareness that accurately represents first-order mental life, considered by itself, is no different from a case that does not. In the HO view, veridical cases of higher-order awareness are *only contingently* related to the first-order states they represent. For HO holds that a higher-order state M* and a lower order state M that M* represents are *distinct existences*; they are not dependent on each other, but M only is typically part of the efficient cause of M*. It is also possible for M* to be brought about by a chain of efficient causation that does not involve M. So, the essential make-up of M* is independent of its lower-order representational content. Therefore, by HO, a phenomenally conscious experience *by itself* provides no grounds for holding any given case of experiential awareness of one's first-order mental life to be veridical. Thus, there is also no reason to hold to the self-citing

[24] The kind of thing I have in mind for the method of "taking up" the content of a higher-order awareness into the content of a representational state of an even higher-order level is a theory of introspective self-knowledge that exploits the semantics of indexicals (roughly, securing the reference of introspective judgment by reference to "*this*" experience) such as that found in Davidson (1987) or Burge (1988).

conception of the warrant for introspective COMS beliefs, and so there is no reason to accept the intuitive non-grippingness of internal-world skepticism.

To make this clearer, consider the following argument (which is meant to be a variant on the argument from illusion for perceptual knowledge of mind-independent objects):

1. 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 being in M.
2. Since, according to HO, M* and M are distinct existences, the same most specific kind of state M* can exist as a veridical or non-veridical representation.
3. Thus, by (2), it is also possible for S's introspective belief that S (herself) is in M to be false, while S is still experientially aware of herself as being in M.
4. And thus, by (2), it is possible for *all* S's introspective beliefs to be false while S (herself) is experientially aware of herself as being in the state she affirms being in.

Premise (1) is a restatement of a key *explanandum* for both HO and SR, an *explanandum* that is presented by the possibility of illusory inner-awareness. Premise (2) articulates the consequence of the metaphysical hypothesis that HO rallies as an *explanans*. From this (3) and (4) follow. Since (4) is equivalent to the claim of the internal-world skeptic, if this argument is correct, it is the case that HO entails the denial of the non-grippingness of internal-world skepticism.

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

2* Since, according to SR, M* is dependent on M, the same most specific kind of state of inner-awareness M* cannot exist as both a veridical and non-veridical representation.

This blocks the way to the common-factor claim in (3), which, in turn, leads us away from the skeptical consequence.

9 Two objections

9.1 'Debunking' explanations

One way for HO to resist the foregoing argument is to accept a *thoroughgoing* fallibilism about inner-awareness and resorts to a "debunking" explanation of non-grippingness of internal-world skepticism: to wit, an explanation that treats it as a pervasive error, due to a persistent cognitive illusion to which human subjects are prone.[25] The debunking theorist does not deny the psychological fact that we are not

[25] See Schwitzgebel (2008) for clear and forceful motivation of the debunking approach. I get the term "debunking" explanation from Horgan et al. (2006).

gripped by the prospects of internal-world skepticism; she only denies that we ought to take this as grounds for maintaining the legitimacy of the notion of privileged access to experience, so that, even though we are not inclined to reject the privileged accessibility of experience, we rationally *ought* to. In other words, this theorist would not deny that non-grippingness of internal-world skepticism is a clear intuition of common sense. But commonsense intuitions are, as Rosenthal says, "data, not self-evident truths" (Rosenthal 2005b, p. 9). As such, they are to be treated as *explananda* of a theory of consciousness that can be subjected to debunking treatment.

To answer this objection fully would require more space that I have available here. So let me register what I take to be sufficient reason to proceed with disjunctive SR without a complete answer. Insofar as we have a viable *non*-debunking explanation of non-grippingness—i.e., an explanation that vindicates the idea that we have a privileged form of access to our own experiential life, which is compatible with all the data that the debunking explanation accounts for—then the burden of proof lies on the side of the theory that is further from commonsense intuition (cf., Horgan et al. 2006, p. 43). This is not to treat the deliverances of common sense intuition as a self-evident truths. It is, rather, to put the point in Sellarsian terms, to assert the superiority of a "proto-scientific" explanation of consciousness that brings more of the manifest image into a "stereoscopic" unity with it than other theories do. That is, it is to assert the widely accepted claim that an explanatory psychological theory that can vindicate key data points from folk psychology without explanatory loss on any other data points (be they "folk-theoretic" or not) is more fruitful than one that does not. And since, *ceteris paribus*, a more fruitful explanatory theory is to be preferred, the non-debunking explanation that the disjunctive SR theory provides enjoys an explanatory edge over HO.

9.2 A distinction without a difference

Another objection is that (2*) does nothing to alleviate the force of skeptical considerations about introspective COMS belief. The idea is that since the difference between a standard state of higher-order awareness, from which we can attain indefeasible warrant for introspective COMS beliefs, and a non-standard state, from which we cannot, is *not* a difference that we can *represent to ourselves in introspection*, it also makes no difference to the skeptical threat posed by the argument from inner-illusion. Rather, it leads us to the same predicament facing HO. Since, given the fallibility of the mechanism of inner-sense, the subject's experiential awareness of herself as being in a certain mental state does not, by itself, issue the subject full assurance that inner-sense has functioned successfully, one might, in any given case, have a non-standard case of experience that experientially *seems* just like its standard counterpart. So, without the guarantee that the mechanisms of inner-awareness work successfully, there is no guarantee for the truth of our introspective belief.

However, HO and SR are not on equal epistemological footing here. To see why, consider a distinction Quassim Cassam (2007, p. 2) draws between two ways one might save the possibility of knowledge in light of (purported) obstacles to it, such

as we encounter in the indiscernibility of standard and non-standard cases of experience. One way employs an "obstacle-*overcoming*" strategy. This strategy begins by acknowledging that there is a genuine obstacle to knowledge, and then attempts to show how we are equipped to overcome it. The other way employs an "obstacle-*dissipating*" strategy, which, unlike the obstacle-overcoming strategy, starts from a view of our cognitive capacities that entails the non-existence of the obstacle on which the skeptical challenge depends. Thus it saves knowledge by showing the skeptical obstacle to be a sham.[26]

HO seems destined to employ an obstacle-overcoming strategy. Since this theory stipulates the complete independence of inner-awareness and the first-order states they represent, to save the epistemic privilege of introspective COMS belief, the HO theorist must specify what *else* needs to be added to inner-awareness for it to provide warrant that also guarantees truth. The disjunctive SR theorist, on the other hand, despite the lack of an introspectively discernible criterion of success, is free to abandon the obstacle-overcoming project by adopting the obstacle-dissipating strategy. Since a state of higher order-awareness is dependent on the lower-order state it represents, there is no place for the worry that our introspective beliefs are radically and systematically false to take root: insofar as we are experientially aware of our first-order mental life, and this awareness is grounded in a standard state of higher-order awareness, the warrant provided by experiential awareness also guarantees the truth of our introspective COMS belief. This is not to say, of course, that one is *completely* free from all doubt about the epistemic adequacy of one's introspective beliefs—indistinguishable pairs of standard and non-standard states are not hereby banished. But one is free from the doubt that, even in a standard case of inner-awareness in experience, the experience by itself still falls short of delivering indefeasible grounds for introspective knowledge.

Acknowledgments I would like to thank Antonio Capuano, Walter Hopp, Kelly Jolley, Guy Rohrbaugh, David Rosenthal, David Woodruff Smith, Josh Weisberg, and Ken Williford for thoughtful comments and suggestions on earlier drafts of this essay. Versions of this material were presented at the Auburn University Philosophical Society and at the Fifth Online Consciousness Conference. I would like to thank those audiences for questions, constructive discussion, and feedback. I would especially like to thank Richard Brown for his helpful comments and for the opportunity to present this paper in this collection.

References

Armstrong, D. M. (1963). Is introspective knowledge incorrigible? *Philosophical Review, 62*, 417–432.

Block, N. (2002). Concepts of consciousness. In D. Chalmers (Ed.), *Philosophy of mind: Classical and contemporary readings*. Oxford: Oxford University Press.

Block, N. (2011). Response to Rosenthal and Weisberg. *Analysis, 71*(3), 443–448.

Brown, R. (2010). Deprioritizing the a priori arguments against physicalism. *Journal of Consciousness Studies, 17*(3–4), 47–69.

Burge, T. (1988). Individualism and self-knowledge. *Journal of Philosophy, 85*, 649–663.

[26] Cf. the distinction between Humean and Cartesian skeptical challenges to Naïve Realism about perception that one can derive from the "argument from illusion" brought out in Martin (2006, pp. 354–355).

Chalmers, D. J. (1996). *The conscious mind: In search of a fundamental theory*. Oxford: Oxford University Press.
Davidson, D. (1987). Knowing one's own mind. *Proceedings and Addresses of the American Philosophical Association, 60*(3), 441–458.
Dennett, D. C. (1991). *Consciousness explained*. Boston: Little, Brown, & Co.
Finkelstein, D. H. (2003). *Expression and the inner*. Cambridge: Harvard University Press.
Flohr, H. (2000). NMDA receptor-mediated computational processes and phenomenal consciousness. In *Neural correlates of consciousness*. Cambridge: MIT Press.
Fodor, J. A. (1983). *The modularity of mind*. Cambridge: MIT Press.
Gallagher, S., & Zahavi, D. (2010). Phenomenological approaches to self-consciousness. Stanford Encyclopedia of Philosophy. http://plato.stanford.edu/entries/self-consciousness-phenomenological/.
Gennaro, R. J. (2006). Between pure self-referentialism and the (extrinsic) HOT theory of consciousness. In U. Kriegel & K. Williford (Eds.), *Self-representational approaches to consciousness*. Cambridge: MIT Press.
Gurwitsch, A. (1985). *Marginal consciousness*. In L. Embree (Ed.) Athens: Ohio University Press.
Horgan, T., Tienson, J., & Graham, G. (2006). Internal-world skepticism and the self-presentational nature of phenomenal consciousness. In U. Kriegel & K. Williford (Eds.), *Self-representational approaches to consciousness*. Cambridge: MIT Press.
Kidd, C. (2011). Phenomenal consciousness with infallible self-representation. *Philosophical Studies, 152*, 361–383.
Kidd, C. (2014). Husserl's phenomenological theory of intuition. In L. Osbeck & B. Held (Eds.), *Rational intuition: Philosophical roots, scientific investigations*. Cambridge: Cambridge University Press.
Kriegel, U. (2006a). The same-order monitoring theory of consciousness. In U. Kriegel & K. Williford (Eds.), *Self-representational approaches to consciousness*. Cambridge: MIT Press.
Kriegel, U. (2009). *Subjective consciousness: A self-representational theory*. Oxford: Oxford University Press.
Kriegel, U., & Williford, K. eds. (2006). *Self-representational approaches to consciousness*. Cambridge: MIT Press.
Kripke, S. A. (1980). *Naming and necessity*. Cambridge: Harvard University Press.
Levine, J. (2001). *Purple haze: The puzzle of consciousness*. Oxford: Oxford University Press.
Lycan, W. G. (1996). *Consciousness and experience*. Cambridge: MIT Press.
Martin, M. G. F. (2006). On being alienated. In T. S. Gendler & J. Hawthorne (Eds.), *Perceptual experience*. Oxford: Oxford University Press.
McDowell, J. H. (1982). Criteria, defeasibility, and knowledge. *Proceedings of the British Academy, 68*, 455–479.
McDowell, J. H. (2011). *Perception as a capacity for knowledge*. Milwaukee: Marquette University Press.
McDowell, J. H. (1998). Singular thought and the extent of "Inner Space". In *Meaning, knowledge, and reality*. Cambridge: Harvard University Press.
McGinn, C. (1996). *The character of mind: An introduction to the philosophy of mind* (2nd ed.). Oxford: Oxford University Press.
Merleau-Ponty, M. (2012). *Phenomenology of perception* (D. A. Landes, Trans.). Routledge.
Moran, R. (2001). *Authority and estrangement: An essay on self-knowledge*. Princeton: Princeton University Press.
Nagel, T. (1974). What is it like to be a bat? *Philosophical Review, 83*(4), 435–450.
Neta, R. (2008). The nature and reach of privileged access. In A. E. Hatzimoysis (Ed.), *Self-knowledge*. Oxford: Oxford University Press.
Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*(3), 231.
Norman, L. J., Heywood, C. A., & Kentridge, R. W. (2013). Object-based attention without awareness. *Psychological Science, 20*, 1–8.
Rosenthal, D. M. (2005a). *Consciousness and mind*. Oxford: Clarendon Press.
Rosenthal, D. M. (2005b). The independence of consciousness and sensory quality. In *Consciousness and mind*. Oxford: Clarendon Press.
Rosenthal, D. M. (2005c). Introduction. In *Consciousness and mind*. Oxford: Oxford University Press.
Rosenthal, D. M. (2005d). Two concepts of consciousness. In *Consciousness and mind*. Oxford: Oxford University Press.

Rosenthal, D. M. (2005e). Sensory qualities, consciousness, and perception. In *Consciousness and Mind*. Oxford: Clarendon Press.
Russell, B. (1912). *The problems of philosophy*. Oxford: Oxford University Press.
Scheler, M. 1992. The idols of self-knowledge. In *Selected philosophical essays*, tran. David R. Lachterman. Northwestern University Press.
Schwitzgebel, E. (2008). The unreliability of naive introspection. *Philosophical Review, 117*(2), 245–273.
Siewert, C. (1998). *The significance of consciousness*. Princeton: Princeton University Press.
Simons, P. (1982). Three essays in formal ontology: Essay I. The formalization of Husserl's theory of parts and wholes. In B. Smith (Ed.), *Parts and moments: Studies in logic and formal ontology*. München: Philosophia.
Smith, D. W. (1986). The structure of (self-)consciousness. *Topoi, 5*, 149–156.
Thomasson, A. L. (1999). *Fiction and metaphysics*. Cambridge: Cambridge University Press.
Weisberg, J. (2011). Misrepresenting consciousness. *Philosophical Studies, 154*(3), 409–433.
Weiskrantz, L. (2009). *Blindsight: A case study spanning 35 years and new developments*. Oxford: Oxford University Press.