Philos Stud (2015) 172:2705–2733
DOI 10.1007/s11098-015-0441-1

Quotational higher-order thought theory

Sam Coleman

Published online: 14 February 2015

Abstract Due to their reliance on constitutive higher-order representing to generate the qualities of which the subject is consciously aware, I argue that the major existing higher-order representational theories of consciousness insulate us from our first-order sensory states. In fact on these views we are never properly conscious of our sensory states at all. In their place I offer a new higher-order theory of consciousness, with a view to making us suitably intimate with our sensory states in experience. This theory relies on the idea of 'quoting' sensory qualities, so is dubbed the 'quotational higher-order thought theory'. I argue that it can capture something of the idea that we are 'acquainted' with our conscious states without slipping beyond the pale for naturalists, whilst also providing satisfying treatments of traditional problems for higher-order theories concerning representational mismatch. The theory achieves this by abandoning a representational mechanism for mental intentionality, in favour of one based on 'embedding'.

Keywords Consciousness · Higher-order thought · Qualia · Representation · Self-representation · Acquaintance

For those who believe consciousness requires higher-order cognitive access to a mental state,[1] two major questions concern, respectively, the relationship between the accessing and accessed *vehicles*, and that between the accessing and accessed

[1] Cf. Gennaro (2012: 282). This *mental operation on a mental state* distinguishes 'higher-order' theories of consciousness from 'first-order' ones (Tye 2000; Dretske 1995) where a mental state is conscious due to its functional role, e.g. being poised to impact beliefs and desires.

S. Coleman (✉)
Department of Philosophy, University of Hertfordshire, de Havilland Campus, Hatfield, Hertfordshire AL10 9AB, UK
e-mail: S.Coleman@herts.ac.uk

contents.[2] Concerning the vehicles, especially interesting is the question how tightly bound the accessed state is with the accessing state. Is the accessed mental state wholly distinct from the accessing mental state, wholly identical to it, or somewhere in between (and if the latter, where exactly in between)? Concerning the contents, especially interesting is the question whether the accessing or accessed content dominates in fixing what the subject finds in consciousness.

The literature on higher-order approaches to consciousness features a superficial variety of positions on the vehicle issue, and an unacknowledged consensus on the content issue. This paper adopts novel positions on both issues, yielding a new higher-order theory of consciousness. Our dialectic springs from the problem of *mistargeted representations*. We see how this problem has generated consensus on the content issue, how this consensus makes the variety of responses to the vehicle issue superficial, and how this convergence is evitable by a different treatment of the difficulty. This leads to a higher-order theory whose distinctive feature, as well as dealing neatly with mistargeting, confers an important virtue lacked by its peers. A statement of this virtue at the outset will appear puzzling: the new theory manages to make our first-order sensory states[3] genuinely conscious. It will become clear the sense in which this is so, and why this virtue is peculiar to the new theory. A third major question for higher-order approaches concerns the *mode* by which the higher-order state accesses its first-order target. The consensus in the field is that this mode is *representation*: we will find reason to challenge this consensus too.

1. Say someone tokens a red London bus-ish visual percept. For Rosenthal's *higher-order thought* (HOT) *theory*,[4] what makes her conscious of that percept is having a thought represent it appropriately.[5] Without this HOT the red London bus-ish percept is doomed to lie dormant in the subject's visual cortex, its sensory qualities unexperienced.[6] HOTs are typically unconscious: what the subject is conscious of is what the HOT represents. For the HOT to be conscious requires it to be targeted by a still higher-order thought. This is Rosenthal's model of introspection.

2. This elegant development of a venerable theory of consciousness[7] has seen several criticisms, perhaps the most important being the *mistargeted representation problem*,[8] influentially aired by Neander.[9] In the good case, the red London bus-ish

[2] There is also the question whether higher-order states are thought- or perception-like. I consider this in §13. For discussion see Van Gulick (2000).

[3] I use 'sensory' broadly, including perceptual and proprioceptive contents.

[4] See Rosenthal (2005).

[5] The subject must not be aware of the state as inferred; the HOT represents *her* as in the relevant state; HOTs must be simultaneous with their targets (see Rosenthal 2005 for details).

[6] Even if the percept is accessible otherwise, e.g. by the subject's cognitive complex: Rosenthal's account of blindsight abilities is along these lines.

[7] Locke and Aristotle, to name but two, held something similar. See Caston (2002).

[8] There is also, notably, the 'rock' problem (Goldman 1993; Stubenberg 1998), but I lack space to examine it. One version asks how a mental state can be made conscious by our thinking about it when rocks aren't made conscious by our thinking about them. My answer is similar to that of Gennaro (2004, 2012) and Van Gulick (2004), but the reader must await the positive view (§§ 9–13) to glimpse how it works. The gist is that rocks cannot form part of the mental complexes needed for conscious states, being outside the head (not to mention rocky!). Consciousness of a state requires it to be 'neurologically grabbable'.

[9] Neander (1998). See also Byrne (1997).

percept is present while a HOT suitably represents there to be such a percept. But sensory state and HOT are wholly separate states for Rosenthal, creating the possibility of two sorts of slippage. In *misrepresentation* the percept is unfaithfully represented: e.g. the HOT represents a *blue* London bus-ish visual percept. In *targetlessness* the percept is missing, nonetheless the HOT represents there to be a red London bus-ish percept.

Rosenthal's challenge is what to say about such cases. Take misrepresentation: he could either say the subject experiences red London bus-ishly, or blue London bus-ishly.[10] If he gives the former response one wonders what the point of the HOT is in HOT theory. But if he gives the latter response, it seems to make the sensory component redundant in generating a conscious state. Concerning targetlessness, the question is whether a conscious state obtains in the absence of the sensory state, with similar dilemmatic options facing Rosenthal in reply. However he responds, Rosenthal seems left maintaining that one component of his HOT theory is surprisingly under-involved in the production of consciousness.

3. Rosenthal is clear about his preferred treatment of these cases. He explicitly entrusts HOTs with generating the 'subjective mental appearances' that comprise conscious episodes. His thinking is somewhat as follows. A conscious mental state is one the subject is aware of herself as being in. Since the best way to construe this awareness is in terms of representation, this means some mental state of the subject represents her to be in the relevant mental state she is aware of herself as having. For instance, a HOT represents her to be in a red London bus-ish visual state. But, with HOTs installed as the providers of subjective mental appearances, why should it matter whether the target state exists? One can surely be aware of oneself as being in a mental state even if there is no such state, as one can be aware as of a unicorn before one without need of a unicorn. How does it feel to be aware of oneself as instantiating a red London bus-ish percept when no such percept obtains? Conscious mental life is the stuff of appearances, Rosenthal emphasises: thus, since the appearance to the subject is the same whether or not she tokens a red London bus-ish percept (providing the same HOT is in place), her experience is indistinguishable across both cases—as hallucinations can be indistinguishable from their veridical counterparts. Corresponding remarks apply to misrepresentation: HOTs govern the subjective mental appearances constitutive of the subject's stream of consciousness.

4. Many commentators have felt dissatisfied with this response.[11] Surely, the sentiment goes, in this *higher-order representational* theory of consciousness,

[10] He cannot apparently deny a conscious state is produced by this combination of HOT and percept, since by hypothesis these sorts of combination produce conscious states, and the sheer fact that the HOT *represents* the sensory state creates room in principle for misrepresentation.

[11] Kriegel (2009: 131) claims Rosenthal's treatment of targetlessness violates the 'obvious truism' that there's something it's like for a subject only if she is in a conscious mental state—recall that HOTs are not standardly conscious, so absent also a first-order state the obvious truism seems contravened, as the subject still enjoys the relevant subjective mental appearances. Rosenthal might reply either by denying the truism, or accepting it but denying the conscious state the subject is in must exist. Kriegel assumes the subject being 'in' the state implies it exists, but this is simply to reject the position Rosenthal takes, where this 'in' has a more intentional flavour (in such cases, anyway).

shouldn't the represented lower-order content matter somewhat more to the content the subject is conscious of? Block has sought to focus these rays of dissatisfaction into a precise objection. Suspecting Rosenthal's approach ends up 'jettisoning the first-order content as constitutive of consciousness',[12] Block alleges an incoherence in HOT theory, due to Rosenthal's 'holding both that an appropriate higher-order thought is sufficient for a conscious state and that being the object of an appropriate higher-order thought is necessary for a conscious state'.[13]

A targetless HOT supports a conscious state, by the sufficient condition. Yet since it has no target, and we can imagine it is not targeted by another higher-order thought, there exists a conscious state without anything being the object of an appropriate HOT, violating the necessary condition. Block concludes HOT theory's necessary and sufficient conditions conflict, leaving it incoherent. But Rosenthal's theory is *not* vulnerable to this objection, as Rosenthal doesn't endorse the stated necessary condition. He is happy that lone, targetless HOTs supply the subjective mental appearances characteristic of consciousness. Block's necessary condition is really only necessary for a *pre-existing sensory state*: its only hope of entering consciousness is via HOT representation.[14] But HOTs are ultimately responsible for generating subjective mental appearances. So Block's attempt to disintegrate HOT theory fails.

Still, there's surely *something* solid within the residue of dissatisfaction around Rosenthal's treatment of HOT mistargeting. This precipitate will by no means prove fatal to Rosenthal's theory, but isolating it will suggest a more fruitful higher-order formula for consciousness.

5. Here's a natural way to capture the naive appeal HOT theory has had for many people as an analysis of consciousness. Consider a red London bus-ish percept, but lodged in a *blindsighter's* brain. Though she may glean much information from it, even being able to report on the stimulus colour when prompted, something is clearly missing for this subject; that would be even were she one of Block's mythical superblindsighters. What's missing, of course, is that she is not *conscious* of the red bus-ish percept: there's nothing it's like for her to token it. HOT theory seems to offer a pleasingly substantive account of the extra something needed to get the visual percept into the subject's stream of consciousness, making her conscious of its qualities. The answer is that the red bus-ish percept would enter her conscious mental life just in case it became suitably represented by a HOT.

Now we're aware, from Rosenthal's treatment of mistargeting, that HOTs call the shots regarding the contents of consciousness. What gets into the subject's stream of consciousness is all and only what her HOTs represent as being the case. Imagine our blindsighter has an operation (currently impossible) to undo her blindsight, making her capable (on Rosenthal's scheme) of tokening the HOTs requisite for visual consciousness of items presented to the hitherto blind field. Things start well: her red London bus-ish percept is HOT-targeted, and she becomes (joy!) conscious

[12] Block (2011: 447). Ultimately I deem Block's suspicion justified—the next section explains how.

[13] Block (2011: 443).

[14] The same goes for HOTs: to be conscious, they require still higher-order representers.

as of a red London bus. But then something malfunctions, and, while its corresponding HOT continues in action, the red London bus-ish percept is destroyed. What happens to the subject's conscious state? Does she experience any change? It seems clear that on Rosenthal's theory the subject should still be conscious red London bus-ishly: for it is HOTs and the presentations they make that determine conscious mental life, and this HOT remains in service, dutifully representing the presence of a red London bus-ish visual percept. So far this is only another illustration of Rosenthal's treatment of mistargeting. In our scenario we have just gone from 'veridical' conscious experience (accurate representation of a mental state) to a targetless case, and what he says about this latter species evidently applies.

Here's the problem. It's clear that, *once* it has disappeared, the visual percept makes no contribution to the stream of consciousness: the subject cannot be conscious of the red bus-ish percept, since it no longer exists. If you're really conscious of *x*, as opposed to being conscious merely *as* of *x*, then *x* surely exists. And Rosenthal's theory—on the natural way of explicating its appeal—was supposed to account for our consciousness of mental states,[15] not merely our consciousness *as of* mental states. We wanted to know what it would take to make our blindsighter conscious of her red London bus-ish percept—what it would take for *that very* visual percept to be like something for her; not merely what it would take to give her *every impression* that some such percept was like something for her. As noted, in the present case the disappearance of the visual percept need make no difference to the subject's conscious mental life, since the HOT keeps supplying the relevant appearances to her stream of consciousness. The complaint, then, is not quite that the visual percept is redundant, or is jettisoned in Rosenthal's account of a conscious state—it may, for all we know, be a necessary ground of that token HOT, with its particular content, arising. The problem, rather, is that—as this case shows—the subject simply *never becomes conscious of the first-order state at all*, even in the good case.

An analogy should bring the point out. Imagine looking down on a smooth, level sheet of blue velvet, a couple of square meters in size (supported by someone at each corner). Held under the sheet, not yet touching it, are various moulded figures, also coated in blue velvet. You're unaware of the shapes. For you to form an idea which shapes are there, it's necessary (and sufficient) for us to press some shape up into the sheet, wrapping a portion of the sheet around it, for you to see. Imagine there's a rabbit-shaped object, and it gets raised into the velvet sheet. The sheet is wrapped around the rabbit, and you see a blue velvet rabbit shape, big ears protruding. You now know there's a blue rabbit–shaped figure, and even have accurate ideas of its surface colour and texture. But for all this you don't actually *see* the rabbit. If we remove the rabbit and its impression remains (perhaps the velvet is rather stiff), your visual experience won't alter. But the rabbit is no longer there, so you can't be seeing it. Yet you are seeing just what you saw before we removed the rabbit—namely, a rabbit-shaped portion of blue velvet. Therefore you

[15] At least, where these exist.

did not see the rabbit when the rabbit was present, pressed into the velvet layer, either. The sheet screened the rabbit off from you.[16]

Corresponding remarks apply to an episode of 'sensory consciousness' on Rosenthal's theory. The fabric of a Rosenthalian HOT is simply too 'thick' for the sensory target ever to penetrate to the subject's stream of consciousness. Managing entirely the content of one's stream of consciousness, HOTs effectively block anything else getting in. At best we consciously experience veridical echoes of sensory states; but of the sensory states we are not conscious. This manifests in the possibility that, on their removal, we are left with just the conscious impression they made when extant, and experience no change. Experiencing no change[17] means the same appearances are contributed to the stream of consciousness as were contributed while the relevant sensory state persisted and was HOT-targeted. But since the appearances remain the same with the sensory state removed, the latter was contributing no appearances to subjective mental life. It's not as if the HOT must now '*pick up some slack*' in generating appearances formerly due to the sensory state. Rather, it has been the HOT governing appearances added to the stream of consciousness all along. In other words, we were not ever quite conscious of the sensory state. Rosenthal sometimes approaches acknowledgment of this verdict, averring that a sensory state 'can contribute nothing to phenomenology apart from the way we're conscious of it'.[18] That 'way', of course, is just the HOT which represents said sensory state. Hence all a sensory state contributes to the stream of consciousness is a proxy that represents it; otherwise put, it directly contributes nothing. *It* doesn't appear to consciousness—therefore we are not conscious of it.[19]

Rosenthal's theory, while far from incoherent, lacks on examination the appeal it appeared to possess: it doesn't account for sensory states becoming conscious, since we are never strictly conscious of them. There's nothing *they* are like for us, given Rosenthal's envisaged structure for conscious states. This constitutes not a refutation, but clarification, of Rosenthal's theory. His concern is what supports the stream of consciousness, and nothing said puts his theory in doubt regarding that objective.[20] Still, we're surely within rights to seek a theory that *does* hold the

[16] We could create a misrepresentation case by changing the shape pressed into the velvet while keeping the rabbit impression intact—either suffices for the point, which is perhaps more vivid with a targetless-style case. Note that velvet sheet—the supplier of appearances—and rabbit are *distinct items*, so this is quite disanalogous to seeing an object by seeing its facing side.

[17] N.b. it's not that we are oblivious as to any change in the appearances, it's that there is *in fact* no change to the appearances (regardless how the subject judges).

[18] Rosenthal (2004, p. 32).

[19] It is decidedly *not* implied here that HOTs are conscious (in the sense of our being conscious of them); though they *supply* appearances to the stream of consciousness, these are appearances not of themselves, but as of sensory states.

[20] As Brown (2012) notes 'what matters, on Rosenthal's account, is explaining the subjective appearances…The goal is not to explain how some token state gets transformed into a conscious state, but rather to explain the way one's own mental life appears to one…how it is that you appear to be in some given state at one time and don't appear to be in it at some other time.' Brown calls the Rosenthal-style account 'non-relational', and a theory that aims to explain how we become conscious of sensory states (not merely as of sensory states) 'the relational view'.

appeal we mistakenly glimpsed in Rosenthal's theory. To adopt a term of Kriegel's, we might hope for genuine *intimacy* with our sensory states, in sensory consciousness.

6. A sensible place to turn is to a *self-representational* (SR) theory, to ensure sensory states make an appearance in the stream of consciousness. Here's a simple SR theory: our red London bus-ish percept, as well as representing red London bus-ishly, represents itself. Since this mental state is the thing HO representing, contributing thereby to the subject's conscious mental life, and since it's identical with the item represented, the percept, one intuits that the mental object of mental representation is not left out of consciousness. The bill for this intuition is paid by observing that the sensory target *cannot* now go missing with conscious appearances unperturbed. Since the conscious state uses itself to represent itself, were the target missing the state determining the contents of consciousness would also be absent: in short there would *be* no appearances to the subject, no conscious state.

Sadly, this structure for a conscious state cannot realistically obtain. Wielding representation to analyse consciousness has multiple purposes. One is to capture the phenomenology; another, to render the correct metaphysics of conscious states—as with our noting self-representation seems equipped to get sensory states into the stream of consciousness. Still a third purpose is that we have an eye on *naturalism*—representation is a relation widely deemed a promising candidate for naturalisation. If consciousness can be explained representationally, a naturalistic explanation of consciousness seems within reach. But caution is required: not every variety of representation is guaranteed to be naturalistically acceptable.[21]

Kriegel notes a problem on this score for simple SR theory. Naturalising representation, by our best guess, means explaining it causally. *Very* roughly, representations are caused by their *representata*.[22] But clearly, a token cannot cause itself. Simple self-representation is a reflexive relation, while causation is anti-reflexive, so simple self-representation cannot be analysed causally. That leaves simple SR theory naturalistically dubious, and to be rejected.[23]

7. Self-representation appeared a promising way to get sensory states into consciousness (to make them like something for the subject). *Simple* SR theory failed, but we, with Kriegel, shouldn't be unduly deterred. Kriegel's (2009) book develops a sophisticated SR model, and it's worthwhile to consider its adequacy.

Kriegel splits the conscious state into parts. There's the sensory component, with 'first-order' environmental content—red-London-bus-ish content, say. There is also a HO representational component, targeting the sensory component: it represents

[21] Nor is every non-representational relation of HO cognitive access guaranteed to be naturalistically unacceptable—we exploit this loophole later (§9 onwards).

[22] Though varying widely in their elaboration, this problematic causal condition is the core of most naturalistic theories of representation. A functionalist alternative is rejected in footnote 23.

[23] Kriegel (2009). Gennaro (2006) presses the same difficulty. A functionalist construal of simple self-representation, with one mental state playing a 'first-order' and a 'higher-order' role, escapes the naturalistic concern but clashes with the need to capture phenomenology: functional properties are dispositional, whereas sensory consciousness is occurrent; so they are non-identical properties (see Kriegel 2009, ch. 6.2).

there to be a sensory component which represents red London bus-ishly. That's what the HO component *directly* represents, but it also *indirectly* represents the composite mental state of which it's part. Kriegel explains indirect representation by an example: a painting directly represents (by depicting) the front of a house, and indirectly represents the entire house thereby. Roughly: you can indirectly represent a whole by directly representing a part when that part is well integrated into, and comprises a decent (or significant) portion of, said whole. A map of the UK couldn't be used to indirectly represent the world (despite the pretensions of certain politicians), but a map with Germany, France and the UK highlighted probably *could* be used to indirectly represent Europe. So, the HO representer, in directly representing the sensory component, indirectly represents the entire mental state. That mental state therefore *self*-represents: one of its parts represents its whole. The HO representer being a (logical)[24] part of this overall state, it indirectly represents itself into the bargain. As before, what the representer represents feeds into the stream of consciousness; so the subject enjoys a conscious state with red London bus-ish content, but also peripheral awareness of the HO representer. The representer is the awareness-supplying component, so in (indirectly) representing itself it grants the subject peripheral *awareness of awareness*. Kriegel claims this 'peripheral inner awareness' is a salient, if elusive, aspect of all conscious experience, and his theory garners support from explaining it.

It's key to the distinctiveness of this account that HO representer and sensory component really are united in a single mental state. Kriegel doesn't want simply to *label* their conjunction a mental state.[25] Moreover, the account of indirect representation relies on genuine integration of the representing part with the indirectly represented whole (plus the requirement that the directly represented part comprise a considerable portion of this whole).[26] How, then, are HO representer and sensory component integrated?

Kriegel distinguishes *complexes* from *sums*.[27] A sum ceases to exist only when one of its members ceases to exist. A complex ceases to exist with all its members intact if they leave the complex-making relationship: the relationship between the parts is essential. A friendship is a complex, since it clearly ceases to exist if the components fall out, though they persist. Losing friends isn't usually fatal! For the complex of a conscious state, says Kriegel, comprising HO and sensory

[24] Kriegel defines logical parthood as 'parthood that is neither spatial nor temporal' (2009: 218).

[25] He fears (2009: 221) his theory may differ only terminologically from HOT theory if no 'substantive' connection can be forged between the components.

[26] This is compressed: Indirect representation requires integration of the directly represented with the indirectly represented. So far this only requires a (directly represented) sensory state to be integrated with the entire mental state of which it's part. But for that mental state to *self*-represent (Kriegel's recipe for consciousness) the HO representer must also be integrated with the whole. Then one part of the complex mental state directly represents another part and indirectly represents the whole, so the mental state self-represents. By this operation the representer also indirectly represents itself, and we get peripheral inner awareness. Its being integrated is crucial: it represents itself by indirectly representing the whole only because it is integrated with that whole.

[27] Following Simons (1987, ch. 9).

components, what integrates them is their *phenomenological* unity, underpinned by a cognitive unity whose mechanism he speculates may be as follows:

> If the brain harbored two synchronized [regarding rate and timing of neuronal firings] representations, one in V4 representing redness and another in (say) the dlPFC [dorsolateral pre-frontal cortex] representing increased firing rate in V4, at the personal level we would experience ourselves to have a single representation that folds within it both an awareness of red and an awareness of that awareness. (2009: 246)

Qua complex, it's crucial to the identity and existence of the conscious state that the two components be thus integrated; if they existed without integration, this wouldn't be the conscious state it is, and plausibly wouldn't exist at all—so Kriegel maintains. At this stage it seems HO representer and sensory component do form a complex, and everything looks in order with Kriegel's account.

8. But this appearance is deceptive. Kriegel's SR structure faces no naturalistic challenge, as it has parts to be causally related. But it can no longer, where simple SR could, surpass HOT theory in making the subject conscious of her sensory state—I now argue. This problem accompanies realisation that Kriegel's two components of a conscious state are not a complex, despite first impressions. The gist of the argument below is this: for Kriegel mental indirect self-representation suffices for consciousness (awareness of awareness, etc). But indirect self-representation doesn't require complexhood. So Kriegel's theory of consciousness is compatible with conscious states failing to be complexes. And if they aren't complexes, we will again find sensory states screened off.

We can approach the difficulty via what Kriegel says about the possibility of mistargeting. Since there *are* parts to the Kriegelian conscious state, we can ask what happens when they're at variance.

Kriegel is explicit about cases where the representer misrepresents the sensory state's content: 'inner awareness is a constituting representation: the qualitative character of a conscious experience is constituted by the…properties it is represented to have'.[28] So, given a red London bus-ish percept, the subject experiences red London bus-ishly just in case the HO component represents a red London bus-ish percept (represents itself indirectly, etc.). The subjective awareness-generating HO component generates also the qualities of which the subject is aware. Hence misrepresentation poses no problem: what the HO representer says, goes, effectively: 'It is perfectly coherent to suppose that a mental state may represent itself to be a certain way when in reality it is not that way',[29] Kriegel reasons.

This might worryingly recall Rosenthal's treatment of mistargeting. Things grow more worrisome when we broach the possibility of targetlessness. Kriegel's book presentation enjoys the fortunate (or canny!) feature that as he discusses mistargeting early on, the reader cannot know that the SR account later endorsed envisages conscious states as having parts: during the early discussion one can be

[28] Kriegel (2009: 137).

[29] Kriegel (2009: 136).

forgiven the impression that Kriegel advocates simple SR theory. Accordingly, when he writes 'it is incoherent to suppose that a mental state may represent itself to exist when in reality it does not exist…Thus a targetless self-representation is logically impossible.'[30] the reader experiences the warm glow of a problem neatly solved: Kriegel's theory appears invulnerable to the targetlessness alleged (by him, among others) to threaten HOT theory. But *now* we know there are parts to Kriegel's SR conscious state. And, surely, one part might conceivably exist without the other. What if a SR conscious state arises, comprising red London bus-ish percept and HO component representing there to be such a percept (indirectly representing itself, etc.), but subsequently the sensory state ceases to exist?

We may take our answer from Kriegel's treatment of a case where a conscious state lacks first-order qualities, but represents itself to have such properties (F):

> This may be taken to constitute a serious embarrassment for self-representationalism. However, I do not see that it is…self-representationalism turns in a verdict that is clear and unproblematic…that [the conscious state] is qualitatively F. (2009: 137)

That's because the HO representer's ascription of qualitative properties to the sensory state constitutes such properties figuring in the stream of consciousness. So the first-order qualitative properties are not needed. Transposing this to the case of a kosher conscious state that then loses its sensory component: Kriegel should find this no embarrassment either—while the HO representer persists, representing there to be a sensory component with such-and-such qualities, just those qualities must feature in the subject's conscious mental life.

Kriegel might block this ominous result if his claim that a conscious SR state is complex can be sustained. Complexhood might be seen to *add* a condition to sumhood: *not only* must the parts of the complex exist, *but* they must be in the relevant relationship. So if Kriegel's SR states are complexes, targetlessness can plausibly be ruled out: The defence would be that a targetless HO state isn't conscious, as consciousness requires indirect self-representation, which requires sensory and HO components existing in a complex.

But it can be seen that indirect self-representation does not demand complexhood. A complex fails if the members persist while their relationship dissolves. The following (counterfactual) case shows the possibility of indirect self-representation given parts no longer united in a whole. Imagine an Apple employee graphically representing the company in a presentation by displaying a picture of Steve Jobs (while Jobs lived), ignorant that Jobs had recently defected to Microsoft. It seems this employee succeeds in indirectly representing Apple by representing Jobs, *and* in indirectly representing herself *qua* (fully-integrated) Apple employee. Here indirect self-representation is enabled by the representing part's integration with the whole indirectly represented, but that whole is indirectly represented through direct representation of a *formerly* integrated component. *Jobs-at-Apple* was never a complex, since the company would have survived Jobs' quitting. So indirect self-

[30] (2009: 136).

representation does not require complexhood. Perhaps it requires there *to have existed* a good level of integration between directly represented component and whole (and for the directly represented part to be significant to the whole), but it doesn't demand the ongoing existence of that relationship.[31] It follows that Kriegel cannot claim consciousness implies complexhood, since consciousness-supporting indirect self-representation does not demand complexhood. Nor does it even demand the persistence of the component whose representation grounds indirect representation of the whole: actually we routinely indirectly represent Apple by representing Steve Jobs, although Jobs has died.[32]

Back to our case: the HO representer can still indirectly represent itself by representing the sensory state with which it was formerly integrated. The HO component is well integrated into the mental state thereby represented, as that state now consists just of itself. We have, overall: first, a HO representer that indirectly represents itself, so there should be peripheral awareness of awareness; second, a mental state with a component representing that mental state, i.e. a self-representing—so conscious—mental state; and finally, as before, the HO representer constitutively represents red London bus-ish sensory qualities. All the ingredients seem in place for a Kriegel-style SR conscious state, given the lone HO representer. Kriegel cannot block targetlessness. The (alleged) phenomenological—*felt*—'unity' of *awareness of awareness* with *sensory content* doesn't guarantee an existent first-order state, or complexity in the vehicle of this content.

Kriegel defends his claim that conscious states are complexes: 'When I have a perceptual experience of the blue sky, the perception of blue and the awareness of that perception are unified by some psychologically real relation whose dissolution would entail the destruction of the experience'.[33] But, given Kriegel's *constitutive HO representing* of experienced qualities, we have unmasked that relation as *constitution*: the conscious state is exhausted by the HO (self)representer. The sensory state, even when existent, enters no complex with the HO component.

In our targetless case it's the constitutive representing of red London bus-ish qualities (as ascribed to the sensory component) that sees them enter the subject's stream of consciousness. Now the sensory component is gone, such representation persists. The qualities that now figure in consciousness are the same, and their

[31] Kriegel's painting of a house doesn't self-represent, but plausibly all that's needed is for the painting to be *inside* the house it depicts. But painting and house are not a complex—no entity is destroyed if the painting is removed. Moreover, if the house burns down, but the painting survives, the painting can still indirectly self-represent by representing its old home.

[32] A trickier case is where the HO representer represents a non-existent sensory component (one *never* integrated with it). Yet if the HO component is 'none the wiser' about the non-existence of its intentional object, and this represented object is putatively integrated with it in the way an existent sensory component would be, plausibly it could still indirectly self-represent. Maybe this case is far-fetched, but it might illustrate the possibility: if someone believed Atlantis (assuming Atlantis never existed) was a large component of the European landmass, they could plausibly indirectly represent Europe by representing Atlantis. If that's right, Kriegel's 'integration' requirement on indirect representing has a somewhat intentional flavour.

[33] Kriegel (2009: 222).

source is the same; namely, the aforementioned constitutive HO representation.[34] Therefore, by parallel with the reasoning applied to Rosenthal's theory, the subject is not conscious of the sensory state's qualities even in the good case. Kriegel's HO representers screen the sensory state off from consciousness just like their Rosenthalian predecessors.

Kriegel implicitly acknowledges this result. Even where a sensory component is present and correct, its '[first-order qualitative] properties are not part of the experience's phenomenal character, indeed are not phenomenologically manifest in any way',[35] he says. Kriegel's account, by following Rosenthal's theory in its division of labour between percepts and their HO representations as regards supplying the appearances characteristic of consciousness, forgoes complexity, and fails to deliver on its promise to make us conscious of our sensory states.[36]

Recall our blindsighter who has blindsight-reversing surgery. Considering the before-and-after contrast in this subject's mental life, who, post-operation, is conscious of the red London bus-ish qualities that lay unexperienced in her visual cortex, we hoped for a theory that would explain what it took to get these qualities—*just these* qualities—into her stream of consciousness. Rosenthal's HOT theory and Kriegel's SR theory cannot satisfy this reasonable hope. These accounts construe conscious states in *non-relational terms*, with HO representers doing all the 'phenomenal labour'. Hence the variety of HO theories of consciousness is only superficial. But it's not asking too much for a HO theory to satisfy our hope, and it could well be expected to construe consciousness-supporting states as complexes. Such a theory follows.

9. Kriegel introduces the notion of a 'display sentence', via the example of a bridge with 'under construction' painted on it.[37] Crossing, you think 'This bridge is under construction'. Kriegel suggests what you actually have before you is the complete *sentence* 'This bridge is under construction', with the subject term supplied by *the bridge*. The bridge is *present* in the sentence, allowing the sentence overall to say something about it. Generally, a display sentence features a constituent whose semantic role is to contribute itself. Searle notes that usually our conversational topic is not within reach, so we use a symbol to stand for—represent—it. When the topic *is* in our vicinity, however, the possibility arises of embedding it within our discourse. In such cases the term pertaining to the item in question does not *re*present it, but, being just the item itself, simply presents it.[38]

[34] One might fear we now have a simple SR state, with accompanying worries about naturalisation. But as Kriegel argues, indirect representation is plausibly non-causal, so there's no problem with it being reflexive. Kriegel has perhaps rescued simple SR theory! The notion of a non-causal, but naturalistically acceptable, grounding for mental intentionality will matter later.

[35] (2009: 110).

[36] Weisberg (2008) notes HOT theory and SR theory agree in having HO representation constitute experienced qualities. He infers SR theory has no advantage regarding intimacy with first-order qualities. But, as he endorses HOT theory, he cannot see this—as I do—as reason to look past both views, for one that would make us genuinely conscious of sensory states.

[37] Kriegel (2009) derives the term 'display sentence' from Zemach (1985), who attributes the idea to Searle (1969).

[38] Searle (1969). Kriegel prefers a representational reading of display, but I side with Searle, as explained below.

Kriegel doesn't note, but it is relevant to us, that the bridge only acquires this presentational role when 'embedded' in the sentence. The bridge doesn't, standing alone, both exist as itself and present itself. That's a function of the context supplied by the semantic apparatus of the sentence wherein it features: within the sentence, it is *used* to present itself.

Kriegel's infectiously over-excited instinct is that display sentences are the key to understanding consciousness, since 'there is something special and unusual going on…which might help us *feel* a "quantum leap" to something that might indeed be sufficient for [consciousness]'[39] Kriegel's proposal is to model conscious states as display sentences of a sort. But whereas he considers that the best way to do this is via his SR theory, I offer a different way of constructing 'mental display sentences' that doesn't screen off first-order sensory states. This way has certain other virtues: notably, it deals neatly with mistargeting, and does much to satisfy our sense that we are in some direct manner 'acquainted' with sensory qualities in consciousness, but without forfeiting naturalism.

The guiding idea is that sensory states are constituents of the relevant consciousness-supporting states, and serve to display themselves within those constructions.[40] This recalls a popular theory of the structure of phenomenal *concepts*; that they 'quote' the experiences they refer to.[41] In Balog's words:

> The idea of an item partly constituting a representation that refers to that item is reminiscent of how linguistic quotation works. The referent of "__" is exemplified by whatever fills in the blank. In a quotation expression, a token of the referent is literally a constituent of the expression that refers to a type which it exemplifies and that expression has its reference (at least partly) in virtue of being so constituted. So, for example, ""dog"" refers to the word spelled d-o-g, a token of which is enclosed between the quotation marks. Although in English we normally quote only expressions of English we can also quote foreign language representations and non-linguistic representations. We can even imagine, perhaps just as a joke, placing something which is not a representation, e.g., a cat, between quotes and thus produce a representation that everyone can understand refers to the type cat. My proposal is that there is a concept forming mechanism that operates on an experience and turns it into a phenomenal concept that refers to either the token experience, or to a type of phenomenal experience that the token exemplifies.[42]

I suggest the right higher-order analysis of consciousness sees a HO state 'quote' a sensory state, forming a larger composite structure wherein the sensory state is displayed. Its being embedded within the HO state and thereby displayed is what

[39] Kriegel (2009: 164).

[40] Zemach (1985: 196) says a displayed item '*is* its [own] sense' in the display sentence. When the item is a sensory state, this idea helps tip us further towards Kriegel's feeling of 'specialness', which may just give the display sentence structure the wherewithal to capture consciousness.

[41] Papineau (2002) proposed the quotational account of phenomenal concepts. Papineau (2007) revised it significantly.

[42] Balog (2012: 33).

constitutes the subject's awareness of the sensory state. In contrast to the Papineau/Balog model for phenomenal concepts, the quoted elements are not yet experiences—for we're explaining what turns sensory states *into* experiences. The suggestion is that it is *mental quotation*, so this is 'quotational higher-order thought'—QHOT—theory. When a sensory state is quoted in the requisite way, the result is a mental display sentence, which provides a conscious state, on the analysis.[43] Another difference is that the subject matter of QHOT theory are *tokens*: the question is what makes one conscious of a token sensory state. Balog's theory is primarily concerned with *types* of (experienced) sensory states: typically we think about experience types ('Ow—pain again'), but we only experience sensory tokens. For Balog token experiences are recruited to represent experience types in thoughts about them. In QHOT theory sensory state tokens are recruited for display in HO quotational structures that supply consciousness of said token sensory content.[44] The final difference is that QHOT theory's quotational structures are *non-representational*: there is no need to represent a token sensory state which is actually present; a type of course cannot be wholly present within a thought, so must be referred to by use of an exemplifying token. More on this feature below (§13).

The quotational higher-order thoughts that supply consciousness are envisaged as very *thin*, best modelled as demonstrating 'frames', with a 'slot' for the sensory state. I propose the required sort of HOT has the frame-like structure 'This state is present: "———"', with the gap between the "———" for the embedding of a sensory state. Let's insert our red London bus-ish percept to yield a complete instance of the state structure:

'This state is present: "red London bus-ish visual quality"'

The thesis is that a subject is conscious of this token red London bus-ish percept just in case she enters such a state.

10. A consciousness-supporting state on QHOT theory is complex. Conscious states comprise sensory content plus subjective awareness of that content. QHOTs, being mere frames, determine no first-order sensory content themselves, all they can do is display the sensory states they embed; a conscious state's sensory content, then, is wholly supplied by the quoted sensory state. In contrast with Kriegel and Rosenthal's non-relational theories, a lone HO state cannot support an experience, since it proffers no sensory content—it simply lacks the resources (more on this below). And without the quotational frame, a sensory state is not displayed, so by hypothesis fails to be the object of awareness. Kriegel's test of a complex is that its components can persist without the complex existing, if they leave the relevant relationship. There is nothing to prevent a given QHOT and sensory state existing separately. But a conscious state—comprising awareness and sensory content—only arises when the sensory state is embedded by the QHOT. Each component supplies

[43] What is proposed is thus similar to Balog's 'joke': placing a non-representation within the QHOT's quotational structure (except sensory states *are also* representations: they represent environmental features).

[44] Zemach sees a use for display sentences to account for thoughts about experiences (e.g. pains, 1985: 196), but it doesn't occur to him that consciousness might be implemented by their means.

a necessary element the other lacks. Therefore the embedding relationship between QHOT and sensory state is essential to a conscious state, and they form a complex.

A distinction is needed to fully understand the previous paragraph. Kriegel and Gennaro both claim *the conscious state* is complex. But what is conscious—what is like something for the subject—is a *content*, and the content of consciousness does not *feel* complex, in the relevant sense. A conscious state does not *feel like* the relating of two components such that if the relation between them fails the overall state is forfeit. Even if we discern elements in phenomenology, even if Kriegel is right that one of these elements is awareness of awareness, phenomenological unity and holism are paramount (something Kriegel himself stresses). What is complex, if something must be complex, is the *consciousness-supporting* state: the vehicle(s) of the conscious content. So the claim about QHOT theory is that the quotational relation between QHOT and sensory state makes them into a complex state, and this complexity is essential to that vehicle supporting a conscious content. Problematically, 'the conscious state' is used in the literature sometimes for vehicles and sometimes for contents. With this distinction made, we should recall that Kriegel isn't entitled to claim his consciousness-supporting states are complex, since lone HO self-representers suffice for whatever 'complexity' is felt in phenomenology.

This is also a good point to address Gennaro's theory. In earlier work he seems to offer a constitutive HO theory, which would include the unwelcome screening-off artefact of Rosenthal and Kriegel.[45] Recently, he has sought to entwine first-order (FO) states more closely with consciousness, such that without one there is no conscious state (CMS). He reasons:

> If we have a [HOT] but no [FO] at all (or vice versa), then what would be the *entire* conscious state does not exist and thus cannot be conscious. A CMS will exist only when its two parts exist and the proper relation holds between them.[46]

But this—effectively the claim that a conscious state is a complex of HOT with FO state (the relationship concerns how the HOT conceptualises the FO state)—relies on the ambiguity just highlighted in 'a conscious state'. Gennaro really means that the *vehicle* that suffices for there to be something it's like for the subject is complex. A conscious state does not feel complex, in the relevant sense, so this cannot be a claim about content. Anyway Gennaro's HOTs are unconscious, so there can't be a *phenomenology* of their being necessarily integrated with FO states in consciousness.[47] Another indication is his claim that the unconscious HOT is 'part of' the conscious state. This clearly cannot be meant in a content-related sense—since an unconscious content is not a conscious content nor part of one. Hence Gennaro is talking at the vehicular level. So we should ask what supports the claim that the

[45] Constitutive HO representation is strongly suggested when he says 'it is the HOTs which bring the intrinsic qualitative properties into the conscious state…When a pain or perception becomes conscious by virtue of becoming the target of an appropriate HOT, it [only] then becomes a qualitative state' (2004: 61).

[46] Gennaro (2012: 61).

[47] As against his argument (2012: 57) that consciousness is 'intrinsic' to a conscious state.

vehicle for an instance of what-it-is-likeness must be complex. Why wouldn't a lone HOT representing an absent FO state suffice for an instance of what-it-is-likeness indistinguishable from a case where the represented FO state exists?

Gennaro believes self-reference is involved in consciousness—his HOTs, by representing FO states with which they are suitably related, allow the complex state to self-represent. The main difference with Kriegel here is that Gennaro's HOTs don't thereby self-represent, so there isn't awareness of awareness—his HOTs remain unconscious (unless introspected). So, in answer to the question above, we can see why Gennaro would say: 'A CMS cannot represent itself (or part of itself) if it doesn't exist in the first place'.[48] But it's clear he is again using 'conscious state' in the vehicular sense. If we *suppose* a conscious state is complex vehicle-wise, then it cannot self-represent with one part missing as it would not, *qua* complex, exist. But this doesn't answer the content-level question. Would a lone HOT provide what-it-is-likeness? We know Gennaro thinks self-representation matters, but, as argued in connection with Kriegel, naturalistically acceptable self-representation is possible for a single state. So self-representationalism won't support Gennaro's claim that a conscious mental state is complex, i.e. that the FO state must exist. Gennaro does not provide good grounds for his thesis that HOT and FO state must both exist for consciousness.

Elsewhere in his theory we find cause to think FO states are surplus to the requirements of subjective mental appearances. His model of introspection sees a third-order HOT target a second-order HOT. But Gennaro allows that this sometimes occurs without a FO state, as in 'dental fear', where patient expectations generate something like an experience of pain. Gennaro holds that, absent a FO pain state, patients 'still subjectively experience those states in an indistinguishable way',[49] i.e. indistinguishably from a genuine pain. This treatment is puzzling. Here a subjective mental appearance as of pain is generated without any FO pain state. One may then reasonably ask why, if two HOTs without a FO state suffice for an experience as of pain, introspectively, *one* HOT with similar content won't suffice for a non-introspective pain-ish experience. Gennaro may point to the presence of an existent conscious state in introspection: the introspected HOT. But in normal experience a HOT doesn't need to be conscious for its content—e.g. that there's pain—to mould subjective mental appearances, so it's unclear why this difference would make the difference. Again, the introspected HOT is *mistaken* for a pain, due to the representational error of the introspective HOT. But this HOT is unconscious, so we still find that *unconscious HOTs can single-handedly fix mental appearances*. Given this approach to introspection, it's hard to see what grounds Gennaro could offer for saying a lone pain-representing HOT wouldn't suffice for a pain-ish experience.

From other things he says there would be considerable pressure on him to openly put HOTs in sole command of subjective mental appearances. He holds that experienced content is wholly fixed by concepts in the HOT (hence the experiences

[48] Gennaro (2012: 63).

[49] Gennaro (2012: 69). See pp. 96–7 for a case involving emotions.

of trained wine-tasters) and still maintains that without a HOT sensory states lack qualitative character.[50] Overall Gennaro's theory is unstable, and its strongest elements will likely push him into a view with constitutive HO representing, screening-off FO states, like Rosenthal and Kriegel.

The trick to turn with a HO theory of consciousness is that matters content-wise, what's experienced, must be explained by matters vehicular, not the other way around. If one finds oneself adding conditions to the vehicular relationship just to preserve some content feature—e.g. to involve FO sensory contents in consciousness—then one is into the realm of the *ad hoc* move. There's nothing in the nature of his vehicles that requires content match between Gennaro's FO and HO states—this emendation exists solely to cater for a content-feature Gennaro desires. The exemplar is simple SR theory: here it couldn't be clearer how content depends on vehicle; no HO or FO vehicle, no consciousness, on the assumption that HO representation enables consciousness, since loss of either vehicle means no vehicle at all, hence no HO representation. Gennaro and Kriegel both wish to retain the services of FO states, but their vehicular setup does not allow this in a principled way. Rosenthal has the virtue of not worrying about the relevance of FO states. If we must reject simple SR theory, we are left with one theory where the relevance of FO states follows from the vehicular setup: QHOT theory, where complexity derives from the nature of the vehicles.

11. Through those features that deliver complexhood, QHOT theory offers a satisfying treatment of mistargeting, without 'screening off' sensory contents like previous accounts. It's clear, first, that misrepresentation cannot occur. A QHOT determines no sensory content: all the sensory content of a conscious state is supplied by the quoted sensory state. If I want quotationally to represent what Florence said yesterday in the heat of argument I can say 'She said this: "Get out and never come back"'. A little closer to our model, I can play a tape-recording of what she said (was I sufficiently self-possessed to make one), saying 'She said this: *click*', i.e. playing the tape-recording at the relevant point. Closer still, I could summon her to repeat what she said, saying 'She said this: "———"', and letting her fire. I'm not yet employing her *very utterance*, though, so we can imagine one further—outlandish—case. Had we a time machine, we could return to the instant Florence was about to shout at me, and I could say (looking on at my wretched past self): 'She said this " ———"', indicating the utterance. Now, with this way of quoting what Florence said—note it is no longer *represented*, but *exhibited*—it's impossible for me to stray from what she said, since her very utterance is used by me to present itself. Likewise, it's impossible for a QHOT to misconstrue the sensory state of which it supplies awareness. For what supplies the sensory content of the quotational conscious state is just *that* very sensory state itself, with *its* content. This result is achieved by the removal the dual layer of sensory contents featured in SR and HOT theories.[51]

[50] Gennaro (2012: 65).

[51] Block (2011) notes that what ails HOT theory is deploying two layers of qualitative content.

What of targetlessness? A lone QHOT won't suffice for a conscious state. Here's why. QHOTs are hypothesised to supply subjective awareness. However, if a QHOT has no sensory target to embed, it could at most arouse subjective awareness of (a state of) *nothing*, since it altogether lacks first-order content. But subjective awareness of a state of nothing at all just isn't subjective awareness, since subjective awareness must be (intentionally) *of something*. An experience literally of nothing is simply no experience. Therefore an empty QHOT won't produce subjective awareness—or a conscious state.[52] QHOTs can only make the subject aware of something if there is something (sensory content) to be aware of. 'This state is present: "'*blank*"' fails to be a thought.[53] Similarly, a linguistic quotational frame without any entrapped token *doesn't* quote, and fails to be a sentence. QHOT theory claims that consciousness is mental quotation.

The theory thus deals neatly with the mistargeting cases that, on a plausible reconstruction, largely motivate Kriegel and Rosenthal's move to 'insulate' first-order states from consciousness, putting HO representers in charge of the subjective mental appearances characteristic of phenomenality. QHOT theory manages this without screening off sensory states. Since such states get into the very fabric of the structures that constitute awareness of them, since they are displayed in those structures simply by being embedded, when we are conscious of a sensory content it's the very first-order sensory state with this content of which we are aware.[54, 55]

12. The last sentence evokes *acquaintance*, a relation of intimacy to sensory qualities many have felt we enjoy, but which seems naturalistically unpalatable,

[52] Some meditators claim to achieve a conscious state of 'nothingness'. But, without doubting the veracity, we can challenge this description of their reports. The meditative state likely features an exceptionally general, diffuse, sense of oneself, or the universe at large, but were it really an experience *of nothing* we should declare it no experience.

[53] Hallucinations purport to make us aware of nonexistent external things. But there still exists a *sensory content*, as of an (in fact nonexistent) external item. In our case, we are imagining there is no such sensory content, either. In that case a QHOT cannot make us aware *of anything*, therefore it cannot make us aware, full-stop, since awareness is always (intentionally) of something.

[54] By dealing with mistargeting while keeping first-order contents as constitutive of consciousness, QHOT theory sidesteps Brown's (2012) argument from mistargeting against the relational view, and for the non-relational view.

[55] Van Gulick's higher-order global space (HOGS) theory (2000, 2004) envisages a sensory state as 'embedded' within a HO state that represents it (Van Gulick holds sensory states are 'recruited' by global brain states, which implicitly represent them). He can thus plausibly offer a similar treatment of mistargeting (e.g., in a targetless case he might claim there is a 'hole' in the global state where sensory quality should be, meaning no sensory consciousness). However he doesn't avail himself of a quotational model, and retains representation. He also holds that HO representers 'transform' the sensory state's content, that making it conscious 'require[s] some change in the state itself rather than just making it the object of a higher-order thought or perception' (2000: §3). I deny quotational embedding alters the sensory state. Once the HO component gets involved in determining experienced sensory content, we risk losing contact with the first-order state again in favour of constitutive HO representing. In this case Van Gulick's theory would lapse into the undesirable situation of the other theories. What these accounts share is *two levels* of qualitative content (cf. Block 2011; Brown 2012), one of which inevitably conflicts with the other, and representation, which creates distance from one's subject-matter. QHOT theory has one level of sensory content, and no representation, so avoids this difficulty. If Van Gulick relinquished the 'two content' model, and constitutive HO representing, QHOT theory might be viewed as a way of developing the HOGS model.

because unanalysable. Kriegel favours his SR account of consciousness over an acquaintance-based account precisely because representation appears naturalisable. Still, he admits the appeal of the idea that we're intimate with our conscious sensory states, so we cannot be mistaken *in consciousness* about their qualities, and in the sense that 'there is not a gap…between the awareness and what one is aware of'.[56] His way of achieving this last result is, we've seen, to make HO representing constitutive of experienced qualities, screening off first-order sensory states from consciousness. The intimacy Kriegel supplies, then, is not the sort we hoped for: we are intimate with the wrong thing, on his theory.

There's little worrying for the naturalist in the idea of quotational display; else physicalism would be under threat from Shakespeare seminars, and Balog—a hardened physicalist—wouldn't appeal to quotation to model phenomenal concepts. But, intuitively, mental quotation *would* make us intimate with first-order qualities: there'd be no gap between awareness and what we are aware of, since sensory qualities would *themselves* be displayed within the quotational conscious state. So QHOT theory captures what's desirable in the idea that we are acquainted with our qualia, yet avoids the screening-off artefact of other accounts, *and* remains naturalistically respectable. Quotational display is that tantalising tad more intimate than representation, without being naturalistically worrisome like acquaintance. To be clear: quotational display *is not acquaintance*, in the sense of an irreducible epistemic relation to conscious contents. Quotational display is *analysable* as the embedding of a sensory state in a QHOT—the key semantic mechanism is constitution. More on this mechanism in the next section.

13. I now address some objections, aiming to clarify and defend QHOT theory.

i. *So a sensory state appears as such within the QHOT, but why couldn't its content be 'distorted', yielding something like misrepresentation?*

Since we're talking about quotational sampling, any inaccuracy would have to derive from something *outside* the quoted item. But then it could not, it seems, tamper with the latter's intrinsic content. Perhaps I could frame Florence's utterance thus 'She did *not* say this: "——"' in the time-travel scenario. Clearly, though, since the quoted item is present *by hypothesis*, this frame does nothing to distort it. But what if a nefarious QHOT arose with content like 'The qualitative *opposite* of this state is here: "—"', directed, say, upon a red percept? Would one see green? Either this QHOT supplies a conscious state, or not. No difficulty arises in the latter case. In the former, recall that a QHOT contributes no *first-order* content (sheer quotation, a higher-order operation, cannot amend the quoted) so if the subject is made conscious at all it can only be of the embedded, red, sensory content. So it seems first-order sensory content cannot be misportrayed in consciousness—even for nefarious QHOTs, first-order inaccuracy is impossible. Perhaps the most that could happen is that the subject acquires an erroneous (and, if conscious, discomfiting) *belief* that what is an occurrent conscious state is distorted in consciousness. Analogously, a sign saying 'The geographical opposite of this is

[56] Kriegel (2009: 109).

here', pointing to a large chasm, can do nothing but make the observer aware of the *chasm*; it cannot make her see a mountain.

ii. *Still, quotation doesn't preclude* targetlessness*: I can point at Beyoncé and report (lying) 'She said this "I'm so into you"'. So it isn't the quotational element of QHOT theory that deals with targetlessness, but a stipulation that the quoted item be present. But this is as* ad hoc *(or not) as Gennaro's matching requirement, criticised earlier.*

Normally when quoting we must employ representations of the subject-matter, as we are unable to display it (even on Balog's theory phenomenal concepts typically use a token of an experience type to stand for *another* such token, or for the type). This opens the door, of course, for distortion (misquoting), even targetlessness (as in lying), since we are at representational distance from the target.[57] But the primary notion in QHOT theory is the *display sentence*, and the variety of quotation involved is peculiar in being display-based: the very subject-matter is present in the quotational construction. It's this variety of quotation which blocks targetlessness, in a principled way. An absent target cannot be displayed, and without display there is—by hypothesis—no consciousness.

iii. *If QHOT theory precludes error about our conscious states it's an implausible theory, since we evidently make such errors.*

QHOT theory implies we cannot be mistaken *in consciousness* about which sensory state is present—it cannot be we are conscious as of a red quale when what's instantiated (in that precise 'location') is a blue quale. The QHOT structure indeed leaves no room for this. But it has long been felt that experience is invulnerable to such error: that no appearance/reality gap obtains for consciousness. QHOT theory promises to vindicate this traditionally troublesome intuition in a naturalistically acceptable form. Notably, HOT and SR theories also achieve this result, in their way—they agree appearance and reality coincide for consciousness—albeit at the cost of constitutive, thus insulating, HO representing. So denying an appearance/reality distinction for consciousness, if it comprises an objection, does not single QHOT theory out. QHOT theory has the added benefit of guaranteeing first-order accuracy. But why would that provide an objection to the theory? For other theories the problem would be that the relationship between HO and FO state is *representational*, and representation—if naturalistic—entails possible misrepresentation. But QHOT theory is non-representational, so this inference does not apply. As we will see below, there is nothing non-naturalistic about QHOT theory's posited mode of higher-order cognitive access, yet this mode ensures FO accuracy.

However, QHOT theory doesn't remove the possibility of error about conscious states—for as soon as we *represent* these, in thought, we presumably move away from a display structure. And now misclassification and misrepresentation become distinct possibilities—and actualities. Though feeling the cold of the dentist's needle, one may well, in expectation of pain, momentarily misclassify the sensation

[57] Cf. Levine (2001: 108).

and react by the classification rather than what's felt. Likewise introspection, if it employs representation, suffers such defects. QHOT theory has no trouble permitting these phenomena. In fact, it *explains* why there is no appearance/reality gap for consciousness, but a healthy one for thought about consciousness: the former, but not the latter, uses a display-based structure.[58]

iv. *What* is *mental quoting, anyway? We understand how linguistic quotation works, but that relies on shared conventions, which can't be operating in the context of making a mental state a conscious state.*

In fact there is surprisingly little agreement about how linguistic quotation operates, though we seem to recognise and understand it when we meet it.[59] So the objection cannot be that linguistic quoting is well understood but mental quotation wholly obscure. Departing from certain elements discussed in connection with linguistic quotation, we can sketch a mental model and its required features. First is a question of reference. In linguistic quotation, a token word is entrapped usually to pronounce about its type—as in '"Socrates" has eight letters'. Where we quote one thing to say something about another—a token is non-identical with its type—then of course *refer-ence* must occur, the token actually present is utilised to speak of another, related, item. With QHOTs, however, a token sensory state is not entrapped to make comment on *some other* entity, not even its qualitative type. Instead, the QHOT's semantic duties begin and end with that token, which is actually present, embedded in the QHOT (I discuss the embedding below). Accordingly, it has been maintained that the entrapped token plays no referential role in linguistic token quotation. Rather than referring to anything, it is simply present in, and *presented by*, the quotational frame. This seems the sort of thing we should say about the sensory states in QHOTs. Since the sensory token is not used to refer to anything (even itself) there is no need of quasi-linguistic conventions to settle reference. The sensory token is just presented. Once we move outside the head we are typically dealing with one thing standing for another, *signs.* Signs are arbitrary, since they (usually) do not contain what is spoken about, so conventions become needed. When a token is being presented no convention is needed to relate it to itself, referentially or otherwise. It is simply *there* and displayed.

[58] A wrinkle: On Balog's model a phenomenal concept *can* be used to think about an occurrent token experience. If that token is, as on QHOT theory, a displayed token sensory state it might seem the quotational ascent that takes us to the phenomenal concept leaves no room for distortion: there would then be, perhaps implausibly, an error-free variety of thought about experiences. Some won't find this implausible—while a theory must allow error in thought about experiences, that doesn't mean it can't compass also an error-free kind. However, this consequence may be avoidable: Balog suggests (see earlier quote) that reference to a token experience goes first via the type it exemplifies—one uses the token to refer to the type, and in falling under this type one thinks again of the token. This long, referential, chain affords ample opportunity for error, e.g. if the token changes while thought is 'on its way back' to it. Alternatively, one may reject the quotational phenomenal concept model of introspection, as I do.

[59] See e.g. Davidson (1979), Washington (1992), Searle (1969), for recent theories and Capellen and Lepore (2012) for a wider survey. The model we are closest to is undoubtedly Searle's.

There is another potential source of conventionality: linguistic quotation works through *quotation marks*, and only by convention do they have their function regarding the entrapped token. It's worth noting here too, however, that some schools of thought deny quotation marks *refer* to the entrapped token. It has been claimed that they simply *demonstrate* or *present* it (n.b.: this isn't the same as saying they are a *demonstrative device*, a referring expression[60]). [61] This, too, is the sort of thing we should say about QHOTs. Again, it may be alleged that even if they don't strictly refer to it, still it is only by convention that quotation marks have their function of demonstrating or displaying the entrapped. In response we may note, a point common in the quotation literature, that quotation doesn't need quotation marks—italics do equally well, and there are many other devices, actual and possible. In speech emphasis indicates a quoted word. This suggests it's the way the entrapped token is *used* that determines whether it is quoted. All our various marks do is make the reader aware she is entering a quotational context; but *that* the context is quotational is not actually a matter of convention. It may be a matter instead of author intentions: but that shows a quotational context is not an essentially conventional entity, but one *determined by the user* of the entrapped token. Correspondingly, in the mental case we would hypothesise the existence of a kind of sub-personal mental state that can use sensory states in this sort of way—for display—as an intrinsic matter, i.e. as a feature of the cognitive system. Its functional role therein must be to embed token sensory states for display to the wider network. I lack a detailed notion of the requisite functional role, but with the need for convention removed there seems no obvious bar to positing such a sub-personal state or function.

The suggestion may be that this is *proto-quotation*, the primordial form underpinning, or logically preceding, the linguistic sorts we engage in. Analogously, linguistic reference is often considered derivative upon the capacity for mental reference. Nobody has uncovered the mechanics of mental reference. But few are in any doubt that there is such a thing. I am proposing something similar for the primitive mental quotation of QHOTs.[62] It's notable that discussions of linguistic quotation sometimes stray into speculating that non-linguistic items can be linguistically quoted—as with Searle's California Jay birdcall.[63] All we need imagine, additionally, is that a non-linguistic item might be non-linguistically quoted—or at least, not in a public language. That would give us the sort of existent which is a QHOT.

Things, sentient agents excepted, do not present themselves—something else is required to present them. That is the need for a second mental state to display a sensory state, which mental display of a mental state is that in which state consciousness consists. The idea is clear enough that the sensory content featured in

[60] See Reimer (1996) §3 for the distinction. It is often held that a demonstrative requires a demonstration to ground it.

[61] E.g. Searle (1969), Reimer (1996).

[62] This would fit a Searlian 'mind first' approach to language.

[63] Searle (1969: 76).

the conscious mental state is supplied by the first-order state, which latter cannot, alone, *present* this content to the cognitive economy. But what corresponds mentally, or in the brain, to the *embedding* of a token in linguistic quotation? One natural suggestion is that the sensory state is *literally* embedded in the QHOT, becoming a proper part of the composite completed QHOT. By contrast, in simple SR theory the sensory state is an improper part, and in Kriegel and Rosenthal's theories no part, of the awareness-supporting state. We may add the condition that (typically) a sensory state causes a QHOT to embed it. Once embedded, however, the primary semantic mechanism is plain part/whole *constitution*: the main precedent for this comes from theories of linguistic quotation. In Searle's theory a token sound is quoted by becoming a literal part of the quoting sentence.[64] Constitution is also a semantic mechanism in the Balog/Papineau theory of phenomenal concepts. Finally, we may recall that Kriegel's indirect self-representation—his key to consciousness—relies on constitution: the directly represented must *partially compose* the indirectly represented.[65] QHOTs make one aware of sensory states, and the 'of' is indeed that of intentionality. But it's false that this 'of' demands to be understood representationally.[66] Something can be an intentional object by being *contained*—literally a content. The content of a completed QHOT includes its embedded sensory state.

A useful analogy is the *picture frame*. A frame's function is to display a particular picture, and the relationship between frame and picture is that the latter is embedded in the former. There is no question that, by convention or in any other way, a given frame could *really* pertain to—be displaying—some picture in the gallery other than the one it physically encloses. Containment, then, has what it takes to be a primitive intentional relation. This sort of relation also has explanatory promise in neuroscience. Following Feinberg,[67] one might envisage disparate brain areas implicated in conscious experience as forming 'nested hierarchies', where 'lower order elements', instead of being dispensable products of earlier processing, actually help to 'make up' the final conscious percept. In a nested hierarchy, Feinberg says, 'the elements comprising the lower levels of the hierarchy are physically combined or nested within higher levels to create increasingly complex wholes'. This description recalls QHOT theory.

A near relation of the Papineau/Balog theory is worth mentioning here. Chalmers 2003 (see also Gertler 2001) develops a model of phenomenal concepts that leans heavily on constitution (what Gertler calls 'embedding')[68] to ground the relevant mental intentionality. Chalmers claims 'direct' phenomenal concepts 'take up' a phenomenal quality into themselves in a way that issues in a peculiarly intimate

[64] Searle (1983: 184). See also Washington (1992).

[65] Cf. also Frege's (1892/1962) view that quotational semantics are fixed simply by the *identity* of the embedded token with the expression talked about.

[66] As against premise 3 of Lycan's 'simple argument' for HO representationalism (2001). Were Lycan correct no direct realist could hold that a perceptual state was about an environmental object just by being partly composed of that object. QHOT theory is thus akin to 'inner direct realism'.

[67] Feinberg (2000) (see p. 79 for the quoted phrases below). I found the view in Gennaro (2012).

[68] See her (2001, p. 308) for an account of embedding.

form of intentional relation.[69] What's especially interesting for our purposes is that Chalmers turns to constitution as a semantic mechanism because of the inadequacy he finds in conventional naturalist—i.e. causation-based representational—accounts of content. He argues that whichever conventional naturalist relation one employs fails to pin down the content of a phenomenal concept (a 'zombie' phenomenal concept remains conceivable). This sort of argument is now a staple of the 'phenomenal intentionality' approach to mental content.[70] Chalmers claims the relation between a phenomenal concept and the phenomenal quality it speaks of is 'tighter' than any causal relation, hence he invokes constitution to turn the trick. We needn't side with Chalmers in adducing the wholesale failure of naturalist approaches to intentionality. However, given their failure *thus far* to reductively analyse aboutness, it would be ironic, in light of Chalmers' resort to constitution, if someone suggested that, regarding QHOT theory, constitution was an inadequate ground for mental intentionality, and urged us instead towards conventional naturalist causation-based mechanisms! Constitution, then, is the 'psychologically real relation'[71] that binds a QHOT and its first-order sensory target. They are sealed as a unit, a real composite existent, by the new powers of the complex: this now displays the embedded sensory content to wider cognitive systems, for use.

Chalmers and Gertler are, admittedly, anti-naturalists. But this isn't due to the centrality of constitution in their model of phenomenal concepts. Chalmers must argue *from* his model towards an anti-naturalist conclusion—in the limited form of observing that his theory would prevent the Papineau-style phenomenal concept-based response to anti-physicalist arguments. It does not follow that, because causation/representation don't suffice to secure the mental intentionality required for consciousness, naturalism is in trouble, simply because constitution carries no threat to naturalism. How could it? What is more mundane? Mother nature uses whatever materials she has to hand—and the constitution relation is an unglamorous, but economic, means of tightly integrating two existents: in our case a quotational mental state and the sensory content it displays. Chalmers, significantly, sharply distinguishes mental-access-via-constitution from the troublesome acquaintance relation.

Obviously more remains to be said, but it needn't be said to make the *model* viable. At this level of theorising we aren't required to produce *detailed* functional analyses, let alone specific proposals for their physical implementation.[72] Science-

[69] He also says the direct phenomenal concept's content 'is partly *constituted* by an underlying phenomenal quality' (p. 235) and talks of a 'slot' in the concept for a quality (p. 243).

[70] See e.g. Loar (1995), Horgan and Graham (2009) for this critique of conventional naturalism.

[71] Kriegel's (2009) phrase. Jehle and Kriegel (2006: 471) say such a relation must involve something 'actually happening' and be 'temporally thick', i.e. more than merely dispositional. Constitution clearly qualifies.

[72] Though it is tempting to co-opt the model Lau (e.g. 2008. See also Lau and Passingham 2006) develops for implementation of HOT theory, where dorsolateral pre-frontal cortex activity represents a signal in the visual cortex (for visual consciousness). On QHOT theory, the idea would be that the dlpfc provides the QHOT, and embeds (for display) the visual cortex activity, via projections from the latter to the former. That would give us a large composite brain state for the completed QHOT, but still some way short of the *global* brain states Van Gulick entertains. However see Gennaro (2012) ch. 9 for interesting critique of Lau's model of HOT implementation.

fiction devices offer an analogy: an imagined item is precisified enough to be functionally distinguished from actual and possible peers, but rarely any further. For instance, one specifies a *teleporter*, as distinguished from a spaceship or train, by saying one travels to one's destination instantaneously and discontinuously (without crossing every intervening spatial point). After that the internal mechanics, and implementation, of teleportation are moot, and these specifics are not part of theorising the model. The model's job is to guide and constrain the work of implementation, if we get round to producing the item.[73] In our case, we have functionally specified the QHOT structure with respect to its peers: in QHOT theory, uniquely, the sensory state is a proper part of the awareness-supporting state and its content, unaltered, is an improper part of the experienced content. By contrast, on non-relational HOT and sophisticated SR theories the sensory state is no part of the awareness-supporting state, and its content is no part of the experienced content.[74] Further, QHOTs make sensory states available to other cognitive systems, by presenting them. That might make the proposal sound like *access consciousness supports phenomenal consciousness*. Actually the relation is the reverse: this network role is the *upshot* of the quotational display, not its basis.[75] As Kriegel suggests, access consciousness is dispositional (cognitive *availability*), and demands a categorical basis, which is likely where phenomenal consciousness comes in. Any good theory of phenomenal consciousness should then have an eye for how this occurrent property supports access consciousness—in this respect QHOT theory excels. In fact, more than making a sensory state available to the wider network, in presenting it the QHOT positively *directs* network attention to it. In that sense the sensory content becomes endorsed by the system.[76] And this is what we find in consciousness: a sensory state placed unavoidably right before us, for use in executive action.

v. *Still, quotation relies on the audience's (or user's)* prior awareness *of the quoted item, so such a device cannot be used to* analyse *awareness.*

Quotation as we're familiar with it involves 'audience' awareness. But to claim consciousness of a mental state is needed so as mentally to quote it demands demonstration of some constitutive tie between quotation and sentience. We may

[73] Consider the development path to *tablets* from similar devices as featured on *Star Trek*.

[74] To complete the roundup: On simple SR theory the sensory state is an improper part of the awareness-supporting state, and its content is (likely) a proper part of the experienced content: we have a single state with 'two-faced' content. On Gennaro's WIV the sensory state is a proper part of the awareness-supporting state, if we take Gennaro at his word, but, though he strives to involve sensory content in consciousness, the tensions we observed in his position will likely result in his rejecting first-order content as part of experienced content. It is even possible that for Gennaro first-order states lack content of their own—since he does not believe in qualitative content existing independently of HO representation. On Van Gulick's HOGS theory the sensory state is a proper part of the awareness-supporting state, and its content—transfigured by HO representing—is a proper part of the experienced content.

[75] There's nothing incoherent in the idea of quotation or display that *happens* to find no audience.

[76] Similarly, regarding Dennett's (1991) idea of consciousness as 'cerebral celebrity', QHOT theory makes quotational display the ground of cerebral celebrity, instead of cerebral celebrity grounding consciousness.

compare the situation of *intentionality*: it used to be thought that intentionality presupposes consciousness (some still believe this). But it has proven profitable to the naturalistic project, notably to cognitive science, to suppose content can exist prior to consciousness. *Given* this assumption—helpful in modeling the interaction of unconscious with conscious contentful brain states—we try to provide naturalistic models of intentionality. But want of a worked-out implementation doesn't (broadly speaking) undermine widespread confidence that intentionality is not consciousness-dependent, and can be naturalised—indeed, that conscious intentionality is explicable in its terms. Similarly, when Newton published *Principia* he was accused of reinstating entelechies; it was felt, regarding the posit of gravity, that the only things that could tend towards one another were agential. It has, again, proven beneficial to naturalism to presume such movements don't demand agency, and attempt a physical model of gravity and other forces. The pattern is this: it often proves useful to naturalists to ascribe some capacity to the non-sentient, non-agential world hitherto attributed exclusively to sentient beings, in the hope, ultimately, of explaining how that world is set up to contain such beings. Having made this ascription, conceptual work is done so the attribute in question is no longer felt to require sentience or agency, and conceptual/empirical work is done towards modeling it naturalistically. QHOT theory advocates the same move regarding quotation, to explain consciousness. We must drop this ambition if it can be shown quotation *presupposes* consciousness, but we're within rights to await this argument and consider it in due course. For my part I don't find the notions of quotation (sampling), display or exhibition are incapable of an agent-free interpretation. A special danger lurks for analyses of consciousness: the closer we get to the right answer the greater the risk of the key ingredient appearing to presuppose consciousness; for, if the analysis is correct, that ingredient entails consciousness. One misstep and we are apt to confuse entailment for presupposition. The claim of QHOT theory is that mental quotation certainly entails, but does not presuppose, consciousness. To be clear, quotation may require mentality—if so non-conscious mentality will do.

vii. *Are QHOTs literal linguistic structures?*

They may be, e.g. if language of thought theory is found plausible. But arguably a mere demonstrating pointer suffices: quotation needn't be linguistic (consider finger quotes). The pointer might be understood, like Rosenthal's HOTs, as an assertoric thought—since it can only display an *occurrent* sensory state to the subject. That might count as a non-linguistic species of assertion, and would provide the indexical elements of linguistic presentation without need of language. But QHOT theory also resembles *higher-order perception theory*, not least in that first-order states are *presented* not represented.[77] It seems possible to read QHOT theory as more HOT- or HOP-like according to one's preference. Still, it has been argued, and I sympathise, that the difference between HOT and HOP theories is ultimately

[77] Van Gulick (2000). Still proponents hold, what QHOT theory denies, that the HO component fixes experienced content (Lycan 1996).

superficial.[78] A major dissimilarity might appear the employment of *concepts* in HOTs—but some consider perception to be conceptual. Again, *perceptions* have been counted *thoughts*, under a wide understanding. So I am content to retain the name quotational higher-order *thought* theory and leave open the issue whether QHOTs involve concepts.[79] Some object that animals, also children, may lack the cognitive sophistication for HO structures. First, this is no special problem for QHOT theory, and I defer to the defences of other HO theorists.[80] Second, QHOT theory if anything makes animal/infant consciousness *easier* to compass: QHOTs are non-representational, and, on one construal, needn't involve concepts.

14. Though not a HO *representational* theory, QHOT theory remains a HO theory of consciousness. Many first-order sensory states lie outside consciousness. It takes a mental operation on mental states—quotational display—to make them conscious.[81]

[78] See e.g. Van Gulick (2000), Gennaro (2004).

[79] One of Gennaro's reasons for including concepts in HOTs is that 'A conscious state *must* be presented to its owner in some way or other' (2012: 86). But QHOTs precisely do not colour the content they make conscious. *But thoughts are composed of concepts, so how is this a higher-order* thought *theory?* That concepts indeed compose thoughts doesn't imply some thoughts aren't constituted differently, as per the 'pointer' described above. It follows that phenomena Gennaro and Rosenthal explain via HOTs' concepts (e.g. experienced wine-tasting) QHOT theory must explain using FO states' concepts. An account is beyond the present paper: but any conceptual footwork conducted at the HO level can be performed down below instead.

[80] E.g. Gennaro (2004)—not least, he reminds us (§1) that animals and infants *have* cortexes.

[81] Picciuto (2011) develops a superficially similar HO account of consciousness, 'the quotational view' (QV). Though QHOT theory and QV agree consciousness involves embedding a sensory state in a HO state, they differ in crucial respects:

i. For Picciuto *phenomenal concepts* are implicated in consciousness, not just in thoughts about experiences as per the Papineau/Balog account. But QHOTs may not be conceptual at all, and certainly don't involve phenomenal concepts (concepts of phenomenal states). If there are no phenomenal concepts (as Tye 2009 argues) then QV fails, whereas QHOT theory is untouched. ii. Picciutto understands his HO quotational structures as 'demonstratives' which *refer* to embedded sensory states and *represent* them as states of the subject. QHOTs demonstrate sensory states but are not demonstratives, nor do they refer to embedded sensory states, and they do not represent them (let alone as states of the subject)—but simply *present* them. iii. Picciutto says quotational embedding 'activates' a sensory state, suggesting an intrinsic modification, like acquisition of an 'inner glow'. On QHOT theory awareness is held just to *consist in* a sensory state's mental quotation, with said state intrinsically unaltered. Piciutto's talk of activation recalls constitutive HO accounts, with associated problems. iv. On QV the quotational HO structure is 'part of' the 'conscious state' (§5.3), meaning we are conscious of the quotational element. On QHOT theory we are conscious only of what the QHOT embeds, i.e. the sensory state. The QHOT itself is not conscious.

This creates a problem for QV. Picciutto's hypothesis is that quotational embedding in a phenomenal concept makes a mental state conscious. But if we are conscious of the quotational element of a completed consciousness-enabling phenomenal concept, then, by the hypothesis, that quotational frame must be embedded in another quotational phenomenal concept (assuming it cannot 'self-quote'). If *this* concept's quotational element is also conscious then it requires embedding in a third phenomenal concept, and we are headed for an unhelpful regress, to explain how the quotational element of one of Picciutto's concepts can be conscious. The other, more plausible, option is that the quotational element of a quotational phenomenal concept does not require quotational embedding to be conscious. But then it follows that being quotationally embedded isn't needed for consciousness, against QV. QV thus faces a dilemma: infinite regress or explanatory failure. Since an infinite regress never arrives, we actually have explanatory failure on both horns. QHOT theory avoids this by denying QHOTs are conscious. Piciutto *might* posit a phenomenal concept, embedding a sensory state, that is then embedded in a further phenomenal concept which is unembedded, so unconscious. This *unwieldy* structure would be erected to

Acknowledgments Thanks to Uriah Kriegel, Tom McClelland, Donnchadh O'Connaill, Raymond Lutra, and an anonymous reviewer for helpful comments, suggestions and objections. Thanks to David Rosenthal for much useful related discussion.

References

Balog, K. (2012). Acquaintance and the mind-body problem. In C. Hill & S. Gozzano (Eds.), *New perspectives on type identity: The mental and the physical*. Cambridge: Cambridge University Press.

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

Brown, R. (2012) Review of Rocco J. Gennaro, *The consciousness paradox: Consciousness, concepts, and higher-order thoughts*. Notre Dame Philosophical Reviews. http://ndpr.nd.edu/news/30848-the-consciousness-paradox-consciousness-concepts-and-higher-order-thoughts/

Byrne, A. (1997). Some like it HOT: Consciousness and higher-order thoughts. *Philosophical Studies, 86*, 103–129.

Capellen, H. and Lepore, E. (2012) 'Quotation' in *The Stanford encyclopaedia of philosophy*. http://plato.stanford.edu/entries/quotation/

Caston, V. (2002). Aristotle on consciousness. *Mind, 111*(444), 751–815.

Chalmers, D. J. (2003). The content and epistemology of phenomenal belief. In Q. Smith & A. Jokic (Eds.) *Consciousness: New philosophical perspectives*. New York: Oxford University Press.

Davidson, D. (1979) 'Quotation', in his *Inquiries into truth and interpretation* (pp. 79–92). Oxford: Oxford University Press.

Dennett, D. (1991). *Consciousness explained*. Boston: Little Brown.

Dretske, F. (1995). *Naturalizing the mind*. Cambridge, MA: The MIT Press.

Feinberg, T. E. (2000). The nested hierarchy of consciousness: A neurobiological solution to the problem of mental unity. *Neurocase, 6*, 75–81.

Frege, G. (1892/1962) On sense and reference. In M. Black & P. T. Geach (Eds.) *Philosophical writings*. Oxford: Basil Blackwell.

Gennaro, R. J. (2004). Higher-order thoughts, animal consciousness, and misrepresentation: A reply to Carruthers and Levine. In R. J. Gennaro (Ed.), *Higher-order theories of consciousness: An anthology*. Amsterdam: John Benjamins.

Gennaro, R. J. (2006). Between pure self-referentialism and the (Extrinsic) HOT theory of consciousness. In U. Kriegel & K. Williford (Eds.), *Consciousness and self-reference*. Cambridge, MA: The MIT Press.

Gennaro, R. J. (2012). *The consciousness paradox*. Cambridge, MA: The MIT Press.

Gertler, B. (2001). Introspecting phenomenal states. *Philosophy and Phenomenological Research, 63*(2), 305–328.

Goldman, A. (1993). Consciousness, folk psychology, and cognitive science. *Consciousness and Cognition, 2*, 364–382.

Horgan, T., Graham, G. (2009) Phenomenal intentionality and content determinacy. In R. Schantz (ed.) *Prospects for meaning*. Amsterdam: de Gruyter.

Jehle, D., & Kriegel, U. (2006). An argument against dispositional HOT theory. *Philosophical Psychology, 19*, 462–476.

Kriegel, U. (2009). *Subjective consciousness*. New York: Oxford University Press.

Lau, H. C. (2008). A higher order bayesian decision theory of consciousness. In R. Banerjee & B. K. Chakrabarti (Eds.), *Models of brain and mind: Physical, computational, and psychological approaches*. Amsterdam: Elsevier.

Footnote 81 continued
support the claim that we are conscious of the HO quotational feature QHOT theory and QV consider key to consciousness: I don't find this phenomenology (cf. Gennaro 2012 Ch. 5).

Lau, H. C., & Passingham, R. (2006). Relative blindsight in normal observers and the neural correlate of visual consciousness. *Proceedings of the National Academy of Sciences of the United States of America, 103*, 18763–18768.
Levine, J. (2001). *Purple Haze*. New York: Oxford University Press.
Loar, B. (1995). Reference from the first-person perspective. *Philosophical Issues, 6*, 53–72.
Lycan, W. G. (1996). *Conscious experience*. Cambridge, MA: MIT Press.
Lycan, W. G. (2001). A simple argument for a higher-order representation theory of consciousness. *Analysis, 61*(269), 3–4.
Neander, K. (1998). The division of phenomenal labour: A problem for representational theories of consciousness. *Philosophical Perspectives, 12*, 411–434.
Papineau, D. (2002). *Thinking about consciousness*. Oxford: Oxford University Press.
Papineau, D. (2007). Phenomenal and perceptual concepts. In T. Alter & S. Walter (Eds.), *Phenomenal concepts and phenomenal knowledge: New essays on consciousness and physicalism*. Oxford: Oxford University Press.
Picuitto, V. (2011) Addressing higher-order misrepresentation with quotational thought. *Journal of Consciousness Studies* 18, 3–4: XX–XX.
Reimer, M. (1996). Quotation marks: Demonstratives or demonstrations? *Analysis, 56*(3), 131–141.
Rosenthal, D. (2004). Varieties of higher-order theory. In R. J. Gennaro (Ed.), *Higher-order theories of consciousness*. Amsterdam: John Benjamins.
Rosenthal, D. (2005). *Consciousness and mind*. Oxford: Oxford University Press.
Searle, J. (1969). *Speech acts*. Cambridge: Cambridge University Press.
Searle, J. (1983). *Intentionality*. Cambridge: Cambridge University Press.
Simons, P. (1987). *Parts: A study in ontology*. Oxford: Clarendon Press.
Stubenberg, L. (1998). *Consciousness and qualia*. Amsterdam: John Benjamins.
Tye, M. (2000). *Consciousness, color, and content*. Cambridge, MA: The MIT Press.
Tye, M. (2009). *Consciousness: Revisited: Materialism without phenomenal concepts*. Cambridge, MA: The MIT Press.
Van Gulick, R. (2000). Inward and Upward: reflection. *Introspection, and Self-Awareness', Philosophical Topics, 28*(2), 275–305.
Van Gulick, R. (2004). Higher-order global states (HOGS): An alternative higher-order model of consciousness. In R. J. Gennaro (Ed.), *Higher-order theories of consciousness: An anthology*. Amsterdam: John Benjamins.
Washington, C. (1992). The identity theory of quotation. *Journal of Philosophy, 89*, 582–605.
Weisberg, J. (2008). Same old, same old: the same-order representation theory of consciousness and the division of phenomenal labor. *Synthese, 160*, 161–181.
Zemach, E. M. (1985). De se and descartes: A new semantics for indexicals. *Noûs, 19*(2), 181–204.