

Vincent Picciuto

Addressing Higher-Order Misrepresentation with Quotational Thought

Abstract: *In this paper it is argued that existing ‘self-representational’ theories of phenomenal consciousness do not adequately address the problem of higher-order misrepresentation. Drawing a page from the phenomenal concepts literature, a novel self-representational account is introduced that does. This is the quotational theory of phenomenal consciousness, according to which the higher-order component of a conscious state is constituted by the quotational component of a quotational phenomenal concept. According to the quotational theory of consciousness, phenomenal concepts help to account for the very nature of phenomenally conscious states. Thus, the paper integrates two largely distinct explanatory projects in the field of consciousness studies: (i) the project of explaining how we think about our phenomenally conscious states, and (ii) the project of explaining what phenomenally conscious states are in the first place.*

1. Introduction

Proponents of traditional higher-order representational theories contend that a mental state is phenomenally conscious in virtue of being appropriately represented by a numerically distinct mental state. I call such theories ‘extrinsic’ higher-order theories. While extrinsic theories have many virtues, they face a plethora of objections.¹ Many of these are soluble given the resources of an extrinsic theory, but one objection in particular is thought by some to be decisive. This is the

Correspondence:
Email:

[1] See Gennaro (2004), Kriegel (2006), Van Gulick (2006), and Carruthers (2009) for discussion of a range of objections to extrinsic theories.

problem of higher-order misrepresentation.² So-called ‘self-representational’ theorists, or what I refer to as ‘intrinsic’ higher-order theorists, are sympathetic to the general higher-order framework but claim that extrinsic views cannot adequately address higher-order misrepresentation.³ Intrinsic theorists claim that their own views can.

This paper assumes the plausibility of the general higher-order framework but argues that whatever other virtues these theories might enjoy, no existing version of either kind of higher-order theory adequately addresses the problem of higher-order misrepresentation. Of the four representative versions of intrinsic theory to be examined below, none can accommodate higher-order misrepresentation without inheriting commitments that further complicate the view, or that collapse the view into explanatory inadequacy.⁴

Drawing a page from the ‘phenomenal concept strategy’, I propose a novel intrinsic higher-order representational theory that does address higher-order misrepresentation.⁵ This is what I call the ‘quotational theory’ of phenomenal consciousness, according to which the higher-order dimension of a conscious state just is the quotational component of a quotational phenomenal concept. Being quotational, such a conscious state is partly constituted by the (non-conscious) experience to which its higher-order component refers. According to the quotational theory of phenomenal consciousness, phenomenal concepts actually help to account for the very nature of phenomenally conscious states. This is a radical departure from the received view of the phenomenal concept strategy — the received view is expressly intended to be independent of any particular account of the nature of consciousness.

The paper has the following structure. In section two I briefly lay out the critical difference between extrinsic and intrinsic higher-order representational theories of phenomenal consciousness. In section three I review the problem of higher-order misrepresentation. In section four I critique existing intrinsic accounts. In section five I present the quotational account. In section six I conclude.

-
- [2] See Byrne (1997), Neander (1997), and Levine (2001). For the purposes of this paper I will assume that higher-order misrepresentation is a genuine problem for extrinsic views. The character of the problem will be spelled out in section three.
- [3] Van Gulick (2004; 2006), Gennaro (1996; 2006), Carruthers (2000; 2005), and Kriegel (2005; 2009) are current defenders of self-representationalism, but the view has a long pedigree and can be found in Brentano (1874/1995) and Husserl (1928/1964), and at least one author claims Aristotle also held a self-representational view (Caston, 2002).
- [4] Weisberg (2008) argues similarly.
- [5] The phrase ‘phenomenal concept strategy’ was introduced by Stoljar (2005).

2. Extrinsic and Intrinsic Higher-Order Theories

Since there is some potential for terminological confusion, two initial clarifications must be made. First, philosophers use ‘phenomenal’, ‘qualitative’, ‘subjective’, and ‘consciousness’ in different ways and it may be unclear exactly what phenomenon ‘phenomenal consciousness’ is even supposed to pick out.⁶ At present I take the phenomenon to be best captured (for better or for worse) by the dreaded phrase ‘what it’s like’ and all of its variations. States that are phenomenally conscious have phenomenal character; they are ‘like something’ for their subjects. Such states have ‘what-it’s-likeness’. As I explain below, following other higher-order theorists I take phenomenal character to be something different than a state’s having mere qualitative character. Throughout the following, for the sake of elegance I will, unless the context requires disambiguation or emphasis, use ‘consciousness’ as shorthand for ‘phenomenal consciousness’.

Second, the typical distinction in the literature between traditional higher-order theory and self-representational theory is somewhat misleading; it suggests that the main difference between the two kinds of theory is that one is ‘higher-order’ (includes a metarepresentational commitment) and that the other is ‘same-order’ (does not include a metarepresentational commitment). But *both* so-called ‘traditional’ higher-order theorists and self-representationalists are committed to some kind of metarepresentationalism. They differ on whether the metarepresentational component to which they appeal is numerically distinct from, and extrinsic to, the conscious state or whether it is intrinsic to the conscious state itself (more on the metarepresentational commitment of both kinds of theory below). For this reason I take the four representative versions of ‘self-representational’ theory discussed below to be members of a subset of higher-order theory and, at the risk of adding to the terminological quagmire of contemporary philosophy of mind, I will use the extrinsic/intrinsic higher-order dichotomy rather than the higher-order/self-representational dichotomy.⁷

[6] See Block (1995). Even theorists who agree on substantive issues use these terms differently. E.g. Rosenthal’s thin *vs.* thick phenomenality, Lycan’s low-order *vs.* higher-order what-it’s-likeness, and Carruthers’ worldly *vs.* experiential subjectivity. See Byrne (2004) for an illuminating discussion concluding that these terminological distinctions are intended to carve nature at the same joints.

[7] For the phrase ‘extrinsic higher-order theory’ I am indebted to Gennaro (2006), wherein he draws a similar distinction between extrinsic higher-order theory and what he calls ‘pure self-referentialism’.

According to the ‘traditional’ extrinsic theorist, a mental state M of a subject S is conscious if and only if M represents and is itself represented by a numerically distinct, extrinsic state M^* . For example, I am in a conscious mental state representing the blue sky if and only if I am in a state representing the blue sky that is itself appropriately represented by a numerically distinct extrinsic state. Extrinsic higher-order theorists argue that state consciousness just is being conscious *of* the relevant state. As David Rosenthal puts it, ‘We are conscious of ourselves as being in qualitative states simply by having HOTs [higher-order thoughts] that represent us as being in such states’ (2005, p. 14). Also, it is important to notice that a higher-order representation M^* does not make its target state M conscious by effecting any intrinsic changes in M . Rather, M is a conscious state merely in virtue of standing in the appropriate representation relation to M^* .⁸

Phrased generally, the core assumption of a ‘self-representational’, or intrinsic higher-order theory is that a conscious mental state is a complex mental state representing both the world and itself (or at least one of its own parts).⁹ More precisely, a conscious mental state has two critical components: a low-order component representing some feature of the subject’s environment (construed broadly to include the subject’s body) and a higher-order component representing the lower-order component. Thus characterized, one can see that self-representationalists build upon the foundation laid by traditional higher-order theorists. As representationalists, both extrinsic and intrinsic theorists are obviously committed to some degree of representationalism regarding consciousness. Both tend to think that representationalism will facilitate the naturalization of consciousness. If content can be explained naturalistically and consciousness can be explained in terms of content, then consciousness itself can be naturalistically, and perhaps reductively, explained (in some sense of ‘reduction’, e.g. explained in non-phenomenal terms that are amenable to scientific explanation). While representationalists might differ on the specific theory of content recruited to do the explaining, all representationalists are committed to the claim that consciousness is best explained (somehow) in terms of content (and/or causal/functional role).

In addition to the shared commitment to representationalism, both extrinsic and intrinsic theorists are committed to some kind of

[8] See also Lycan (1996; 2001).

[9] There are various ways in which the relation between the relevant parts of a conscious state might obtain. See Kriegel (2006) for a survey of the possibilities.

metarepresentationalism regarding consciousness. This commitment is best captured by William Lycan's (2001) 'simple argument' for higher-order theory. The point of Lycan's argument that is most relevant to the current discussion is the idea that awareness or perception (whether 'conscious' or not) requires representation. But just as a putative non-conscious percept requires a 'first order' of representation, a putative conscious percept requires a 'second order' of representation—a representation of that non-conscious percept. This second commitment common to both extrinsic and intrinsic views brings out a related commitment that may be held with varying degrees of strength. Both kinds of higher-order theorist are committed to some version of what Neander (1997) calls the 'division of phenomenal labor', according to which first-order representational properties are supposed to determine, and explain, qualitative character (how it is that a state *M* represents anything sensory in the first place) and higher-order representational properties are supposed to determine, and explain, phenomenal character (the what-it's-likeness of *M*'s subject being in *M*). Again, since according to extrinsic views the higher-order representation effects no intrinsic change in its target state, it is important to notice that the higher-order properties that determine the phenomenal character of the state *M* are the properties of a numerically distinct, extrinsic state *M**.¹⁰

3. Higher-Order Misrepresentation

As stated in the last section, extrinsic theorists argue that a mental state is conscious in virtue of there being an appropriate representation of that state by some other mental state, all of which is supposed to be explainable naturalistically. But since a naturalistic representational system can always malfunction, higher-order representations can misrepresent their targets. This opens up two possible complications for extrinsic higher-order theory.

It should be emphasized that the complications to be discussed are specifically about *higher-order* misrepresentation: they are due to the fact that the constraints of the theories under discussion leave open the

[10] This is how the division is typically drawn, but, as stated above, not all extrinsic theorists use the same terminology, nor do all intrinsic theorists put it quite this way. For example, Kriegel would characterize the division in the following way. First-order properties *represented in a conscious experience* determine the qualitative character. Higher-order properties determine the subjective character, and the complex itself (the qualitative character integrated with the subjective character) determines the phenomenal character of the state. On the other hand, for Kriegel, non-conscious perceptual states have something quite like, but that falls short of, qualitative character: 'schmalitative character' (2009, pp. 109–110).

possibility of a higher-order state (or component) misrepresenting its first-order target, and as this paper assumes, that opens up difficulties for these views. That is not to say that we do not sometimes mistake certain states for others. For example, we do seem sometimes to mistake tickles for pains, as when one waits in anticipation for the dentist's drill; or pleasant tastes for unpleasant ones, as when one expects cola but unknowingly sips orange juice. But these kinds of mistakes can be explained as *first-order* misrepresentations that are partly conceptualized erroneously prior to becoming conscious. Mistakes of that sort, which are distinct from higher-order misrepresentation, must be accommodated by a representational theory, not ruled out. Higher-order misrepresentation, on the other hand, is problematic and any adequate higher-order account of consciousness must either rule it out or explain why it need not be ruled out.

The first complication is that a higher-order state can misrepresent an actual first-order state, e.g. a first-order state that actually represents a circle could be misrepresented as representing a square. The difficulty for extrinsic higher-order theorists then is to explain what such a state will be like. On the extrinsic view, phenomenal properties are determined by the higher-order state, but that means that the state representing the circle would be just like a state representing a square. The two states would be indistinguishable.¹¹ But even worse, according to extrinsic views, for my circle-representing state to be phenomenally conscious (have circle what-it's-likeness) *just is* for it to be appropriately represented by a higher-order state. In the case under consideration, though, that higher-order state is misrepresenting it as a square. But why then would that make my *circle*-representing state *phenomenally* conscious? There is no circle (state)-representing higher-order state. If state consciousness amounts to a subject being aware of *that* state and phenomenal consciousness is taken to be a kind of state consciousness, then the extrinsic theorist must conclude that I am aware of my circle-representing state (my circle-representing state is phenomenally conscious, or has circle what-it's-likeness) in virtue of my being aware of what seems to be an altogether different state: a square-representing state. Here is another way to put it that brings out the incoherence of the consequence illustrated by the previous sentence. The extrinsic theorist must assert that my circle-representing state is phenomenally conscious (read 'like circle' for me) but like square for me.

[11] Kriegel (2009) develops this line of argument.

The second and even more devastating complication is that a higher-order state could represent an altogether non-existent first-order state. For example, a subject may have a higher-order representation that she is in a state representing a square when she is in fact not in any corresponding lower-order state representing anything at all. The extrinsic theorist must either claim that in such a case that the subject would experience nothing at all or accept that the subject would experience squarishness. The first option seems *ad hoc*, though. For by hypothesis of extrinsic theory, to be state conscious just is to be aware of the relevant state, and to be aware of a state just is to have a higher-order representation that one is in that state, then we would need some explanation for why in this case it would not be like anything for the subject to be in the non-existent state that she (her system) is representing herself to be in.

Byrne (1997) argues that extrinsic theorists must accept the second option; if phenomenal consciousness just is being aware (having a higher-order representation) of some first-order state, as the extrinsic theorist argues, then because the subject higher-order represents that she is in a state representing a square, the extrinsic theorist must accept the conclusion that it would be squarish for the subject to be in the higher-order state. The problem is that extrinsic theories are theories of *state* consciousness, and to avoid regress, the state that is supposed to be rendered conscious by its relation to a higher-order representation is the target state, not the higher-order state itself. But in this case there is no such target state, so which state is *the* conscious state? In such a case, if it must be like something for the subject, and it being like something amounts to the subject being in a conscious state, then the conscious state would have to be the higher-order state (for that is the only other relevant state the subject is in). However, this is not a conclusion the extrinsic theorist can accept; it amounts to admitting that consciousness can be explained in terms of a single state, and that undermines the core relational strategy of extrinsic views.¹²

Of course, extrinsic theorists do not quietly accept that these consequences follow from their views, nor that such consequences are as urgent as their objectors claim, even if they do follow. For example, Rosenthal and Lycan have both defended extrinsic higher-order theory against misrepresentation charges. Rosenthal (1997, p. 744; 2004, p. 40) argues that the higher-order state is sufficient to explain phenomenal awareness. The higher-order state determines what it's like

[12] See Neander (1997) and Levine (2001) for similar misrepresentation arguments against extrinsic higher-order theory.

for the subject, even if there is no such first-order target state. For Rosenthal, you either have a HOT (and are in a conscious state) or you do not.

Lycan (1996, p. 20) argues similarly. He says the mere theoretical possibility that certain kinds of higher-order misrepresentation can occur is not as serious a problem as it initially appears. Even if such cases are possible, and perhaps such cases do in fact happen, Lycan claims they may be quite rare. According to Lycan, both mismatching and targetless higher-order misrepresentation are theoretically possible, but in fact, quite rare (*ibid.*)

But even if such cases of higher-order misrepresentation are rare, as Lycan claims, one might take the fact that extrinsic theories allow for their mere theoretical possibility to count strongly against such views. For if state consciousness just is one's awareness (or seeming awareness) of being in a state, as Rosenthal argues, the question 'Which state is *the* conscious state in the first place?' remains unanswered by simply accepting the possibility that targetless higher-order thoughts issue in state consciousness. Moreover, it will not do to claim that such extrinsic views are naturalistic theories that merely describe phenomenal consciousness, because typically an inconsistency in the described constraints of a (naturalistic) theory at least suggests that something has gone wrong somewhere in the description. Indeed, the intrinsic theorists discussed in the next section find these responses wanting. They each take higher-order misrepresentation to be a major problem for which extrinsic theorists cannot account and which intrinsic theories are designed at least in part to rule out.

4. Intrinsic Higher-Order Attempts to Address Higher-Order Misrepresentation

4.1 Van Gulick's HOGS

According to Van Gulick's 'higher-order global states' (HOGS) model, a conscious mental state is a complex global state whose organization and content 'embodies a heightened degree of *reflexive self-awareness*' (2006, p. 24). For Van Gulick, the self-awareness of conscious states is built up out of implicit meta-intentionality. To account for implicit meta-intentionality, Van Gulick appeals to a teleosemantic version of consumer semantics, according to which the content of a mental state is partly determined by the specific role it has been adapted to play. Van Gulick argues that sub-personal states which underlie more complex states (even conscious states) are already meta-intentional. Such states are meta-intentional in so far as

they have ‘specifically adapted to the intentional nature and content of the states and processes to which they apply’, but such states need not be as cognitively complex as extrinsic higher-order theorists tend to assume, e.g. they are exemplified by the kinds of sub-personal states that underlie certain basic learning processes (*ibid.*, p. 21–22).¹³

With the above mentioned view of meta-intentionality in place, Van Gulick appeals to the ‘global workspace’ theory of consciousness. On the most developed global workspace theory, the mind is organized around the global availability or ‘broadcast’ of information in the brain to specialist sub-systems. The result of a global broadcast is to make the contents being broadcast widely available to various processing systems, which, according to such theories, thereby makes the states representing those contents conscious.¹⁴

Van Gulick attempts to scale-up low-level meta-intentionality to the meta-intentionality that constitutes consciousness. The basic idea is to take a world-representing state and recruit it into the global workspace. Give it ‘cerebral celebrity’ and in virtue of its links to other systems downstream to which it has adapted, the global state becomes conscious in virtue of the higher-order content that is implicitly represented by its links to various other states or systems.¹⁵ Moreover, on Van Gulick’s view the higher-order component of a conscious state is not provided by a wholly distinct extrinsic state. Both the lower-order components and the higher-order component are parts of the same state.

The crucial question is: could a conscious HOGS higher-order misrepresent? Because the original state is recruited into the global state, which thereby renders the state conscious, the answer seems to be ‘no’. This would provide a straightforward way of ruling out the possibility of a higher-order state misrepresenting an actual lower-order target. It would also provide a straightforward answer to the question introduced above that the extrinsic theorist seemed unable to answer. Namely, when targetless higher-order representations arise, which

[13] See also Van Gulick (1988) for a discussion of Garcia’s rats, and Garcia and Koelling (1967).

[14] See Baars (1988). For Baars, global broadcast was supposed to explain consciousness, and perhaps it does explain *one* kind of consciousness, *viz.* what some philosophers have dubbed *access* consciousness, or that a mental state is conscious in the sense that it is accessible to inform reasoning, decision-making, and generate verbal report. However, many philosophers deny that Baars’ theory explains the ‘what it’s like’ aspect, or that a state is *phenomenally* conscious. For the distinction between access and phenomenal consciousness see Block (1995) and Chalmers (1996). For a critique see Church (1998).

[15] For the origins of the cerebral celebrity idea see Dennett (1991; 1992; 2001), and Dennett and Kinsbourne (1995a,b).

state, exactly, is *the* conscious state? Van Gulick's answer is that in such cases there is no conscious state, because there is no state that gets globally integrated in the first place. You just cannot have the relevant sort of higher-order state without some first-order component.

While Van Gulick's view seems straightforwardly to rule out higher-order misrepresentation, the explanation of consciousness that we are left with is not so straightforward; it is explanatorily inadequate in a much more significant way. First, one's conscious awareness of a mental state seems to require an explicit representation of that state.¹⁶ On Van Gulick's view, though, there need be no actual occurrent explicit higher-order representation, but merely the global availability of some content. This seems to require commitment to the claim that consciousness, a seemingly categorical property, is grounded in a dispositional property: the property of being globally available to a wide range of states or consumer systems. That claim requires further argument to establish. Indeed, some find the claim implausible.¹⁷ Thus, even if Van Gulick's HOGS model does rule out higher-order misrepresentation, the model we are left with does not sufficiently explain consciousness, which is the explanatory goal of the model in the first place.¹⁸

4.2 Gennaro's WIV

According to Gennaro's 'wide intrinsicity view' (WIV), we should individuate conscious states widely. That is, *contra* extrinsic theorists, who contend that the conscious state of an M/M* relation is the lower-order state M, Gennaro claims that we should consider the two states to be a sort of global state. To put it visually, we draw a conceptual circle around M and M* and consider that conjunction to be *the* conscious state. Gennaro says, 'My WIV does not treat the conscious rendering state as entirely distinct from CS... Rather, it treats conscious states as complex states with both CS and the meta-psychological states

[16] Weisberg (2008) argues similarly, but he argues that implicit representation does not capture the common sense understanding of awareness. I agree, but I think the more serious problem is that it does not adequately *explain* consciousness in terms of common sense, folk psychology, a cognitive-scientific theory, nor a philosophical theory.

[17] This is a problem raised for Carruthers' account by Jehle and Kriegel (2006, to be discussed below), but it is equally a problem for Van Gulick's account.

[18] It might be thought that one can attempt to address higher-order misrepresentation independently of the so-called 'hard problem' of consciousness (as an anonymous referee suggests). However, since theories attempting to address the hard problem are just the theories for which the problem of higher-order misrepresentation arises, it will not suffice to account for higher-order misrepresentation at the expense of an inadequate account of phenomenal consciousness, which is the primary explanandum in the first place.

as parts. Conscious states are individuated widely so as to treat the meta-psychological state as intrinsic to the conscious mental state' (1996, p. 16).¹⁹ And so, on Gennaro's view, we are supposed to have one state (in some sense of 'one state') with both low-order world-representing parts and a higher-order state-representing part, and that makes the view an intrinsic theory.

Kriegel (2005; 2006) maintains that Gennaro's theory makes consciousness purely 'notional'. We are merely to 'take' or 'consider' both states in conjunction to be *the* conscious state. If that is the case, then Gennaro's view smacks of arbitrariness; we could 'take' any two states we so desire to be a state of the system. Given that Gennaro claims the two states must be 'importantly related', though, presumably Gennaro has something more robust than mere consideration in mind (Gennaro, 1996, p. 16). I take it that what he means by 'individuating states widely', or 'considering' two states in certain cases to be one global state, is considering *from the perspective of a theory* that attempts to accurately describe the actual structure of a conscious state. If that is what Gennaro has in mind, then we should understand him as saying that we ought to consider what might appear to be two (local) states actually to be one (global) state, *if it is theoretically fruitful*. But notice that the extrinsic theorist, too, asserts that a higher-order state and its lower-order target must be 'importantly related' somehow.²⁰ It is just that intrinsic theorists argue for a tighter (constitutive) relation between the two. The problem is that the crucial theoretical task of adequately explaining exactly how the lower-order and higher-order components must be 'importantly related' is something that neither the extrinsic theorist nor Gennaro complete, and so, while Gennaro's account might gesture in the right direction, it leaves unanswered the crucial question facing both kinds of theorist.

At one point Gennaro claims that the 'understanding unconsciously "synthesizes" the raw data of experience to produce the resulting conscious state' (2006, p. 237). But it is hard to see how a vague notion of 'synthesis' will suffice. What is synthesis, exactly? Can it be explained naturalistically? One option that Gennaro suggests is that synthesis is realized by neural feedback loops (2004, pp. 62–63), but neural feedback loops are merely how the process might be instantiated in the brain. They do not provide the kind of representationalist account that Gennaro seems to be after. Moreover, it is not at all clear how neural

[19] While Gennaro uses the expression 'complex', the WIV provides no account of what distinguishes a complex from a mere conjunction. For the distinction between conjunction and a complex see Simons (1987).

[20] See e.g. Rosenthal (1997, p. 744).

feedback loops would be related to a conscious complex anyway, such that they would *explain* the required relation adequately.

This is not the place to defend Gennaro's account against the charge of being merely notional, though, because, even if wide individuation could be made more robust by neural feedback loops or some other process, Gennaro tells us nothing at all about why the wide individuation of what are seemingly two importantly related states would rule out higher-order misrepresentation; e.g. there is no explanation for why a low-order state representing a circle and a higher-order representation of that state as representing a square could not be taken to be importantly related, and thereby, taken to be components of the same global state. If so, we would still need an account that rules out such mismatches, or explains what such states are like in a satisfying way. Thus, even if Gennaro can answer Kriegel's objection that his view is merely notional, Gennaro's view does not adequately address the problem of higher-order misrepresentation.

4.3 Carruthers' Dual-Content Theory

In certain ways Carruthers' view is part extrinsic theory and part intrinsic theory. On the one hand, according to Carruthers, a state is conscious if and only if it is available to a distinct extrinsic state. On the other hand, the lower-order state itself acquires higher-order content. So while on a typical extrinsic view higher-order content is the content of a distinct extrinsic representation M^* , on Carruthers' view the higher-order content is actually integrated into (but implicitly represented by) the low-order state M itself. More specifically, for a low-order state M to be conscious, there need not be an actual occurring higher-order state M^* that targets M . Rather, M is conscious in virtue of being available to a higher-order thought producing faculty (the mind-reading system). The resulting state M (in virtue of its availability to the mind-reading system) acquires and implicitly represents higher-order content, and thus, possesses dual-content: the state possesses first-order world-representing content as well as higher-order experience-representing content.

Consider a simple example of a first-order non-conscious perception of red, the content of which can be symbolized as '[red]'. When that red-representing state (call it ' M ') is made available to the mind-reading system, which is capable of producing higher-order thoughts, M acquires a higher-order content something like [experience of red] or [seems red]. Now M has two contents: {[red], [seems red]}.

On Carruthers' view, higher-order contents are integrated into the original lower-order state M, but it is important to notice that the higher-order content is represented *implicitly*, simply in virtue of the availability of M to the relevant downstream consumer system, in particular, the mind-reading system. There is no explicit (structured) higher-order representational content that actually gets integrated into M. Availability is supposed to be what enables M to acquire higher-order content on the assumption of some version of consumer semantics, according to which for a state to have content is at least in part a matter of how the relevant downstream cognitive systems, or 'consumers' might use that state.

Since phenomenal character is parasitic on qualitative character in that the very qualitative content of the target low-order representation M itself comprises one of the dual-contents that render M conscious, Carruthers' account seems to rule out higher-order misrepresentation. However, upon further reflection it is not clear that the account does; for it leaves open the possibility that a downstream system might 'misconsume' a lower-order content, such that the higher-order content acquired in virtue of that use would not match the lower-order component of the state.

Moreover, the truth of Carruthers' view relies on the truth of some version of consumer semantics: the mind-reading system generates higher-order thoughts about other mental states, but for a mental state to be conscious on a given occasion the system need not *actually* generate any such thought. To actually acquire higher-order content, the target state need only be *available* to the mind-reading system. Without the assumption of consumer semantics Carruthers' view crumbles.²¹ It is in this sense that Carruthers' theory is 'dispositional'. It therefore faces the same problem as Van Gulick's theory: it bases consciousness, a seemingly categorical property, on availability, an apparently dispositional property. Some philosophers have argued that the dispositional component underlying the commitment to consumer semantics renders Carruthers' view implausible.²² Even if Carruthers' view is not *implausible*, it is quite difficult to wrap one's head around the idea that when one undergoes an occurrent conscious state, that the seemingly categorical property of being conscious is actually grounded in a dispositional base, *viz.* being merely available.

[21] It should be noted that Carruthers does not merely assume consumer semantics without argument. He defends it and claims elsewhere that it should be endorsed as a whole by cognitive science. See Botterill and Carruthers (1999).

[22] See Jehle and Kriegel (2004), also referenced above in the discussion of Van Gulick's account.

4.4 Kriegel's COI Hypothesis

Uriah Kriegel has done the most to articulate and advance an intrinsic account in recent years. Contrary to Van Gulick, Gennaro, and Carruthers, Kriegel requires that for a state to be conscious, the relevant higher-order content must be explicitly represented and actually integrated with some corresponding low-order content into a unified state by a 'psychologically real' process. This is what Kriegel calls 'cross-order information integration' (COI).²³ According to Kriegel, the content of a conscious experience is complex in that it is comprised of both low-order world-representing contents and higher-order experience-representing contents. This sort of content might be 'produced simply through the integration of information carried by what are *initially separate* mental states. When the contents of these separate mental states are appropriately integrated, a (single) mental state arises which has just the right sort of representational content' (Kriegel, 2005, p. 46, my emphasis). The result of cross-order integration is a single complex mental state that constitutes the conscious state. Without the relevant parts appropriately integrated, there simply is no conscious state.

Kriegel (2009) offers several arguments for the constraints on the general self-representational framework. For the purposes of this paper I will not rehearse nor dispute those arguments; I will assume their plausibility. Instead I will focus my criticism on the psychologically real process to which Kriegel appeals: the binding process. My main argument against Kriegel in this paper is that binding does not support his more general account. In particular, it does not rule out the possibility of higher-order misrepresentation. Therefore, Kriegel must either explain exactly why the kinds of misrepresentation cases that his account allows are not problematic (as such cases are supposed to be for extrinsic views) or he must give some other account of the psychologically real process that integrates the low-order and higher-order contents requisite for conscious states, one that would rule out or adequately explain the seemingly problematic cases of higher-order misrepresentation.

As mentioned above, for Kriegel, the relationship between the relevant low-order/higher-order pair must be psychologically real. Kriegel argues that the binding process is one plausible way to account for such a process. The binding problem in cognitive science

[23] See Kriegel (2005). This is one reason why Kriegel considers Gennaro's view to be merely notional. For Kriegel, on Gennaro's view there is no psychologically real process that binds the two states together such that we would have good theoretical reasons to consider the states as one.

is the problem of explaining how the brain binds together information that is processed in distinct areas of the brain to generate a cohesive, unified representation. For example, shapes, colours, and motion are each detected and represented in different areas of the brain. But when, say, on the tennis court someone sees a roundish, yellowish object moving toward her side of the net, she perceives this event as one cohesive event, not as several events at once. To address the binding problem, mind/brain science theorists posit some process to explain how binding works. That process is the binding process.

One widely accepted view of the binding process is found in von der Malsburg (1981). According to his view, when distinct groups of neurons in structurally distinct parts of the brain fire within milliseconds of each other, their content is bound together into a seemingly single cohesive event. According to von der Malsburg's view, synchronous firing in time represents the cohesion of the features represented as belonging together. So to explain the bound features of the approaching tennis ball, the realizers of the representations of roundish, yellowish, and motion toward, fire nearly synchronously in time, and thus are represented as being of the same object.

Kriegel acknowledges that his appeal to binding is just an hypothesis (it is just one possible neural realization of his account), but he thinks it provides some empirical basis for the cross-order integration that his account requires. Just as there seems to be a binding process that binds various environmental features of perceived objects, so too might there be a process (or an extension of the same process) that binds together representations of the environment with representations of those representations.

Here is one way that Kriegel's story might unfold. At t_1 a subject S enters into a first-order (non-conscious) state M representing red. Subsequently, at t_2 M remains active. At t_2 (or nearly synchronously at t_3) M^* is formed and activated. M^* represents M , and, since M^* was activated nearly synchronously in time with M , the two states are bound together into a unified state M^C . M^C is the conscious state.

Contrary to extrinsic theorists, according to Kriegel's view, the state that is assumed to be the conscious state is not the first-order state M , but the complex state M^C itself. In that regard, if there is no M , then there is no complex state to be the conscious state in the first place. This appears to answer the problem of targetless higher-order misrepresentation.²⁴ However the problem of mismatching low-order/higher-order states re-emerges when considering that M and M^* are

[24] Though, see Weisberg (2008) for a persuasive case that it does not.

two ‘initially separate mental states’. There is nothing about a process of binding (in terms of nearly synchronous activation) that would rule out integrating, say, a first-order circle-representing state with a higher-order state representing that state as square-representing. If that is the case, then the question ‘What is M^C like for S?’ remains unanswered. Kriegel might claim that M^C ‘is like’ whatever M^* determines, but that move seems *ad hoc*; for it becomes unclear what explanatory role the qualitative character (of M) is even contributing to the explanation of phenomenal consciousness. Perhaps Kriegel could weaken the division of explanatory labour claiming that qualitative character is determined not by the first-order representational properties as represented by the low-order component M in a conscious state, but rather, by how the low-order state is represented by the higher-order component M^* in a conscious state. This would not clear up the question of why qualitative character as represented by low-order components is even needed in the first place, though. It would also make it the case that conscious states could always have first-order (qualitative, not schmalitative) components of which the subject is completely unaware, e.g. it might be the case that a subject is in a conscious state actually first-order representing a circle, but higher-order representing it as a square. But then, my circle-representing state is, strictly speaking, conscious, or at least contributing somehow to the complex state’s being conscious, even though I am completely unaware of it.

It is true that Kriegel’s binding hypothesis is intended as one possible neural realization of his general intrinsic account and is logically independent of it. But the devil is in the details, and even if binding is logically independent of the general account, it does tell us something about how we are to understand the finer details of the account. The problem is that the binding story does not adequately support the conceptual constraints of Kriegel’s general intrinsic account. In particular, binding does not tell us enough about how higher-order misrepresentation would be ruled out, nor does it explain why it need not be ruled out.

This concludes my assessment of the existing versions of intrinsic higher-order theory. Each version has attractive elements, but none adequately addresses the problem of higher-order misrepresentation, nor explains why it need not be addressed. In the next section I introduce the quotational version of intrinsic higher-order theory that does adequately address higher-order misrepresentation. To begin the discussion I briefly review the origin and original purpose of quotational concepts in the phenomenal concepts literature.

5. The Quotational Theory of Phenomenal Consciousness

5.1 Quotational Phenomenal Concepts and the Phenomenal Concept Strategy

Quotational concepts first appear in David Papineau's (2002) *Thinking About Consciousness*.²⁵ In that book, Papineau employs quotational concepts as a way of deflecting various anti-physicalist arguments (e.g. the knowledge argument, explanatory gap arguments, and zombie/invert arguments) by attributing the apparent mysteriousness of phenomenally conscious states to the special concepts we use to think about those states, *viz. phenomenal* concepts. This strategy has come to be known as the phenomenal concept strategy. Proponents of the strategy claim that phenomenal concepts are 'conceptually isolated' from non-phenomenal concepts, e.g. ordinary theoretical (physical, functional, intentional) concepts. Given conceptual isolation, they argue, it is no surprise that, say, Mary — expert in the neuroscience of colour vision in Jackson's (1986) knowledge argument — learns something new upon stepping out of her black and white room and perceiving red, or that there are explanatory gaps (Levine, 1983; 2001), or that zombies/inverts are conceivable (Chalmers, 1996). The task, then, for proponents of the strategy is to explain conceptual isolation in a way that is consistent with physicalism, and there are different ways that philosophers have tried to do this. One such way is Papineau's quotational account of phenomenal concepts.²⁶

It should be emphasized that Papineau's project was different than mine. Papineau aimed to develop a view of concepts that would be able to address the above sorts of anti-physicalist arguments. My aim is to suggest an account of the nature of conscious states. Given the constraints of this paper, I will not defend the phenomenal concept strategy itself, nor will I defend the quotational account against other views of phenomenal concepts. Rather, I will assume the phenomenal concept strategy has promise. I will also assume that the quotational account is at least one of the plausible contenders for an account of phenomenal concepts. Of the plausible contenders no one account emerges as being clearly superior, so the role of 'best account' is up for grabs. With these assumptions in mind, I will co-opt quotational

[25] Papineau (2007) has since abandoned his quotational account in favour of a different kind of constitutive account.

[26] Other notable accounts of phenomenal concepts are Loar's (1990), Carruthers' (2000), and Tye's (2000) recognitional accounts, and Perry's (2001) and O'Dea's (2002) indexical accounts.

concepts and *mutatis mutandis* put them to work in an account of the nature of consciousness. While anti-physicalists and physicalists alike have recently argued that the strategy is doomed and should be abandoned, I, on the other hand, maintain that the strategy can be expanded.²⁷ The details of my own project will be presented below, but first, here is how Papineau initially characterized quotational concepts.

According to Papineau's original quotational account, the concepts we use to think about conscious states (phenomenal concepts) have a specific quotational structure. A phenomenal concept's structure consists of a perceptual experience prefixed by what he calls an 'experience operator', represented schematically, e.g. by 'THE EXPERIENCE <blank>', where '<blank>' is filled in by an actual occurrent experience, or an imaginative creation/recreation thereof. Construed as quotational, phenomenal concepts somehow 'embed' the experiences to which they refer, just as linguistic quotation expressions embed or inscribe within quotation marks the signs or expressions to which their quotation marks refer. There are several accounts of just how linguistic quotation works.²⁸ Without committing to any of these at this point, we can say that the linguistic quotation expression 'swan' does something like embed or inscribe within its quotation marks an instance of the sign or expression (put crudely, a token of the word) *swan*. By doing so, by displaying or exhibiting that inscription, the quotation marks directly refer to what is inscribed between them. Similarly, on Papineau's original quotational view of phenomenal concepts, when I think about, or conceptualize, my conscious visual perception of a swan, I deploy a concept of my conscious experience, which has a structure that is closely analogous to a linguistic quotation expression, e.g. THE EXPERIENCE <swan>, where '<swan>' is the swan experience itself.

On this account, phenomenal concepts are conceptually isolated in that they embed the experiences to which they refer. Because non-phenomenal concepts do not embed experiences, there will be no *a priori* links between phenomenal concepts and non-phenomenal concepts. Consequently, the sorts of phenomena emphasized in anti-physicalist arguments such as the knowledge argument are just the kind of phenomena that one would predict given the conceptual isolation of phenomenal concepts. Proponents of the strategy can then

[27] See Chalmers (2006) and Tye (2009) for arguments against the phenomenal concept strategy.

[28] See Lepore and Cappellan (2007) for an overview and a novel account.

argue that the conceptual isolation of quotational phenomenal concepts explains Mary's situation (as well as explanatory gaps, or zombie/invert conceivability), and that the anti-physicalist conclusion that physicalism is false does not follow. Thus, as evidenced by the previous discussion, Papineau's appeal to phenomenal concepts, if it is successful, provides a method of responding to the anti-physicalist, but it is most certainly not intended to provide any specific physicalist account of the nature of consciousness itself.

5.2 Putting Quotational Concepts to Work in a Theory of Consciousness

In section 4.4 I argued that binding does not rule out the relevant kind of higher-order misrepresentation, and for that reason Kriegel's account of cross-order integration in terms of binding is insufficient. However, Kriegel does make a convincing case for the claim that we need to invoke some psychologically real process to account for cross-order integration. The quotational account of consciousness supplies the requisite psychologically real process in terms of the subject's conceptualization of certain of her own perceptual states, using a specific kind of concept with a specific kind of structure. Assuming that a process of conceptualization is a psychologically real process, the right kind of conceptualization might be just what intrinsic theorists require.²⁹ In that regard the quotational account lends plausibility to the intrinsic higher-order framework.

Papineau's original account was an account of how we conceptualize states that are already conscious by some independent measure. With the idea of quotational concepts of experience in place, we can extend the conceptualization process to include perceptual states that are conscious, not by some independent measure, but in virtue of being conceptualized by quotational concepts in the first place. If one is going to acknowledge quotational concepts, at all, then we must ask on what grounds we may conclude that quotational concepts are used to think only about states that are not conscious prior to being quoted. The answer seems to be that to establish the exclusion of non-conscious experiences as possible objects of quotational concepts we would need some further argument. But there is no reason to think there would not

[29] It would be helpful to have a better grasp of to what a 'psychologically real process' must amount. While this subject has been neglected in the literature, I cannot explore it in this paper. Jehle and Kriegel (2006) do give some indication of what they have in mind by 'psychologically real'. A psychologically real process must amount to something 'actually happening' with the subject, and it must be 'temporally thick'; it cannot be merely dispositional (*ibid.*, p. 471).

also be quotational mental state concepts that embed or inscribe experiences that are not conscious prior to being embedded.

I take Papineau's account (as well as the received view of phenomenal concepts overall) to be focused on what I will call 'introspective' phenomenal concepts. Introspective phenomenal concepts target experiences that have already been rendered conscious (on the received view, by some independent process — on my view by some other quotational state). However, there is no argument given for why a concept of quotational structure must be introspective in the above-mentioned sense. It is merely assumed that Papineau's experience operator must take only conscious experiences as objects.

There are many data now that support the existence of non-conscious perceptions that seem to guide our actions in significant ways and a convincing case has been made for countenancing such states as non-conscious *experiences*.³⁰ These are experiences that possess qualitative character, but for which there is no phenomenal character. One can easily (and plausibly) envision, then, an iteration of quotational concept that operates over non-conscious experiences. This I characterize as a 'basic' phenomenal concept. The philosophically interesting claim then is this: *if* there are quotational concepts at all and there are both conscious and non-conscious experiences, then quotational concepts can operate over (take as objects) either kind of experience. The state resulting from the subject deploying a quotational concept that operates over a non-conscious experience can be characterized as having two crucial components: a higher-order (state-representing) component constituted by the mental (conceptual) 'quotation marks', and a low-order component constituted by the previously non-conscious state to which the higher-order component refers. Such a thought would 'embed' within it the very non-conscious state to which it would refer and the resulting state would be just the kind of state required by intrinsic theorists.

Quotational concepts can be schematized similarly to the way they are in Papineau's original account as having two crucial constituents. The first is what I call the conceptual sketch, or structure, which is something like 'THE STATE <blank>' or perhaps 'THIS STATE <blank>'. The second component is the 'experience', or more precisely, the content (specifically a contentful perceptual state), which is either an occurrent perceptual state or a state that is imaginatively cre-

[30] For the data on non-conscious percepts guiding action, see Milner and Goodale (1995) and Jacob and Jeannerod (2003). For a case that such states qualify as 'experiences' see Carruthers (2000).

ated/recreated from memory that fills in the '<blank>'.³¹ The structure of the concept can be innately specified and developed through maturation, or acquired through learning, but it is not something that must be acquired by the subject undergoing a specifically phenomenally conscious experience (as phenomenal concepts must, according to the received view). The content, on the other hand, must be acquired through perceptual experience of some sort (conscious, non-conscious, imagined, or actual) for the simple reason that the content of the quotational concept *just is* an 'experience'. For example, when I consciously perceive a white swan, according to the view under discussion, I undergo a first-order visual representation of the white swan that is quoted by a (quotational) thought that embeds the very first-order white-swan-representing state to which it refers. The resulting thought is a metarepresentational state that represents my white-swan-representing state as a white-swan-representing state that I am undergoing, or more simply as 'my state'.³² But it also displays or activates that very white-swan-representing contentful state.

5.3 What Makes Quotational States 'Quotational' States?

When I say a quotation embeds 'the very state' to which it refers, I mean the quotational concept actually latches onto *the very state M* itself. This can be spelled out in the following way. At t_1 the subject S undergoes a lower-order representational state M. At t_2 S undergoes the quotational state M*, which contains, not a representation of M, but M itself. Call this construal of the quotational theory 'QT'.

(QT) The first-order component of a quotational state M* is M itself, and M* is the conscious state.

Here is one characteristic that distinguishes the quotational view from other intrinsic accounts. On Van Gulick's view, cross-order integration is due to teleosemantic links and global accessibility. Gennaro leaves the question unanswered. On Carruthers' and Kriegel's views, a first-order representation M and an (initially separate) higher-order

[31] 'Experience' is a loaded term. Papineau clearly intended it to mean *conscious* experience. But if one acknowledges the reality of non-conscious experiences, then 'experience' is always ambiguous between the conscious and non-conscious types. For precision's sake, I think it best to eliminate the expression 'experience' from the schematization of quotational concepts in my account. I replace Papineau's 'experience' operator with a more generic 'state' operator.

[32] This need not require that the subject consciously think the thought 'my state', nor that the subject is cognitively sophisticated enough to articulate such a thought in inner-speech. It might well be the default assumption built into the mind-reading system that any incoming state is the subject's own state.

representation of M are bound together into a unified state with dual-content, symbolized as something like {[blue], [representing/seems blue]}. But on the quotational view there is no independently existing higher-order representation of M prior to the subject deploying a quotational thought; that is, there is no initially distinct higher-order representation of the first-order state that must then be bound together with the first-order state. Rather, while quotational states are indeed metarepresentational, the higher-order component provided by the quotational concept can be thought of as a kind of demonstrative. That is, not a demonstrative that merely points to some state or other, but rather, one that points *and* presents the very state to which it points. In that regard, the low-order content has a dual role: its information encodes some aspect of the world, but also, that very same token is represented as representing the world in virtue of being enveloped by a quotational state.

According to both Papineau's original view and the account that I am laying out, quotational concepts 'embed' or 'contain' or 'inscribe' or 'latch onto' or 'envelope' the states to which they refer. What could it mean for one state to embed (contain, inscribe, latch onto, or envelope) another? Are quotational concepts a *deus ex machina* or a mere metaphor? How could a mental state or a concept have the kind of nested structure that quotational states are purported to have?

First, even if quotational concepts are metaphorical, metaphor has an important explanatory role in naturalistic explanation. Mere (perhaps 'literary') metaphors are not intended to motivate a research programme. Metaphors in natural science are (Boyd, 1979, p. 489). Though they may differ on exactly how metaphors function, both scientific realists and anti-realists think that metaphors have an important role in scientific explanation, and some metaphors are more apt than others.³³

More importantly, though, quotational concepts are in fact not merely metaphorical. The purported 'nesting' of states is metaphorical in that there might not be spatio-temporal nesting of brain states. However, as long as the quotation *function*, which is independent of how that function is conveyed symbolically or realized neurologically, is sufficiently robust to make both the concept and its content constitutive of a single representation, one can appreciate the aptness of the metaphor (both of quotational thought and linguistic quotation) that one representation (and its semantic value) is 'contained within' the other.

[33] See Boyd (1979) and Black (1962).

Consider: even the symbol used to indicate the function of linguistic quotation need not have involved literal nesting or embedding in the first place. We could have symbolized the function of quotation by the use of ‘underscore’ as in ‘swan’ or by preceding each ‘quoted’ term with a ‘q’ as in ‘qthe qwhite qswan’. Neither case involves literal nesting, but both retain what matters most about quotation: the necessary, or constitutive, relation between what quotes and what is quoted. In both cases what is quoted is presented in virtue of being quoted, unlike a mere indexical that might point but leave its actual referent unknown (imagine the case of the blindfolded person who thinks ‘I don’t know where here is’).³⁴ Thus, it is explanatorily useful to claim that the ‘underscore’ or ‘q’ expressions ‘contain’ their semantic values in the sense that they present them necessarily, even if the containment is *somewhat* metaphorical.

The crucial point is this. Quotational thoughts need not actually involve nested states. What matters is that the link between the quotational thought deployed and its target representational content be sufficiently robust (psychologically real), such that the target state is a constituent of the quotational representation (and is itself presented as a state the subject is in). I have suggested that one useful way to think about quotational concepts is in terms of demonstratives that necessarily display their referents.

5.4 The Quotational Theory, Higher-Order Misrepresentation, and the Phenomenal Concept Strategy

So how does the quotational view actually handle problematic cases of higher-order representation? Recall that (QT) says that the higher-order state contains within it, or latches onto, the very same token first-order state that it comes to represent. Given (QT), the structure of the quotational concepts deployed during the conceptualization of a given perceptual state make it such that higher-order misrepresentation is not possible. That is, higher-order misrepresentation is not just theoretically possible but rare, as it is on extrinsic views. Rather, it is theoretically impossible for the higher-order component to diverge from its first-order target, because the higher-order representation partly consists of that very first-order target.

Moreover, on the quotational view, consciousness is not merely notional, as it is on Gennaro’s wide-intrinsicality view. Nor must one

[34] Lepore and Cappellen (2007) capture this idea of containment/proximity with their Strong Disquotational Schema: only ‘e’ quotes ‘e’. Furthermore, ‘e’ could not have quoted anything but ‘e’ (*ibid.*, p. 151). According to Lepore and Cappellen, quotation is not a pragmatic feature of language, it is a *sui generis* mode of referring (*ibid.*, p. 5).

be committed to the view that consciousness, a seemingly categorical property, can be grounded in a dispositional property such as being available to downstream consumer systems *a la* Carruthers and Van Gulick. Nor does it reintroduce higher-order misrepresentation by appeal to a binding process of cross-order information integration, as Kriegel claims. Thus, regarding higher-order misrepresentation, the quotational view has significant advantages over both extrinsic and competing intrinsic views.

One question that naturally arises is whether the quotational concepts at work in the quotational theory of consciousness really are quotational *phenomenal* concepts. In other words, one might accept the quotational account of consciousness but deny that the quotational concepts that partly constitute conscious states are phenomenal concepts. This issue is largely terminological, for there is only one main substantive reason why the quotational concepts employed by the quotational theory ought not to be considered ‘phenomenal’, and that is if by employing such concepts in a theory of the nature of consciousness, the phenomenal concept strategy itself is undermined. As stated in the opening passages, the phenomenal concept strategy is intended to defend physicalism against various anti-physicalist arguments, and the core insight of the strategy is thought to be that it distances the nature of consciousness itself from the concepts we use to think about conscious states.³⁵ According to the view that I have introduced, the conceptual distancing between the explanation of consciousness itself and the apparent mystery of consciousness seems to have been shortened: quotational concepts are partly constitutive of phenomenally conscious states themselves. Thus, one might argue that I have reintroduced the mystery that the strategy intended to explain away. It needs to be shown, then, that the quotational account of consciousness retains whatever explanatory power the phenomenal concept strategy has, else we would have to give an alternate defence of physicalism.

Fortunately (if the strategy has promise at all), no such alternate defence of physicalism is required; nothing about the quotational theory of consciousness itself undermines whatever explanatory power the strategy has. Consider one example: the quotational theory of consciousness retains a physicalist explanation of Jackson’s Mary scenario. Assuming that Mary has progressed through a normal process of maturation (with the exception of living in a solely black and white world), even before she exits her room she would be able to think (and

[35] See Balog (2009).

would have undergone) other quotational thoughts. That is, she would have undergone other phenomenally conscious states, e.g. states representing black and states representing white. What she gains when she steps outside the room is not the quotational structure itself, but rather, a specific experience of red that can then immediately be integrated into that structure, and thus, rendered conscious. And once Mary has undergone the *conscious* experience of red, she can subsequently quote that conscious state, deploying an *introspective* iteration of a quotational thought, whereby she might think something like ‘Ahhhh. That is what it’s like to (consciously) experience red’. Regarding Mary’s situation, nothing about the original alleged explanatory power of the phenomenal concept strategy has been lost. Furthermore, there is no reason to think the explanatory power of the strategy would be lost regarding explanatory gap arguments or zombie/invert intuitions. Thus, if the phenomenal concept strategy is successful, it remains so even under my proposed revision.³⁶

Moreover, it is not merely that the phenomenal concept strategy itself has nothing to lose from endorsing the quotational account of consciousness; it has something to gain as well. First, unifying an account of the way we think about conscious states with an account of what constitutes state consciousness at all renders the phenomenal concept strategy less *ad hoc*. On the quotational account, it is not just that we have unique concepts that we use only to think about states that are already conscious by some independent process (why would we need unique concepts just for that?), but more importantly, we have those unique concepts *and they partly constitute conscious states* in the first place. The uniqueness of consciousness itself calls out for unique concepts in a way that thinking about states that are conscious by some independent process does not. Also, it offers an explanation for exactly why it is that conscious states *seem* mysterious. Similarly, they seem mysterious not simply because we use unique concepts to think about them, but rather because the unique concepts we use to think about them partly constitute what makes those states conscious in the first place. That is, conscious states themselves are at least *prima facie* mysterious, however, on the quotational view that mysteriousness is explained rather than merely explained away.

[36] Again, the issue here is not whether the strategy actually has any promise. The issue is whether or not my revision undermines whatever explanatory power the strategy might be thought to have. Whether or not the quotational view can retain more sophisticated physicalist responses to anti-physicalist objections (e.g. Chalmers, 2006, and Carruthers and Veillet, 2007) is a reasonable concern. However, that concern is one that must be addressed elsewhere.

6. Conclusion

In sum, if higher-order misrepresentation is a genuine problem, it is a genuine problem for both traditional extrinsic higher-order theories and for existing intrinsic theories. The intrinsic theories appearing hitherto in the literature either succeed in ruling out higher-order misrepresentation but fail to provide a satisfying explanation of consciousness in the first place, inherit commitments that further complicate the view, or reintroduce the problem. The quotational account avoids these consequences. It adequately addresses higher-order misrepresentation while maintaining whatever explanatory power the phenomenal concept strategy is presumed to have. According to the received view of the strategy, phenomenal concepts are thought to be ‘phenomenal’ concepts because they refer to phenomenal states. On the other hand, according to the quotational view of consciousness, phenomenal concepts are taken to be ‘phenomenal’ concepts because they partly constitute phenomenal states. There is, though, nothing about the phenomenal concept strategy that is undermined by the quotational account of consciousness. Rather, while this paper is not intended as an explicit defence of the phenomenal concept strategy, the quotational account has the ancillary advantage of providing the strategy with additional support.³⁷

References

- Armstrong, D. (1968) *A Materialist Theory of Mind*, New York: Humanities Press.
 Armstrong, D. (1981) *The Nature of Mind and Other Essays*, Ithaca, NY: Cornell University Ithaca Press.
 Baars, B. (1988) *A Cognitive Theory of Consciousness*, Cambridge: Cambridge University Press.
 Balog, K. (2009) Phenomenal concepts, in McLaughlin, B., Beckermann, A. & Walter, S. (eds.) *Oxford Handbook in the Philosophy of Mind*, Oxford: Oxford University Press.
 Black, M. (1962) *Models and Metaphors*, Ithaca, NY: Cornell University Press.
 Block, N. (1995) On a confusion about the function of consciousness, *Behavioral and Brain Sciences*, **18**. Reprinted in Block, N.J., Flanagan, O. & Guzeldere, G. (eds.) (1997) *The Nature of Consciousness: Philosophical Debates*, Cambridge, MA: MIT Press.
 Botterill, G. & Carruthers, P. (1999) *The Philosophy of Psychology*, Cambridge: Cambridge University Press.

[37] Many thanks to audiences at the 2010 Pacific APA, the 2010 Southern Society for Philosophy and Psychology, and especially to Elizabeth Picciuto, Peter Carruthers, Josh Weisberg, Mike Bruno, and an anonymous referee for insightful comments on earlier drafts.

- Boyd, R. (1979) Metaphor and theory change: what is 'metaphor' a metaphor for?, in Ortony, A. (ed.) *Metaphors and Thought*, Cambridge: Cambridge University Press.
- Brentano, F. (1874/1995) *Psychology from an Empirical Standpoint*, Rancurello, C.A., Terrell, D.B. & McAlister, L.L. (trans.), London: Routledge.
- Byrne, A. (1997) Some like it HOT: Consciousness and higher-order thoughts, *Philosophical Studies*, **86**, pp. 103–129.
- Byrne, A. (2004) What phenomenal consciousness is like, in Gennaro, R.J. (ed.) *Higher-Order Theories of Consciousness*, pp. 19–44, Amsterdam: John Benjamins.
- Carruthers, P. (2000) *Phenomenal Consciousness: A Naturalistic Theory*, Cambridge: Cambridge University Press.
- Carruthers, P. (2005) *Consciousness: Essays from a Higher-Order Perspective*, Oxford: Oxford University Press.
- Carruthers, P. (2009) Higher-Order Theories of consciousness, in Zalta, E.N. (ed.) *The Stanford Encyclopedia of Philosophy (Fall 2009 Edition)*, [Online], <http://plato.stanford.edu/archives/fall2009/entries/consciousness-higher/>
- Carruthers, P. & Veillet, B. (2007) The phenomenal concept strategy, *Journal of Consciousness Studies*, **14** (9–10), pp. 212–236.
- Caston, V. (2002) Aristotle on consciousness, *Mind*, **CXI**, pp. 751–815.
- Chalmers, D. (1996) *The Conscious Mind*, Oxford: Oxford University Press.
- Chalmers, D. (2006) Phenomenal concepts and the explanatory gap, in Alter, T. & Walter, S. (eds.) *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*, Oxford: Oxford University Press.
- Church, J. (1998) Two sorts of consciousness?, *Communication and Cognition*, **XXXI**, pp. 57–72.
- Dennett, D. (1991) *Consciousness Explained*, Boston, MA: Little, Brown and Co.
- Dennett, D. (1992) The self as the center of narrative gravity, in Kessel, F.S., Cole, P.M. & Johnson, D.L. (eds.) *Self and Consciousness: Multiple Perspectives*, Hillsdale, NJ: Lawrence Erlbaum Associates.
- Dennett, D. & Kinsbourne, M. (1995a) Time and the observer: The where and when of consciousness in the brain, *Behavioral and Brain Sciences*, **XV** (2), pp. 183–247.
- Dennett, D. & Kinsbourne, M. (1995b) Escape from the Cartesian Theater, *Behavioral and Brain Sciences*, **XV**, pp. 234–246.
- Dennett, D. (2001) Are we explaining consciousness yet?, *Cognition*, **LXXIX**, pp. 221–237.
- Garcia, J.A. & Koelling, R.A. (1967) A comparison of aversions induced by X rays, toxins, and drugs in the rat, *Radiation Research Supplement*, **VII**, pp. 439–450.
- Gennaro, R. (1996) *Consciousness and Self-Consciousness*, Amsterdam: John Benjamins.
- Gennaro, R. (2004) *Higher-Order Theories of Consciousness: An Anthology*, Amsterdam: John Benjamins.
- Gennaro, R. (2006) Between pure self-referentialism and the extrinsic HOT theory of consciousness, in Kriegel, U. & Williford, K. (eds.) *Self-Representational Approaches to Consciousness*, Cambridge, MA: MIT Press.
- Husserl, E. (1928/1964) *Phenomenology of Internal Time-Consciousness*, Heidegger, M. (ed.), Churchill, J.S. (trans.), Bloomington, IN: Indiana University Press.
- Jackson, F. (1986) What Mary didn't know, *Journal of Philosophy*, **LXXXIII**, pp. 291–295.

- Jehle, D. & Kriegel, U. (2004) An argument against dispositionalist HOT theory, *Philosophical Psychology*, **XIX**, pp. 463–476.
- Kriegel, U. (2005) Naturalizing subjective character, *Philosophy and Phenomenological Research*, **LXXI**, pp. 23–57.
- Kriegel, U. (2006) The Same-Order Monitoring Theory of consciousness, in Kriegel, U. & Williford, K. (eds.) *Self-Representational Approaches to Consciousness*, Cambridge, MA: MIT Press.
- Kriegel, U. (2009) *Subjective Consciousness*, Oxford: Oxford University Press.
- Lepore, E. & Cappellan, H. (2007) *Language Turned on Itself: The Semantics and Pragmatics of Metalinguistic Discourse*, Oxford: Oxford University Press.
- Levine, J. (1983) Materialism and qualia: The explanatory gap, *Pacific Philosophical Quarterly*, **LXIV**, pp. 354–361.
- Levine, J. (2001) *Purple Haze*, Oxford: Oxford University Press.
- Loar, B. (1990) Phenomenal states, *Philosophical Perspectives*, **IV**, pp. 81–108.
- Lycan, W. (1996) *Consciousness and Experience*, Cambridge, MA: MIT Press.
- Lycan, W. (2001) A simple argument for a higher-order representational theory of consciousness, *Analysis*, **LXI**, pp. 3–4.
- Neander, K. (1997) The division of phenomenal labor, in Tomberlin, J. (ed.) *Language, Mind, and Ontology, XII of Philosophical Perspectives*, pp. 435–457, Oxford: Blackwell.
- O’Dea, J. (2002) The indexical nature of sensory concepts, *Philosophical Papers*, **31**, pp. 169–181.
- Papineau, D. (2002) *Thinking About Consciousness*, Oxford: Oxford University Press.
- Papineau, D. (2007) Phenomenal concepts are not demonstrative, in McCabe, M.M. & Textor, M. (eds.) *Perspectives on Perception, Philosophical Research*, **VI**, Frankfurt: Ontos Verlag.
- Perry, J. (2001) *Knowledge, Possibility, and Consciousness*, Cambridge, MA: MIT Press.
- Rosenthal, D. (1997) A theory of consciousness, in Block, N., Flanagan, O. & Güzeldere, G. (eds.) *The Nature of Consciousness: Philosophical Debates*, pp. 729–753, Cambridge, MA: MIT Press.
- Rosenthal, D. (2004) Varieties of Higher-Order Theory, in Gennaro, R.J. (ed.) *Higher-Order Theories of Consciousness*, pp. 19–44, Amsterdam: John Benjamins.
- Rosenthal, D. (2005) *Consciousness and Mind*, Oxford: Oxford University Press.
- Simons, P. (1987) *Parts: A Study in Ontology*, Oxford: Oxford University Press.
- Stoljar, D. (2005) Physicalism and phenomenal concepts, *Mind and Language*, **XX**, pp. 469–494.
- Tye, M. (1995) *Ten Problems of Consciousness*, Cambridge, MA: MIT Press.
- Tye, M. (2000) *Color, Consciousness and Content*, Cambridge, MA: MIT Press.
- Tye, M. (2009) *Consciousness Revisited: Materialism Without Phenomenal Consciousness*, Cambridge, MA: MIT Press.
- Van Gulick, R. (1988) A functionalist plea for self-consciousness, *Philosophical Review*, **XCVII**, pp. 149–188.
- Van Gulick, R. (2004) Higher-Order Global States (HOGS): An alternative higher-order model of consciousness, in Gennaro, R. (ed.) *Higher-Order Theories of Consciousness: An Anthology*, Amsterdam: John Benjamins.
- Van Gulick, R. (2006) Mirror mirror — is that all?, in Kriegel, U. & Williford, K. (eds.) *Self-Representational Approaches to Consciousness*, Cambridge, MA: MIT Press.

Weisberg, J. (2008) Same old, same old: The Same-Order Representation Theory of consciousness and the division of phenomenal labor, *Synthese*, **CLX**, pp. 161–181.

Paper received July 2009; revised June 2010.