

# Phenomenal Concepts and Higher-Order Experiences

PETER CARRUTHERS

*University of Maryland, College Park*

Relying on a range of now-familiar thought-experiments, it has seemed to many philosophers that phenomenal consciousness is beyond the scope of reductive explanation. (Phenomenal consciousness is a form of state-consciousness, which contrasts with *creature-consciousness*, or *perceptual-consciousness*. The different forms of state-consciousness include various kinds of *access-consciousness*, both first-order and higher-order—see Rosenthal, 1986; Block, 1995; Lycan, 1996; Carruthers, 2000. Phenomenal consciousness is the property that mental states have when it is *like something* to possess them, or when they have subjectively-accessible *feels*; or as some would say, when they have *qualia* (see fn.1 below).) Others have thought that we can undermine the credibility of those thought-experiments by allowing that we possess purely recognitional concepts for the properties of our conscious mental states. This paper is concerned to explain, and then to meet, the challenge of showing how purely recognitional concepts are possible if there are no such things as *qualia*—in the strong sense of intrinsic (non-relational, non-intentional) properties of experience. It argues that an appeal to higher-order experiences is necessary to meet this challenge, and then deploys a novel form of higher-order thought theory to explain how such experiences are generated.

## 1 Introduction: thought-experiments and qualia<sup>1</sup>

There is now an array of familiar philosophical thought-experiments that are supposed to establish, not just that there are qualia that can vary independently of functional and intentional properties, but that qualia are non-physical in nature (Kripke, 1972; Jackson, 1982, 1986; Chalmers, 1996, 1999; Siewert, 1998). For example, I can think, ‘*This* type of experience [pain] might have occurred in me, or might occur in others, in the absence of any of the usual causes and effects of pains. There could be someone in whom *these* experiences occur but who isn’t bothered by them, and where those experi-

---

<sup>1</sup> Some philosophers use the term ‘qualia’ in a weaker, theoretically-neutral, sense than I do here, to mean whatever it is that renders a mental state phenomenally conscious (e.g. Lycan, 1996; Sturgeon, 2000). Taken in this sense, to deny the existence of qualia would be to deny the very existence of phenomenal consciousness. But in the strong sense that I propose to adopt throughout this paper, one can be a qualia irrealist—denying that experiences possess any mental properties that are intrinsic and non-intentional—while continuing to insist that some of our experiences possess subjectively accessible feels.

ences are never caused by tissue damage or other forms of bodily insult. And conversely, there could be someone who behaves and acts just as I do when in pain, and in response to the same physical causes, but who is never subject to *this* type of experience.'

Equally, I can think, '*This* type of experience [as of red] might have occurred in me, or might normally occur in others, in the absence of any of its actual causes and effects. So on any view of intentional content that sees content as tied to normal causes (i.e. to information carried) and/or to normal effects (i.e. to teleological or inferential role), *this* type of experience might occur without representing *red*.'

Even more radically, indeed, it seems that I can think, '*This* type of state [an experience] might not have been, or might not be in others, an *experience* at all. Rather it might have been / might be in others a state of some quite different sort, occupying a different position within the causal architecture of cognition.' Or I can think, 'There might have been a complete physical / causal / intentional duplicate of me who failed to undergo *this* experience, or indeed any experience at all.'

What do these thought-experiments really establish? One popular response has been to claim that they show only that there is something distinctive about the way in which we conceptualize our experiences, not anything about the nature of those experiences themselves (Loar, 1990, 1997; Papineau, 1993, 2002; Sturgeon, 1994, 2000; Tye, 1995, 2000; Carruthers, 2000). Such thought-experiments only show, it is said, that we have some concepts of experience that are purely recognitional, in the sense of having no conceptual ties with physical concepts, or with concepts of causal role and/or concepts of intentional content. (Some use the term 'phenomenal concepts' in this regard.) The consensus amongst the authors listed above is that metaphysical claims about the nature of phenomenal properties can't be established by means of thought-experiments that turn crucially on the existence of purely recognitional concepts of experience. For we might well possess concepts of this type even if phenomenal properties are actually physical properties, or causal-role properties, or intentional properties of one sort or another.

It is true that not *every* philosopher who takes a physicalist / reductive attitude towards phenomenal consciousness is prepared to allow the existence of purely recognitional concepts of experience. Thus Dretske (1995), for example, thinks that we can only really refer to our own experiences indirectly, via the properties (*redness*, *roundness*, or whatever) that those experiences are experiences of. So when I think, '*This* type of experience [as of red] might have been F', I am really thinking, 'My experience *of this* [red] might have been F'. Dretske's difficulty, however, is then to provide a satisfactory

explanation of how it is that I seem to be capable of thinking, ‘*This* type of experience [as of red] might not have been an experience of *this* [red]’.

I shall not pursue the point here. I propose simply to assume, for the moment, that there are purely recognitional concepts of experience (returning to the issue briefly in section 4). I shall assume, that is, that it is possible to form a concept of a type of experience that consists in nothing more and nothing less than the capacity to recognize that type of experience when it occurs in one’s own mental life. Such concepts will have no conceptual connections with any of our physical, causal-role or intentional-content concepts—not even with the concept *experience*, if this is functionally specified by the characteristic place of experiences in the causal architecture of cognition. Our task will be to see how the existence and features of such purely-recognitional concepts are best explained if there are no qualia.<sup>2</sup>

This paper will be concerned to confront an important challenge. This is to explain how purely recognitional concepts are even so much as *possible* if qualia (in the strong sense) *don’t* exist. For as we shall see, it can be difficult to understand how purely recognitional concepts of experience—of the sort that we manifestly seem capable of possessing—are possible unless qualia are the properties being recognized. This challenge will form the topic of the remainder of this paper. Sections 2 and 4 will outline and develop it, while sections 3 and 5 will show how the challenge can best be met. I shall be arguing, first, that purely recognitional concepts of experience need to be grounded in higher-order experiences of our (first-order) perceptual states; and second, that the most plausible version of higher-order experience theory is not the ‘inner sense’ theory of Armstrong (1968, 1984) and Lycan (1996), but rather one that can be derived from dispositional higher-order thought theory, of the sort defended by Carruthers (2000).

I should emphasize at the outset, however, that although the higher-order theories in question (whether of ‘inner sense’ or of higher-order thought) are normally intended and presented as reductive explanations of phenomenal consciousness, this is not their role in the present context. Our task is to explain how purely recognitional concepts of experience are possible without invoking qualia (thus blocking some of the main arguments *against* the reductive

---

<sup>2</sup> I shall therefore set to one side the various direct attacks on purely recognitional concepts of experience that have been offered by such disparate authors as Wittgenstein (1953) and Fodor (1998). My view is that no version of Wittgenstein’s famous argument against private concepts can succeed without drawing on anti-realist assumptions about the nature of the mind, or the nature of concepts, or both. And my view is that Fodor’s argument—that recognitional concepts cannot be actual because such concepts *don’t compose*—makes a false assumption. This is, that if recognitional concepts are possible at all, then it is possible for concepts composed out of recognitional concepts to be recognitional also. But I shall not attempt to substantiate these points here. My interest is in the *conditional* question: supposing that there can be purely recognitional concepts of experience, what then follows?

explicability of consciousness), not to propose any particular reductive explanation. Someone who rejects higher-order thought theory as a reductive account of phenomenal consciousness, as such, might still be able to accept the present proposals for explaining the existence of purely recognitional concepts of experience.<sup>3</sup>

I should also emphasize that the dispute between qualia-theorists and their opponents cuts across the debate between non-physicalists and physicalists about phenomenal consciousness. For there are those who believe in intrinsic qualia who are nevertheless physicalists about such properties (McGinn, 1991; Block, 1995). I shall have nothing to say to such people here. While I believe it would be a bad thing for aspiring physicalists if they were required to believe in qualia, I shall not attempt to substantiate this claim in the present context. My focus is on those who either reject or don't want to be committed to qualia in the strong sense, but who nevertheless wish to make appeal to purely recognitional concepts in blocking the arguments against physicalism. My targets therefore include Sturgeon (2000), Tye (2000) and Papineau (2002); indeed, they include any physicalist who thinks that we can hope for more than bare identities between phenomenally conscious states and physical states, and who wishes to propose a reductive story in terms of some combination of causal roles and/or intentional contents.

In fact the main goal of the paper is to argue for the following conditional claim: *If there are purely recognitional concepts of experience (with the properties that we believe such concepts to have), and there are no intrinsic qualia, then there are higher-order experiences that serve to ground the application of those concepts.* A subsidiary goal is to contrast two accounts of higher-order experience. One is inner-sense theory, which is briefly criticized in section 3. And the other is that provided by dispositional higher-order thought theory, sketched in section 5. Many questions about the latter account remain, of course; and no attempt is made here at a full defense. My goal is to say just enough to indicate how higher-order experiences may be possible without inner sense (see Carruthers, 2000, for further development and discussion).

---

<sup>3</sup> I shall return to this point once again briefly at the end of the paper. Note, however, that one way in which the issue of reductive explanation impacts upon our present topic is this. It seems unlikely that anyone would want to endorse inner-sense theory who did *not* think that it provided a successful reductive explanation of phenomenal consciousness. For there would seem to be no *other* motives for believing in an organ of inner sense. In contrast, since dispositionalist higher-order thought theory doesn't need to appeal to anything that most people don't believe in anyway, many will be able to accept the present account of purely recognitional concepts who don't accept higher-order thought theories as reductive accounts of phenomenal consciousness as such.

## 2 How are purely recognitional concepts of experience possible?

A concept is recognitional when it can be applied on the basis of perceptual or quasi-perceptual acquaintance with its instances. And a concept is *purely* recognitional when its possession-conditions (in the sense of Peacocke, 1992) make no appeal to anything other than such acquaintance. A concept is purely recognitional when nothing in the grasp of that concept, as such, requires its user to apply or appeal to any other concept or belief. A purely recognitional concept *of experience* is then a higher-order recognitional concept, which applies to another mental state (*viz.* an experience), and whose possession-conditions don't presuppose any other mental-state concepts (not even the concept *experience*).

Now, in one sense it is relatively easy to understand how we might possess purely recognitional concepts of experience, even in the absence of any qualia. Suppose that experiences are events that fall into a variety of distinct kinds, whether physical, functional, or intentional. Then it is easy enough to imagine that there could be a causal mechanism that would generate, from the presence of one of these states of kind *K*, the judgment that one is in *K*.<sup>4</sup> Since the mechanism is a causal one, it might qualify as a kind of quasi-perceptual acquaintance. But since it is also *brute-causal*—in the sense that its operation is independent of any of the subject's other mental states—it can also count as purely recognitional. (I should stress that by a 'brute-causal' account of purely recognitional concepts, I don't mean a causal/ informational account of the *content* of those concepts, as opposed to a teleosemantic or inferential-role one. I just mean an account of the application-conditions of those concepts that doesn't make appeal to any other mental states besides the one the concept is applied to.) So where's the problem? Why should the existence of purely recognitional concepts of experience put any pressure on us to allow the existence of intrinsic qualia?

One source of difficulty with the above proposal is this. Although there may be no conceptual connection between recognitional concepts of types of experience and related functional or intentionally characterized concepts, it seems that there *are* such connections with other purely-recognitional concepts. For example, subjects can know *a priori* that the state that they recognize when deploying a particular recognitional concept *is* an experiential state, provided that the latter, too, is picked out by a recognitional concept (albeit a more abstract one). Possessing a generalized recognitional concept of *experience*, and possessing the purely recognitional concept *this type of state* [a particular kind of experience], subjects can know, as soon as they

---

<sup>4</sup> See Papineau (1993) where a model of our capacity to recognize our own experiences of just this kind is presented and defended. I shall return to consider alternative possible models, including Papineau's most recent views, in section 4 below.

reflect, that the items picked out by *this* are actually experiences. But of course, the mere fact that one concept tracks instances of kind *K* (or has as its function to track instances of that kind, or whatever), while another tracks instances of kind *E* (where the extension of *K* is included in *E*) wouldn't enable a thinker to know *a priori* that all *K*s are *E*s.

Qualia-theorists, in contrast, can easily explain these *a priori* connections amongst our recognitional concepts of experience. They can claim that we have available a generalized concept of *experience* that is the concept of a state that just *feels* a distinctive way to the subject, being available for immediate introspective recognition. For a qualia-theorist, the concept of experience is primarily that of a state possessing certain kinds of introspectible properties. On this account 'experience' stands to '*this state*' [a quale] just as 'color' stands to 'red'—perhaps it is by first possessing the capacity to recognize qualia of various kinds, and abstracting, that one gets the generalized concept of experience. So, possessing this sort of concept of experience, and also being capable of recognizing *this state* [quale], subjects can of course tell that what they have just recognized is an experience (*viz.* a state possessing one of a range of distinctive inner feels).<sup>5</sup>

In contrast, if we deny the existence of qualia, then the story is much less easy to tell. One option would be to claim that the generalized concept of experience is functional-intentional. But although one's functionalist concept of experience may include the fact that experiences are apt to issue in, or to be available to, purely recognitional judgments, in any particular case where one makes such a judgment one will only be able to tell that it is an *experience* that one has recognized as a result of a meta-conceptual inference. That is, only if one knows *that* the concept that one has just employed is a purely recognitional one, will one be able to know that the item recognized is an experience. This looks highly counter-intuitive.

So, one who denies the existence of qualia must somehow claim that the concepts *this* [type of experience] and *that* [experience in general] are both purely-recognitional, while at the same time allowing for the idea that the subject can discern the relationship between their instances *a priori* in something like the way that one can discern the relationship between instances of *red* and of *color*. Qualia theorists can claim that qualia are directly *present to* the concept-wielding mind, being available to purely-recognitional classifica-

---

<sup>5</sup> I should emphasize that the appeal to a process of *abstraction*, here, is for purposes of illustration only. The real point concerns, not the genesis of our concepts, but rather our awareness of that which grounds their application. Possessing a recognitional concept of *this* [type of experience], and possessing a recognitional concept of *that* [type of state, namely experience in general], I can see by mere reflection that anything of *this* type is of *that* type, in something like the way that someone possessing a recognitional concept of *red* and a recognitional concept of *color* can see by reflection that anything red is colored.

tion while also being objects of immediate awareness. If qualia are rejected, in contrast, then some account has to be given of the acquaintance-relation that underpins purely-recognitional applications of experience-concepts. And the only candidate on the table at the moment, is a brute-causal account (others will be considered in section 4). But this doesn't have the resources to explain the character of our awareness that the grounds for applying the recognitional concept *this* [type of experience] are included amongst the grounds for applying the recognitional concept *that* [experience in general].

Another problem for the brute-causal account being mooted here, is that there seems to be a particularly intimate connection between the content of the recognitional judgment, '*This* [experience] is a *K*', and the specific nature and/or content of the state that grounds that judgment. What I recognize when I deploy a recognitional concept of experience is in some sense presented to me (albeit non-conceptually) *as* an experience. I do not merely find myself judging '*This* is a *K*', as it were blindly, or for no reason. Rather, I think that I am aware of, and can inspect and reflect on the nature of, the event that evokes that recognitional judgment.<sup>6</sup>

How is this possible? Again, qualia-theorists can provide an answer—it is because the property that my recognitional concept picks out is both intrinsic and directly present to the concept-wielding mind. Qualia are supposed to be properties that we are aware of, and that we can come to have immediate recognitional capacities for by virtue of that awareness. In contrast, it is much less obvious what a defender of the brute-causal account can say here. For if the property, whose instantiation causes an application of the recognitional concept *K*, is a physical, or functional, or intentional one, then it is far from clear how such properties could figure in the right way in the content of awareness. Indeed, given that the connection between the concept and its instances is supposed to be *brute-causal*, it seems plain that the account does *not* have the resources to capture the relevant mode of presentation of those instances.

So what is it that I am recognizing when I apply a recognitional concept of experience, if not a quale? How can what I am recognizing be presented to me *as* an experience, given that it doesn't have to involve any conceptualization of it as such, unless what I recognize possesses the distinctive and (on this view) defining properties of phenomenally-conscious experience (i.e. qualia)?

---

<sup>6</sup> Another way of putting this point is that the brute-causal account cannot adequately capture the distinction between the *sense* of a recognitional concept (or its *mode of presentation* of its instances), and its referent. When I apply a recognitional concept to my experience, that experience seems to be presented to me in a distinctive way, and it is this mode of presentation that grounds my application of the relevant concept.

### 3 HOEs to the rescue?

Inner-sense theorists, who believe that we are subject to higher-order experiences (HOEs), have answers to these problems (Armstrong, 1968, 1984; Lycan, 1996). On this view, humans not only have first-order non-conceptual and/or analog perceptions of states of their environments and bodies, they also have second-order non-conceptual and/or analog perceptions of their first-order states of perception. Humans (and perhaps other animals) not only have sense-organs that scan the environment / body to produce fine-grained representations which can then serve to ground thoughts and actions, but they also have *inner* senses, charged with scanning the outputs of the first-order senses (i.e. experiences) to produce equally fine-grained, but higher-order, representations of those outputs (i.e. to produce higher-order experiences).<sup>7</sup>

Now, there are important issues here concerning the proper characterization of the contents of perception. In particular, should we say that those contents are *non-conceptual* (as Tye, 1995, argues), or should we merely say that they are *analog* (that is, being more fine-grained than any concepts we possess; which is the view defended in Carruthers, 2000)? The former entails the latter, but the latter need not entail the former. For there remains the possibility that perceptual contents might be *both* fine-grained *and* imbued with concepts. Important as these issues are, they need not detain us here. Since experiences are analog (or 'fine-grained') on either view, I propose to adopt this way of talking. And in what follows I shall adopt the convention of marking terms referring to perceptual contents with a sub-scripted 'a' for analog.

An experience *as of* red, say, is a state with the first-order analog content  $red_a$ . A higher-order experience targeted on that very state, will be one with the second-order analog content *seems  $red_a$*  or *experience of  $red_a$* . Such a higher-order experience can then serve to ground a higher-order recognitional concept. This can either be a recognitional application of the theoretically embedded concept *experience of red*, or it can be a concept that is purely recognitional. A purely-recognitional concept of experience that is deployed in the presence of, and guided in its application by, a second-order analog content will be a recognition *of* a state of *experience of  $red_a$* , but without conceptualizing it *as* what it is—an experience of red.

A higher-order experience is just that—an experience whose non-conceptual / analog content represents the non-conceptual / analog content of a first-order experience. A higher-order experience of an experience of red will be a state with the analog content *experience of  $red_a$* . An application of a higher-

---

<sup>7</sup> Since some people hear the term 'experience' in a sense that entails phenomenal consciousness, I should emphasize that this is not the intention here. No higher-order experience theorist believes that higher-order experiences are themselves (normally) phenomenally conscious. Those who have trouble with this terminology should substitute 'higher-order perceptual state' throughout.



order recognitional concept that is driven by the content of this higher-order experience will therefore have at least a non-accidental connection with the experiential status of what is recognized, in the same sort of way that recognitional applications of the concept *red* that are driven by the content of a first-order state with the analog content *red<sub>a</sub>* have an intrinsic connection with the redness of what is recognized.

Consider someone who has experiences with the analog content *red<sub>a</sub>*, and who also possesses the recognitional concept *red*, where the latter is guided in its application by the former. Their application of the concept *red* will not be brute-causal or 'blind', but will rather be guided by their awareness of the redness recognized. And if they also have experiences with the contents *green<sub>a</sub>*, *blue<sub>a</sub>*, and so on, they may also possess a more generalized recognitional concept *color*. (And given that they *also* have experiences with the contents *smooth<sub>a</sub>*, *loud<sub>a</sub>*, *sour<sub>a</sub>*, and so on, they may even have a generalized recognitional concept of *perceptible property*.)

So, too, then, for someone who has higher-order experiences with the analog content *experience of red<sub>a</sub>*, and who also possesses a purely-recognitional concept *this* [experience of red]—their application of the concept *this* will not be brute-causal, either, but will rather be guided by their higher-order perceptual awareness of the experience recognized. And given that they also have higher-order experiences with the contents *experience of green<sub>a</sub>*, *experience of smooth<sub>a</sub>*, *experience of loudness<sub>a</sub>*, and so on, they may be capable of possessing a generalized recognitional concept *that* [state of experience in general]. In which case, anyone deploying the higher-order recognitional concept *this* will be able to discern, *a priori*, the connection with their higher-order recognitional concept *that*, in just the same way that someone deploying the recognitional concept *red* will be able to discern the conceptual connection with their generalized recognitional concept *color*.

It seems that higher-order experiences provide us with just what we need in order to answer the qualia-theorist's challenge. They enable us to explain how we can possess purely-recognitional concepts of experience whose application can be grounded in awareness of the properties recognized, and in such a way that there can be *a priori* connections discernable amongst such concepts themselves. And higher-order experiences, too, provide just the necessary modes of presentation that intuition seems to require for our recognitional concepts of experience. But these benefits are provided at a considerable cost. For there are powerful objections to theories of inner sense.

One objection is this. If there really were such an organ of inner sense, then it ought to be possible for it to malfunction, just as our first-order senses sometimes do (Sturgeon, 2000). And in that case, it ought to be possible for someone to have a first-order percept with the content *red<sub>a</sub>* causing a higher-order percept with the content *seems orange<sub>a</sub>*. Someone in this situa-

tion would be disposed to judge, 'It is red', immediately and non-inferentially (i.e. not influenced by beliefs about the object's normal color or their own physical state), which would normally be sufficient grounds for us to say that the object seems red to them. But at the same time they would be disposed to judge, 'It *seems* orange'. Not only does this sort of thing never apparently occur, but the idea that it might do so conflicts with a powerful intuition. This is that our awareness of our own experiences is *immediate*, in such a way that to *believe* that you are undergoing an experience of a certain sort *is* to be undergoing an experience of that sort. But if inner-sense theory is correct, then it ought to be possible for someone to believe that they are in a state of *seeming-orange<sub>a</sub>* when they are actually in a state of *seeming-red<sub>a</sub>*.

Another objection to inner-sense theories is developed by Carruthers (2000). It is that, on the one hand, the computational demands placed on an organ of inner sense would surely be considerable (perceiving perceptions is going to be a task no easier than perceiving physical objects); and yet, on the other hand, there is no plausible story to be told about the powerful evolutionary pressures that would have been necessary to provide the incentive to build and maintain such an organ.

I shall not pursue these objections to inner-sense theory here. Rather, I shall show shortly (in section 5) that there may be a way of getting all of the benefits of this theory without any of the costs, by deploying a particular version of higher-order thought theory. First, however, I shall return to consider the alternatives in more detail. Are there any ways to explain our capacity for purely recognitional concepts of experience that *neither* appeal to intrinsic qualia *nor* to higher-order experiences?

#### 4 Can we do without HOEs?

Loar (1997) claims that phenomenal concepts (*viz.* purely recognitional concepts of experience) pick out the physical properties to which they refer *directly*, without the mediation of anything else. (Since Loar's account is designed to defend physicalism, the physical properties in question are those that are identical to, or that realize, the phenomenal properties of our experiences.) Put differently, he says that the physical properties of the brain that are also phenomenal properties *provide their own modes of presentation*—when identifying such a property recognitionally, there is no distinction between the property recognized and its mode of presentation to the subject.

It would seem, on the face of it, that this is just another version of the brute-causal account discussed earlier. Our recognitional judgments of experience are directly caused by (the physical states that are) our experiences, without the mediation of any further mental state. And it is apparent, too, that the account is subject to just the same difficulties as before. In particular, it can-

not accommodate the powerful intuition that we are *aware of*, and can introspect and contemplate, that which grounds our applications of our purely recognitional concepts. For on Loar's account, *nothing* grounds their application except the physical state-types that cause them. Nor can the account explain the *a priori* connections between recognitional concepts of particular experience-types and a recognitional concept of experience in general.

There are two alternative readings of Loar's (1997) position, however. One is that he intends to allow his anti-physicalist opponents the existence of qualia, in the strong sense adopted in this paper. Loar himself does not use this language; but he does stress that he can grant his opponents all of their initial intuitions and still block their argument to an anti-physicalist conclusion. Loar may simply be concerned to defend the view that qualia are (strictly identical with) physical states of the brain. And if this *is* his position, then he does not fall within the scope of my arguments here. As I emphasized in section 1 above, my goal is to establish that purely-recognitional concepts of experience *without qualia* require higher-order experiences.

The second possible alternative reading of Loar is that he is assuming some sort of representationalist or intentionalist reduction of phenomenal properties. For at the outset of his paper he allows that the phenomenally conscious properties of visual experience might coincide with 'internally determined intentional structure, so that it is an introspectable and non-relational feature of a visual experience that it represents things visually as being thus and so.' (Loar, 1997, p.597.) This looks, on the face of it, like an identification of phenomenal properties with narrowly-individuated intentional content. If so, then the position is no longer consistent with the existence of qualia, and is vulnerable to the arguments I shall present against Sturgeon immediately below.

Sturgeon (2000) develops an account of the relation between phenomenal consciousness and our recognitional concepts that is neutral as to the nature of the former, and yet that purports to explain the so-called 'explanatory gap' between them. But he also wants to claim that the explanatory gap is not in itself metaphysically troubling—once we understand the nature of the gap and how it arises, we should see that people committed to physicalism and/or naturalism needn't be concerned by it. If this is to be successful, however, then it is crucial that he should have an account of the relationship between phenomenal consciousness and our recognitional concepts, together with a view of the nature of the latter, that will work *whatever* the true nature of phenomenal consciousness should turn out to be.

Since Sturgeon (2000) is plainly sympathetic towards a form of intentionalist—or representationalist—approach to phenomenal consciousness, let us work through the commitments of his account under that assumption. So suppose that intentionalism is the truth about phenomenal consciousness;

suppose that phenomenally conscious states are just states possessing a certain sort of (analog and/or non-conceptual) intentional content. Can his account be made to work, in that case, without introducing higher-order experiences?<sup>8</sup>

Well, if there aren't any higher-order analog contents involved, then all that exists to ground a purely-recognitional judgment of *this* [experience of red], is the analog intentional content  $red_a$ . This is a first-order intentional content, appropriate to ground a first-order judgment of *red*. How does it give rise to the higher-order judgment *this*? 'No problem', Sturgeon may say, 'It causes it'. But this would just be a return to a form of brute-causal account.

The point is that our judgments of *this* [experience of red] seem related to the experience (which is, on the intentionalist hypothesis under consideration, none other than the first-order analog content  $red_a$ ), in just the sort of manner that judgments of *red* are related to redness. That is, they are recognitional judgments grounded in some sort of non-judgmental analog awareness of their objects. When I make judgments about my own experiences, they seem to be presented to me in something like the way that redness is presented to me when I make judgments of color—I am aware of a fineness of grain in what I recognize that slips through the mesh of my conceptual net, for example. But the first-order analog content  $red_a$  isn't the right *sort* of content to ground an awareness of the experiential state itself. It can ground recognition of redness, but not experienced-redness. What I am aware of, by virtue of being in a state with the analog content  $red_a$ , is redness, not experienced-redness. And all the fineness of grain in its content has to do with redness, not with the experience itself.

It does then seem that Sturgeon can't transcend a brute-causal account, if he tries to operate without appeal to either higher-order experiences or to qualia. In contrast, as soon as higher-order analog contents are admitted, the problems go away. A recognitional judgment of *red* is grounded in the analog content  $red_a$ , and a recognitional judgment of *this* [experience of red] is grounded in the analog content *experience of red<sub>a</sub>* or *seems red<sub>a</sub>*, which takes the experience of red as its object in something like the way that  $red_a$  takes redness as its object.

Papineau (2002) proposes a somewhat different theory. He now argues that purely recognitional concepts of experience can be formed on the back of our first-order recognitional concepts of colors, textures and so on, by prefacing such concepts with an operator of the form, 'The experience: ...'. We can

---

<sup>8</sup> Sturgeon wishes to remain neutral on *whether* intentionalism is true. But he also wants to claim that his account is *consistent with* intentionalism. What I shall argue is that such consistency requires higher-order experiences in order to succeed. He therefore can't avoid our question by pleading that he is neutral on the question of the truth of intentionalism.

set such an account a dilemma, however, depending on how the content of the experience-operator is said to be fixed.

Suppose, on the one hand, that the embedding concept of experience is in broad terms theoretical. Suppose, that is, that experience is here characterized in terms of causal role, or intentional content, or both. But in that case the experience-operator can't do the necessary work of explaining the content of phenomenal concepts. This is because the latter concepts can be free of any *a priori* connections with any causal-role or intentional-content concepts (hence the conceivability of zombies etc.). As we noted at the outset of the paper, I can think, 'Items of *this* type [experiences of red] might normally have been caused by decisions to speak', and so on.

Then suppose, on the other hand, that the embedding concept of experience is itself purely recognitional, in the sense that it refers directly to the property of being an experience, without theoretical mediation. (Papineau, 2002, speculates that its content might be fixed through some form of teleosemantics.) This enables us to meet one of our desiderata, at least—we can explain how it can be *a priori* for users of purely recognitional concepts of experience that what they are recognizing are experiences. This is because the concept 'experience' is actually a component in all such recognitional judgments.

The trouble, though, is that the account still can't accommodate our sense that we are directly *aware of* what grounds the application of a phenomenal concept, in a way that need involve no *a priori* connections with non-phenomenal concepts. For notice that on this account, when I recognize in myself a particular type of experience (as of red, say), what is actually going on is that I make a judgment of 'red' while prefacing it with an experience-operator. In which case it would surely have to be *a priori* that experiences of this type have something to do with redness. But in fact I can think, 'Experiences of *this* type might normally have been caused by greenness, or might even have occupied the causal role now occupied by pains'.

Tye (2000) occupies a position that seems to vacillate between those of Sturgeon and Papineau. Tye is a first-order representationalist about phenomenal consciousness itself (see also his 1995). He maintains that phenomenally conscious states are those with a certain sort of intentional content (non-conceptual and abstract), provided that they are poised in the right sort of way to have an impact on conceptual thinking and belief-formation. And he sometimes appears to suggest that phenomenal concepts are purely recognitional concepts that can be applied in the face of just such intentional contents. For example, he writes, 'The phenomenal concepts I apply and the features to which I apply them are the same in both the perceptual and the introspective cases.' (1995, p.167.) That is, whether I am judging *red* or *experience of red*, just the same concepts are involved. But this won't do.

Recognitional judgments of color are one thing, recognitional judgments of experiences of color quite another. And the latter cannot be grounded in first-order contents representing colors alone, as we saw in our discussion of Sturgeon above.

In other passages, on the other hand, Tye appears to suggest that the concept *experience* will always be a component in any recognitional judgment of experience. For example, he writes, 'Introspective awareness of phenomenal character, I maintain, is awareness-*that*—awareness that *an* experience with a certain phenomenal character is present.' (2000, p.52.) (Note that for Tye the phenomenal characters of an experience are the characters *represented* in the content of that experience—redness, greenness, or whatever.) But then this is really no different from the view of Dretske (1995), which we briefly mentioned and set aside in section 1 above—despite Tye's rhetoric concerning 'introspection' and 'recognitional concepts', the view is that we know of our experiences via awareness of the objects of our experience. Moreover, it runs up against the main difficulty we noted for Papineau: if phenomenal concepts like *this* [experience of red] are really concepts of the form, 'This experience *of this* [redness]', then there is the problem of explaining how I can nevertheless think, '*This* [experience of red] might not have been an experience *of this* [redness], and might not have been *an experience* at all, but rather a decision to speak.'

There are things Tye could say here, of course, perhaps drawing a distinction between *roles* and *role occupiers* (following Lewis, 1980). That is, he could explain the possibility represented when I think of circumstances in which *this* [type of experience] would exist but without any of its actual normal causes and effects, by saying that this is to think of the brain-state that actually occupies the causal role in question (experience of red, say) occurring, but in some other role ('mad experience of red'). And he could explain the possibility represented when I think of circumstances in which all the normal causes and effects of an experience of red are present, but without *this* [type of experience] occurring, by saying that this is to think of the causal role in question being occupied by a different type of physical state than in the actual circumstances ('Martian experience of red').

Such maneuvers cannot do justice to the original intuitions, however. For when we entertain thoughts of the form, '*This* type of experience could be / could have been F', we do not seem to be thinking thoughts of the form, 'The type of state that actually occupies such-and-such a causal role could be / could have been F'. Indeed, it seems possible to think, '*This* type of experience....' without *any* specification of a causal role figuring in the content of the thought (not even one governed by an actually-operator). In fact the phenomenology of such cases isn't that I think *through* a causal role to the type of state (whatever it is—presumably a brain state) that actually occupies that

causal role. It is rather that I think of a type of state that doesn't need to be specified by partly-indexical description, because I am directly aware of it. The referent of the phrase '*This* type of experience' seems to be present to consciousness, not 'hidden' beneath a causal role as the actual bearer of that role. And higher-order experience theory explains how this can happen.

In conclusion, it would appear that we have no real alternative, if we wish to explain how purely recognitional concepts of experience are possible without appealing to qualia, but to frame our account in terms of higher-order experiences. I now propose to sketch how this can be done without having to appeal to an organ of 'inner sense', by deploying a form of dispositionalist higher-order thought theory—a task that will occupy us through the final section of the paper.

### 5 How to get HOEs from HOTs (for free)

There are a number of different higher-order thought (HOT) theories on the market. The account to be presented here unites elements of a number of them, and then combines that account with an appeal to some or other form of consumer-semantics to explain how higher-order experiences (HOEs) will automatically be generated by the operations of a HOT faculty.

Rosenthal (1986, 1993) provides an account in terms of the *actual* occurrence of higher-order thoughts. For a state to be conscious is for it *actually* to be targeted by a higher-order thought at the time, where that thought is non-inferentially produced. Dennett (1978, 1991) offers a *dispositionalist* account, claiming that conscious status resides in *availability* to higher-order thought; but he also distinctively claims that these thoughts are to be expressed in natural language (so consciousness is essentially language-involving). Carruthers (1996) agrees with Dennett in offering an account that is dispositional, while dropping the alleged connection with natural language. But Carruthers also claims that the higher-order thoughts in question must themselves be available to higher-order thought (hence explaining conscious experience in terms of availability to *conscious* thought—he calls this 'reflexive thinking theory'). The present account shares the dispositionalism of Dennett (1991) and Carruthers (1996). But it rejects the language-involvement of the former, while also rejecting the latter's claim that the higher-order thoughts involved should be conscious ones. So it agrees with Rosenthal in allowing that the higher-order thoughts in virtue of (availability to) which a state is conscious will characteristically be *non-conscious*.

According to dispositionalist higher-order thought theory (Carruthers, 2000), the conscious status of a mental state or event consists in its non-inferential availability to a 'theory of mind' or 'mind-reading' system capable of higher-order thought. And a conscious experience, in particular, will be an experience that is available to cause higher-order thoughts about the occur-

rence and content of that very experience. We can then utilize some or other form of consumer semantics (either teleosemantics, or some form of functional or inferential-role semantics) in order to explain how our experiences acquire higher-order analog contents by virtue of their availability to higher-order thought.<sup>9</sup>

According to all forms of consumer semantics, the intentional content of a state depends, at least in part, on what the down-stream consumer systems that can make use of that state are disposed to do with it. And there is independent reason to think that changes in consumer-systems can transform perceptual contents, and with it phenomenal consciousness. (See Hurley, 1998, for presentation and discussion of a wide range of examples that are interpretable in this light.)

Consider the effects of spatially-inverting lenses, for example (Welch, 1978). Initially, subjects wearing such lenses see everything upside-down, and their attempts at action are halting and confused. But in time—provided that they are allowed to move around and act while wearing their spectacles—the visual field rights itself. Here everything on the input side may remain the same as it was when they first put on the spectacles; but the planning and action-controlling systems have learned to interpret those states inversely. And as a result, intentional perceptual contents become normalized.<sup>10</sup>

If consumer semantics is assumed, then it is easy to see how mere dispositions can transform contents in the way that dispositionalist higher-order thought theory supposes. For notice that the consumer-system for a given state does not *actually* have to be making use of that state in order for the latter to carry the appropriate content—it just has to be *disposed* to make use of it should circumstances (and what is going on elsewhere in cognition) demand. So someone normalized to inverting spectacles does not actually have to be acting on the environment in order to see things right-side-up. He can be sitting quietly and thinking about something else entirely. But still the spatial content of his perceptual states is fixed, in part, by his dispositions to think and move in relation to the spatial environment.

Consider, here, the implications of some form of inferential role semantics, in connection with a different example. What is it that confers the con-

---

<sup>9</sup> For exposition and defense of different forms of teleosemantics, see Millikan (1984, 1986, 1989) and Papineau (1987, 1993). For some varieties of inferential role semantics, see Loar (1981, 1982), McGinn (1982), Block (1986), and Peacocke (1986, 1992).

<sup>10</sup> I should emphasize that while consumer semantics provides a possible and plausible explanation of the inverting-lenses phenomenon, this isn't actually forced on us. For there remains the possibility that the righting of visual experience may be caused by feedback from motor systems to the visual system, giving rise to alterations in the internal operations of the latter. This would then be an explanation in terms of changes on the input side of conscious experience, rather than an account in terms of changes on the output side using consumer-semantics.



tent  $P\&Q$  on some complex belief-state of the form ‘ $P\#Q$ ’? (The sign ‘#’ here is meant as a dummy connective, not yet interpreted.) In part, plainly, it is that one is disposed to infer ‘ $P$ ’ from ‘ $P\#Q$ ’ and ‘ $Q$ ’ from ‘ $P\#Q$ ’ (Peacocke, 1992). It is constitutive of a state with a conjunctive content that one should be disposed to deduce either one of the conjuncts from it. But of course this disposition can remain un-activated on some occasions on which a conjunctive thought is entertained. For example, suppose that I hear the weather-forecaster say, ‘It will be windy and it will be cold’, and that I believe her. Then I have a belief with a conjunctive content even if I do nothing else with it. Whether I ever form the belief that it will be windy, in particular, will depend on my interests and background concerns, and on the other demands made on my cognitive resources at the time. But my belief still actually has a conjunctive content—it has it categorically—in virtue of my inferential dispositions.

According to dispositionalist higher-order thought theory, then, the availability of our perceptual states to a ‘theory of mind’ or ‘mind-reading’ faculty is sufficient to transform the intentional contents of those states. Where before, in the absence of such a faculty, the states had merely first-order contents—containing analog representations of worldly color, texture, shape, and so on—now all of those states will have, at the same time, higher-order analog, experience-representing, contents. Each state that is an analog representation with the content  $red_a$  is at the same time an analog representation with the content *experience of  $red_a$* , in virtue of the fact that the theory-of-mind system contains concepts of experience that can be applied to those very states.

We are now in position to explain how purely-recognitional concepts of experience are possible, obtaining all of the advantages of ‘inner sense’ theory without any of the associated costs. Here is how the story should go. We begin—both in evolutionary terms and in normal child development—with a set of first-order analog contents available to a variety of down-stream consumer systems. These systems may include a number of dedicated belief-forming modules, as well as a practical reasoning faculty for figuring out what to do in the light of the perceived environment together with background beliefs and desires. One of these belief-forming systems will be a developing mind-reading system.

When our mind-reading faculty has reached the stage at which it confers on us an understanding of the subjective nature of experience, and/or a grasp of the is/seems distinction, then we will easily—indeed, trivially—become capable of second-order recognitional judgments of experience, with these judgments riding piggy-back on our first-order recognitional concepts (in something like the way that Papineau, 2002, outlines, as discussed in section 4 above). So if subjects had a recognitional concept *red*, they will now

acquire the concept *seems red*, or *experience of red*, knowing (a) that whenever a judgment of 'red' is evoked by experience, a judgment of 'seems red' is also appropriate on the very same grounds; and (b) that a judgment of 'seems red' is still appropriate whenever a disposition to judge 'red' has been blocked by considerations to do with abnormal lighting or whatever. Note that at this stage the higher-order concept in question is still a theoretically embedded one, with conceptual connections to worldly redness (it is, after all, a seeming *of red*). What one recognizes the state *as* is a state whose normal cause is worldly redness, and so on.

This change in the down-stream mind-reading consumer system is sufficient to transform all of the contents of experience, rendering them at the same time as higher-order ones. So our perceptual states will not only have the first order analog contents *red<sub>a</sub>*, *green<sub>a</sub>*, *loud<sub>a</sub>*, *smooth<sub>a</sub>*, and so on, but also and at the same time the higher-order analog contents *experience of red<sub>a</sub>*, *experience of green<sub>a</sub>*, *experience of loudness<sub>a</sub>*, *experience of smoothness<sub>a</sub>*, and so on. The subject will then be in a position to form recognitional concepts targeted via just these higher-order contents, free of any conceptual ties with worldly redness, greenness, loudness, and smoothness. And once possessed of such concepts, it is possible for the subject to wonder whether other people have experiences of *this* sort when they look at a ripe tomato, to conceive of worlds in which zombies perceive red without undergoing *this experience*, and so on.

Here we have an account of our purely recognitional concepts of experience that appeals to higher-order experiences, but without the need to postulate any sort of organ of inner sense. So (in contrast with inner-sense theory) there should be no problem in telling some sort of evolutionary story concerning the emergence of higher-order experience. This now reduces to the problem of explaining the emergence of our 'theory of mind' capacity, and some or other version of the 'Machiavellian intelligence' hypothesis might suffice here (Byrne and Whiten, 1988, 1998). Moreover, it should also be obvious why there can be no question of our higher-order analog contents getting out of line with their first-order counterparts, on this account—in such a way that one might be disposed to make recognitional judgments of *red* and *seems orange* at the same time, for example. This is because the content of the higher-order experience *seems red<sub>a</sub>* is parasitic on the content of the first-order experience *red<sub>a</sub>*, being formed from it by virtue of the latter's availability to a 'theory of mind' system.

Before closing I should stress once again that although the present account of how purely-recognitional concepts of experience are possible is drawn from higher-order reductive theories of phenomenal consciousness, that is not how it is being used in the present context. First-order theorists of phenomenal consciousness like Dretske (1995) and Tye (1995) might agree with the

present use of higher-order thought theory to explain the possibility of purely-recognitional concepts of experience, while rejecting that theory as an account of phenomenal consciousness as such. They merely need to claim that phenomenal consciousness is already present in creatures that lack any capacity for higher-order thought, and also perhaps in perceptual states in us that are unavailable to such thought.<sup>11</sup>

In fact the present account should be acceptable to a wide range of different theorists, provided only that they are prepared to endorse some form of consumer semantics as one determinant, at least, of intentional content. For it should then be plain that higher-order experiences with higher-order analog contents can come to exist by virtue of the availability of first-order analog contents to a faculty of higher-order thought, without any need to postulate 'inner scanners' or any organ of inner sense. And it can be by virtue of the existence of such higher-order experiences we come to form purely-recognitional concepts of experience, grounded in those higher-order analog contents. In any case, anyone of reductionist sympathies who does *not* endorse consumer semantics in general (or the particular use being made of it here), and who is reluctant to believe in the existence of an organ of inner sense, is still left with the challenge of explaining how purely recognitional concepts are possible without qualia.<sup>12</sup>

---

<sup>11</sup> In fact, since Dretske (1995) endorses a form of teleosemantics, while Tye (1995) opts for a form of pure causal-covariance (or input-side) semantics, the present proposals could be acceptable to Dretske but not to Tye. For an extended critique of first-order theories of all sorts, as well as for full defense of dispositionalist higher-order thought theory as a reductive account of phenomenal consciousness, see Carruthers (2000).

<sup>12</sup> I am grateful to Scott Sturgeon for a conversation and series of questions that prompted this paper; and to Fred Dretske, Scott Sturgeon, and two anonymous referees for comments on an earlier draft.

## References

- Armstrong, D. 1968. *A Materialist Theory of the Mind*. Routledge.
- Armstrong, D. 1984. Consciousness and causality. In D. Armstrong and N. Malcolm, *Consciousness and Causality*, Blackwell.
- Block, N. 1986. Advertisement for a semantics for psychology. *Midwest Studies in Philosophy*, 10, 615-678.
- Block, N. 1995. A confusion about a function of consciousness. *Behavioral and Brain Sciences*, 18, 227-247.
- Byrne, R. and Whiten, A. eds. 1988. *Machiavellian Intelligence*. Oxford University Press.
- Byrne, R. and Whiten, A. eds. 1998. *Machiavellian Intelligence II: Evaluations and extensions*. Cambridge University Press.
- Carruthers, P. 1996. *Language, Thought and Consciousness*. Cambridge University Press.
- Carruthers, P. 2000. *Phenomenal Consciousness: a naturalistic theory*. Cambridge University Press.
- Chalmers, D. 1996. *The Conscious Mind*. Oxford University Press.
- Chalmers, D. 1999. Materialism and the metaphysics of modality. *Philosophy and Phenomenological Research*, 59, 473-496.
- Dennett, D. 1978. Toward a cognitive theory of consciousness. In C. Savage, ed., *Minnesota Studies in the Philosophy of Science*, 9.
- Dennett, D. 1991. *Consciousness Explained*. Allen Lane.
- Dretske, F. 1995. *Naturalizing the Mind*. MIT Press.
- Fodor, J. 1987. *Psychosemantics*. MIT Press.
- Fodor, J. 1998. *In Critical Condition*. MIT Press.
- Hurley, S. 1998. *Consciousness in Action*. Harvard University Press.
- Jackson, F. 1982. Epiphenomenal qualia. *Philosophical Quarterly*, 32, 127-136.
- Jackson, F. 1986. What Mary didn't know. *Journal of Philosophy*, 83, 291-295.
- Kripke, S. 1972. Naming and necessity. In G. Harman and D. Davidson, eds., *Semantics of Natural Language*, Reidel.
- Lewis, D. 1980. Mad pain and Martian pain. In N. Block, ed., *Readings in Philosophy of Psychology*, vol. I, Harvard University Press.
- Loar, B. 1981. *Mind and Meaning*. Cambridge University Press.
- Loar, B. 1982. Conceptual role and truth-conditions. *Notre Dame Journal of Formal Logic*, 23, 272-283
- Loar, B. 1990. Phenomenal states. *Philosophical Perspectives*, 4.
- Loar, B. 1997. Phenomenal states. In N. Block, O. Flanagan and G. Güzeldere, eds., *The Nature of Consciousness*, MIT Press.
- Lycan, W. 1996. *Consciousness and Experience*. MIT Press.

- McGinn, C. 1982. The structure of content. In A. Woodfield, ed., *Thought and Object*, Oxford University Press.
- McGinn, C. 1991. *The Problem of Consciousness*. Blackwell.
- Millikan, R. 1984. *Language, Thought, and Other Biological Categories*. MIT Press.
- Millikan, R. 1986. Thoughts without laws: cognitive science with content. *Philosophical Review*, 95, 47-80.
- Millikan, R. 1989. Biosemantics. *Journal of Philosophy*, 86, 281-297.
- Papineau, D. 1987. *Reality and Representation*. Blackwell.
- Papineau, D. 1993. *Philosophical Naturalism*. Blackwell.
- Papineau, D. 2002. *Thinking about Consciousness*. Oxford University Press.
- Peacocke, C. 1986. *Thoughts*. Blackwell.
- Peacocke, C. 1992. *A Study of Concepts*. MIT Press.
- Rosenthal, D. 1986. Two concepts of consciousness. *Philosophical Studies*, 49, 329-359.
- Rosenthal, D. 1993. Thinking that one thinks. In Davies and Humphreys, eds., *Consciousness*, Blackwell.
- Siewert, C. 1998. *The Significance of Consciousness*. Princeton University Press.
- Sturgeon, S. 1994. The epistemic view of subjectivity. *Journal of Philosophy*, 91, 221-235.
- Sturgeon, S. 2000. *Matters of Mind*. Routledge.
- Tye, M. 1995. *Ten Problems of Consciousness*. MIT Press.
- Tye, M. 2000. *Consciousness, Color and Content*. MIT Press.
- Welch, R. 1978. *Perceptual Modification*. Academic Press.
- Wittgenstein, L. 1953. *Philosophical Investigations*. Blackwell.