

Mind, forthcoming

Phenomenal Consciousness, by Peter Carruthers. Cambridge: Cambridge University Press, 2000. Pp. xx + 347. H/b £37.50.

This book is one of the latest in what is now a rather long line of attempts to “naturalise” phenomenal consciousness. The theory on offer is a dispositionalist version of the higher-order thought (HOT) theory: simplifying a bit, an experience E is phenomenally conscious just in case it is disposed to cause the belief that the subject is having E (p. 227). Carruthers does a lot of useful spadework before the theory gets a full-dress defence. Chapters 2 and 3 attempt to overcome various obstacles to Carruthers’ project of explaining phenomenal consciousness in broadly physicalistic terms—the argument from conceivability, Jackson’s knowledge argument, McGinn’s mysterianism, and so forth. Chapter 4 argues that for the purpose of the reductive project, “intentional content should be characterised *narrowly*” (p. 88). Chapters 5 and 6 examine, respectively, the strengths and weaknesses of “first-order representational” (FOR) theories—the kind proposed by Dretske (*Naturalizing the Mind*, Cambridge, MA: MIT Press, 1995) and Tye (*Ten Problems of Consciousness*, Cambridge, MA: MIT Press, 1995). The extended argument for the dispositionalist HOT theory occupies chapters 7-9. Chapter 10 criticizes three rival dispositionalist HOT theories that, unlike Carruthers’ theory, constitutively tie phenomenal consciousness to language. These rivals are Carruthers’ earlier theory in *Language, Thought, and Consciousness* (Cambridge: Cambridge University Press, 1996);

and the accounts in Dennett's "Toward a Cognitive Theory of Consciousness" (reprinted in *Brainstorms*, Brighton: Harvester Press, 1981, pp. 149-73), and *Consciousness Explained* (Boston: Little, Brown, 1991). The next and final chapter concerns some objections to Carruthers' theory that can be extracted from Dennett's defence (in *Consciousness Explained*) of his "multiple drafts" model of consciousness.

Carruthers' writing has always been refreshingly straightforward, and the present book is no exception. The level of argument is generally high, and a good deal of light is shed on many topics. The discussion of narrow content is problematic, however. The main illustrative example concerns "Peter, Paul and Mary [who] are each undergoing subjectively indistinguishable experiences *as of* a cat bristling with rage" (p. 104). Peter and Paul are confronted with the cats Tiger and Stripe, respectively, while Mary is hallucinating a cat. Carruthers maintains that all three subjects are entertaining "thoughts of the very same type" that they each would express with the words "That cat is dangerous" (pp. 104-5). In other words, the three subjects are entertaining thoughts with the same "narrow content". This narrow content is not a *proposition*; rather, it (presumably) determines a function from contexts to (typically) propositions: "While narrow-content types do not have truth-conditions intrinsically, they will normally acquire one or another truth-condition when tokened in a particular context." (p. 109) So plugging in Peter's context to the relevant function yields the proposition that Tiger is

dangerous, while Paul's context yields the proposition that Stripe is dangerous; Mary's context, on the other hand, does not yield a proposition at all (p. 109).

It is quite unclear, though, what the rest of this function looks like. Suppose Tom is looking at Tiger who is asleep on a mat; Dick is looking at Tiger who has been disguised as a skunk; and Harry is hallucinating a cat bristling with mild irritation; they all utter "That cat is dangerous". Are any of these subjects entertaining thoughts with the same narrow content as the thoughts of Peter, Paul and Mary? As far as I can see, nothing in Carruthers' official explanation provides an answer. This and related complaints are by now familiar (see, especially, Stalnaker, "On What's in the Head", reprinted in his *Context and Content*, Oxford: Oxford University Press, 1999), but they are ignored. This is doubtless because Carruthers thinks he doesn't need an answer to the "general question" of the "individuation conditions" for narrow content (p. 110). But if we are only given a small hint, Carruthers is not entitled to assert that "narrow content...figures in the laws and nomic generalisations of intentional psychology" (p. 110).

With narrow content supposedly thus secured, Carruthers deploys it against the threat posed by Block's "Inverted Earth" (*Philosophical Perspectives* 4, 1990, pp. 53-79); here he departs from Dretske, Tye, and also his closest ally, Lycan (*Consciousness and Experience*, Cambridge, MA: MIT Press, 1996). In Block's much-discussed story, an Earthling has colour-inverting lenses implanted and is transported to a near-duplicate of Earth, where everything is painted in its complementary colour, and the inhabitants speak

a language like English but with “inverted” colour words. Tomatoes are green on Inverted Earth (and are called “red”), but because of the inverting lenses the subject notices no change. If we further suppose (as Block argues) that the (wide) intentional contents of the Earthling’s colour experiences shift with time, so that his experiences of tomatoes come to represent them as *green*, then a representational approach to phenomenal consciousness is in trouble. After the shift, the subject’s tomato-experiences have quite different contents from his earlier tomato-experiences back on Earth, yet they are alike in phenomenal character. With Block, Carruthers holds that such a shift in (wide) content would occur. Against Block, he denies that “we need to appeal to qualia to explain the sense in which his experience remains subjectively the same for him...Rather, we can explain the subjective continuity of his experience by appeal to the narrow contents of that experience.” (p. 109) (Block, incidentally, has a rather different conception of narrow content: see “Inverted Earth”, p. 79, n. 26.)

At least in the case of colour experience, it is perfectly clear how Carruthers intends us to *use* the expression “narrow content”: a subject’s experience has “narrow-red” content just in case his experience would be described by Dretske, Lycan and Tye as an experience as of a red object or, equivalently, his experience would be described by Block as having a certain (non-intentional) quale. But this takes us no closer to understanding narrow content, because an experience with narrow-red content is *not* supposed to be, simply, an experience that represents something as red (this is shown by

Carruthers' treatment of the Inverted Earth example), or an experience with a certain quale.

Further clarification is promised when Carruthers ends his defence of narrow content by comparing it to McGinn's "weak externalism" (*Mental Content*, Oxford: Blackwell, 1989): "[t]he gap between weak externalism and belief in narrow content is small; and for present purposes I am happy to allow it to be crossed." (p. 112) Specifically, "the difference between weak-externalist content and narrow content is that the former, but not the latter, requires the existence of the relevant properties *in* [i.e. instantiated in] the world in which the content-to-be-characterised figures." (p. 113)

Unfortunately, this comparison does not help. In the first place, Carruthers' explanation of "weak externalism" is not quite accurate. It is the doctrine that if a subject is in a mental state with the content that *x* is *F*, then the property *Fness* exists (perhaps without being instantiated), and, for certain fillings for "*x*" (e.g. "that dog"), the object *x* exists (see McGinn, pp. 36-43). So the weak externalist is free to hold what Carruthers says he must deny, that "someone can have an experience *as of* a cube...in a world in which there are no cubes." (p. 113) Given this, by the quotation in the previous paragraph it would appear to follow that there is *no* gap between weak externalism and belief in narrow content. But that cannot be correct: the weak externalist's contents are just ordinary *propositions*—and Carruthers' narrow contents are not.

Although Carruthers relies heavily on narrow content, the difficulties just mentioned need not seriously affect his case for the dispositionalist HOT theory. Someone suspicious of narrow content who has another way of dealing with “inversion” arguments for qualia could agree with most of the argument in the book (occasionally deleting the word “narrow”). So for the rest of this review we can forget about narrow content. We can also set aside another complication, Carruthers’ claim that “perceptual contents are *analog* as opposed to digital, at least in relation to the concepts we possess” (p. 11). (This is related to, but is not quite the same as, the more common claim that perceptual contents are “non-conceptual”.)

The argument for Carruthers’ version of the dispositionalist HOT theory proceeds in three basic stages. First, the field is whittled down to representational theories of phenomenal consciousness. Second, FOR theories fall to the objection that they “cannot really explain the *feel*, or ‘what-it-is-likeness’, of phenomenally conscious experience.” (p. 147) That leaves higher-order representational (HOR) theories, for instance Lycan’s higher-order experience (HOE) theory and Rosenthal’s “actualist” HOT theory (e.g., “Two Concepts of Consciousness”, *Philosophical Studies* 49, pp. 329-59, 1986); the third stage knocks out these rival HOR theories, leaving the dispositionalist HOT theory as the only one standing. It is impossible to summarise, much less evaluate, all these arguments in a few pages; somewhat at random, I shall briefly discuss the second stage, and then raise a question about Carruthers’ own theory.

The objection to FOR theories uses a distinction between “*worldly* subjectivity”—“what the *world*...is like for an organism”—and “*experiential* subjectivity”—“what the organism’s *experience of the world*...is like for the organism.” (pp. 127-8) I am not completely confident that I understand this distinction (which is apparently a distinction primarily between *properties*), but if I do, it may be illustrated as follows. Suppose an organism has colour vision. Then what the *world* is like for the organism is, *inter alia*, coloured. In other words, the organism’s perceptual experience *represents* the world as coloured. So we may say that a *worldly-subjective property* (relative to this organism) is simply a property *represented by* the organism’s perceptual experience: redness, for example. *Experientially subjective properties*, on the other hand, are just familiar Nagelian “what it’s like” properties. For someone with normal colour vision, experiences as of tomatoes and strawberries are saliently alike in respect of what it’s like to undergo them; all these experiences therefore share an experientially subjective property. This property is a *property of* experiences, and is of course not redness—worldly subjective properties are plainly different from experientially subjective properties. (For a similar distinction, see Lycan, pp. 76-7.)

With this distinction in hand, the problem with FOR theories is supposed to be that while they “can almost certainly provide a successful explanation of *worldly* subjectivity...it is very difficult to see how the further, additional, subjectivity of (some) *experience* can be explained without introducing higher-order representations (HORs)

into the account.” (pp. 147-8) That is, the FOR theorist can provide a naturalistic account of, say, colour vision—*worldly* subjectivity. But he cannot provide a naturalistic account of *experiential* subjectivity—*what it’s like* for an organism with colour vision. And, of course, this is to say in other words that the FOR theorist cannot explain phenomenal consciousness.

The FOR theorist may reply that to explain worldly subjectivity just *is* to explain experiential subjectivity, because the experientially subjective property distinctive of experiences as of tomatoes and strawberries is identical to the property of *being an experience that represents something as red*. Carruthers’ response—compressing a long line of argument—is that many experiences (including those that represent redness) do not *have* experiential subjectivity. Carruthers gives a wide range of examples, from Armstrong’s absent-minded driver to recent experimental work on the existence of a “nonconscious” action-guiding visual sub-system (here Carruthers draws heavily on Milner and Goodale, *The Visual Brain in Action*, Oxford: Oxford University Press, 1995). One might object that content-bearing events available to guide movement should be called “experiences” only if they are phenomenally conscious. However, as Carruthers notes, this just redescribes the problem as one of explaining in naturalistically acceptable terms why some content-bearing events are “experiences” (p. 167). FOR theorists typically sort those content-bearing events that are phenomenally conscious from the rest by appealing to functional role: in order to be phenomenally conscious an experience

must be “poised” (Tye’s term) to affect other cognitive systems in certain ways. Carruthers raises serious challenges to a number of variations on this theme.

The moral of the attack on FOR theories can be put as follows. Don’t try to explain phenomenal consciousness just in terms of content and functional role, for it will inevitably be obscure why the favoured functional role makes the crucial difference. As Carruthers puts it in connection with Kirk’s FOR theory (*Raw Feeling*, Oxford: Oxford University Press, 1994): “It is mysterious why an analog state’s being made available to another set of cognitive systems...should suddenly confer on it the properties of phenomenal consciousness—properties which it did not, by hypothesis, possess prior to being so available.” (p. 170)

This is a fair point. But, as it turns out, Carruthers seems vulnerable to his own objection. For Carruthers’ dispositionalist HOT theory explains phenomenal consciousness using basically the *same* meagre materials—content plus functional role. The HOT apparatus is in a sense irrelevant.

To explain. On Carruthers’ theory, what makes an experience phenomenally conscious is its *disposition* to be targeted by a HOT, which naturally prompts the following question. “How can something which hasn’t actually happened to a perceptual state (namely, being targeted by a HOT) confer on it—categorically—the dimension of subjectivity?” (p. 242) Carruthers answers by claiming that on the “inferential role

semantics” he endorses, “first-order analog representations of the environment” take on an “enriched dual content” when they are disposed to cause the appropriate HOTs:

Each experience of the world...becomes at the same time a representation that just such an experience is taking place; each experience with the content *red_a*, say, [i.e. the “narrow analog” content *red*] is at the same time an event with the content *seems red_a* or *experience of red_a*. And [these experiences] have these contents categorically, by virtue of the *powers* of the HOT consumer system, in advance of any HOT actually being tokened. (p. 242)

That is, if an experience with the (narrow analog) content *red* is disposed to cause the appropriate HOT, this makes the experience actually have the “dual content” *red* and *experience of red*. So, according to Carruthers, when I look at a tomato, my experience represents *that the tomato is red and I am having an experience that represents that the tomato is red*, or something along these lines. And it is the content of the experience—specifically, the fact that it has “both ‘objective’, or world-body-representing content and ‘subjective’, or experience-representing content” (p. 244)—that explains why there is something it’s like for me to look at the tomato. The only role of the HOT system is to explain why my experience has that dual content in the first place;

if some *other* account of the dual content were forthcoming, the fundamental explanation of phenomenal consciousness in terms of “objective” and “subjective” content would be unchanged.

But now—waiving worries about whether phenomenally conscious experiences really do have these “dual contents”—we seem to be back in exactly the same bind that afflicted FOR theorists. Given Carruthers’ broad conception of an experience, whether or not there in fact *are* experiences with dual content that are phenomenally *non-conscious*, surely there *could be*. Once Carruthers has accustomed us to non-conscious experiences (or, if you insist, “quasi-experiences”) with (only) the content *red*, it is hard to see why there couldn’t be non-conscious experiences with the dual content *red* and *experience of red*. So an experience with dual content won’t necessarily be phenomenally conscious, and therefore an extra naturalistic ingredient is required. Functional role is the obvious candidate, but if the earlier objection to FOR theories is correct, this proposal won’t work.

It should come as no surprise if Carruthers has not pulled off his ambitious naturalising project. However, he has succeeded admirably in producing a rich and fascinating book, empirically well-informed and philosophically astute. Anyone interested in consciousness needs to read it.

ALEX BYRNE

Massachusetts Institute of Technology