19 August 2020

Global Workspace Theory and Animal Consciousness

Jonathan Birch

Centre for Philosophy of Natural and Social Science,

London School of Economics and Political Science,

Houghton Street, London, WC2A 2AE, UK.

Email: j.birch2@lse.ac.uk

Webpage: http://personal.lse.ac.uk/birchj1

This is a preprint of an article that will appear in *Philosophical Topics.*

7082 words

Abstract: Carruthers has recently argued for a surprising conditional: if a global workspace theory of phenomenal consciousness is both correct and fully reductive, then there are no substantive facts to discover about phenomenal consciousness in non-human animals. I present two problems for this conditional. First, it rests on an odd double-standard about the ordinary concept of phenomenal consciousness: its intuitive non-gradability is taken to be unchallengeable by future scientific developments, whereas its intuitive determinacy is predicted to fall by the wayside. Second, it relies on dismissing, prematurely, the live empirical possibility that phenomenal consciousness may be linked to a core global broadcast mechanism that is (determinately) shared by a wide range of animals. Future developments in the science of consciousness may lead us to reconsider the non-gradability of phenomenal consciousness, but they are unlikely to lead us to accept that there are no facts to discover outside the paradigm case of a healthy adult human.

Key words: animal consciousness, phenomenal consciousness, consciousness science, global workspace theory, phenomenal concepts

1. Carruthers' surprising conditional

For more than thirty years, scepticism about attributions of phenomenal consciousness[1] to (states of) non-human animals has been a major theme in Peter Carruthers' work (Carruthers 1989, 1992, 1999, 2000, 2004, 2005, 2018a,b, 2019, 2020). Yet the nature of his scepticism, and the motivation for it, look to have recently changed. For a long time, scepticism about animal consciousness seemed an inevitable consequence of Carruthers' support for a higher-order thought (HOT) theory of consciousness. The HOT theory posits that consciousness requires a capacity for thinking about first-order representations, and it is doubtful whether any non-human animals possess such capacities. So much the worse for animal consciousness, said Carruthers (1989, 2000, 2005). So much the worse for the plausibility of the HOT theory, said the critics (Jamieson and Bekoff 1992; Tye 1995; Dretske 1995).

A few years ago, Carruthers (2017) abandoned the HOT theory, rejecting all his earlier reasons for adopting it. The PDF file of this paper on his website is called "Carruthers recants". Despite this, Carruthers has not become any more sympathetic to attributions of consciousness to non-human animals. In two recent articles, he has developed a new sceptical argument that is independent of the HOT theory (Carruthers 2018a, b).[2] This new argument

[1] If a mental state is phenomenally conscious, there's something it's like to be in that state. The state is felt by a subject. This is not much of a definition, but it is hard to do better. See Schwitzgebel (2016) for discussion of the concept of phenomenal consciousness and the possibility of defining it any more precisely.

[2] This argument is also presented in *Human and Animal Minds* (Carruthers 2019, pp. 140-155), where it shares the stage with a second, partially independent argument based on the

instead derives its force from the empirical plausibility of the global workspace theory (GWT) of consciousness developed by Baars, Dehaene and colleagues (Baars 1989, 2005, 2017; Dehaene et al. 1998; Dehaene and Naccache 2001; Dehaene and Changeux 2011; Dehaene 2014).

The global workspace theory posits the existence in the brain of a global broadcast mechanism that integrates representations from perceptual systems, affective systems and memory systems and broadcasts the integrated content back to the input systems and onwards to a wide variety of consumer systems, including mechanisms of voluntary report, planning, reasoning and decision-making (Dehaene and Changeux 2011, p. 209). Although many of the finer details of the theory have changed over the years, this core commitment has remained intact.[3] The representations currently being broadcast are said to be in the global workspace. Interpreted as a theory of phenomenal consciousness and not just cognitive access, the theory states that a representation becomes phenomenally conscious when it enters the workspace, whereas more localised processing outside the workspace occurs without phenomenal consciousness. The empirical case for the existence of a global broadcast mechanism (reviewed in Dehaene 2014) rests on well-established experimental paradigms (especially

---

semantics of ascriptions of consciousness to non-human animals (pp. 155-161). I discuss that second argument briefly in Section 7.

[3] Baars's (1989) focus was on the cognitive architecture of the global workspace, with only speculative remarks about its neural implementation, whereas Dehaene and colleagues' "global neuronal workspace" theory offers an account of the neural implementation of the workspace that draws extensively on neuroimaging data. See Section 5 for more discussion of the relation between the Baars and Dehaene et al. versions of GWT.

backward masking and the attentional blink) that allow us to compare, using neuroimaging, the processing that results from unconscious perception of a stimulus with the processing that results from conscious perception of the same stimulus.

As Carruthers is well aware, there is no consensus behind GWT as an adequate theory of phenomenal consciousness in humans. One naturally wonders: Could there not be phenomenally conscious states outside the workspace? What methodological approach could settle this question, given that states outside the workspace would not be available to voluntary report? This is the territory of the long-running "overflow" debate (see Overgaard 2018; Phillips 2018 for critical reviews of recent work). The debate is ongoing, and few would argue that the "overflow" side has been decisively defeated (see, e.g., Bronfman et al. 2014, 2019 for a recent case for overflow).

There is, however, a serious possibility that the overflow debate will eventually be resolved in favour of the "no overflow" view, and thus a serious possibility that GWT will emerge as at least an extensionally adequate theory of phenomenal consciousness in humans: a theory that places the boundary between the phenomenal and the non-phenomenal in the right place. It is worth thinking, then, about what *would* follow for consciousness in non-human animals, were GWT to build up overwhelming empirical support as a theory of human phenomenal consciousness.

Here Carruthers enters the fray with a surprising conditional claim:

> If a global workspace theory of phenomenal consciousness is correct, and is fully reductive in nature, then we should stop asking questions about consciousness in

> nonhuman animals—not because those questions are too hard to answer, but because there are no substantive facts to discover. (Carruthers 2018a, p. 47)

I will refer to this as Carruthers' "surprising conditional". It highlights the sense in which the nature of Carruthers' scepticism, and not just its motivation, has changed. Carruthers used to think there were facts of the matter about consciousness in non-human animals—and that consciousness was determinately not present. Now, he thinks there is simply no fact of the matter.

Much hangs on what Carruthers means by "fully reductive", and this will receive detailed discussion below. For now, it is enough to note that Carruthers sees a realistic possibility of GWT emerging not only as an extensionally adequate theory of phenomenal consciousness in humans but also as a theory that fully explains away the appearance of an explanatory gap or hard problem of consciousness. This dissolution of the hard problem/explanatory gap is built in to the antecedent of his conditional.

Carruthers' argument deserves careful scrutiny, and my aim is to provide that scrutiny. I will first explain, briefly, the argument *for* the conditional (although readers should really consult Carruthers 2018a, b and 2019 on this). I will then present two problems for the argument. The first is that it rests on an odd double-standard about the ordinary concept of phenomenal consciousness. The second is that Carruthers rejects, without good reason, the possibility that phenomenal consciousness could be linked to a core global broadcast mechanism that is determinately present in a wide range of animals.

2. Simplifying the surprising conditional

To give myself a clearer target, I want to simplify Carruthers' surprising conditional. The conditional as stated above generates scope for confusion. It is a conditional about what we *should* do if GWT is correct and fully reductive (we should stop asking questions about animal consciousness), but the controversial and interesting part is the idea that "there are no substantive facts to discover". My target will be the following conditional, which I will read as a material conditional:

> *If GWT is a correct and fully reductive theory of phenomenal consciousness in humans, then, for all non-human animals, there is no fact of the matter as to whether or not they possess phenomenally conscious states.*

The debate is not about the *truth or falsity* of this conditional. I agree with Carruthers that the conditional is probably true, but only because I think the antecedent is false, and a material conditional with a false antecedent is vacuously true. That doesn't mean I think Carruthers is correct on the issues at stake.

What is really at stake is whether the conditional is *non-vacuously* true, and whether, accordingly, there is a sound modus ponens argument for a "no fact of the matter" view about animal consciousness. Such an argument would go like this:

> *Premise 1:* GWT is a correct and fully reductive theory of phenomenal consciousness in humans.

Premise 2: If GWT is a correct and fully reductive theory of phenomenal consciousness in humans, then, for all non-human animals, there is no fact of the matter as to whether or not they possess phenomenally conscious states.

Conclusion: For all non-human animals, there is no fact of the matter as to whether or not they possess phenomenally conscious states.

Carruthers appears sympathetic to this argument but, wary of continuing debate about Premise 1, stops short of endorsing it. I take his central claim to be that the above argument is *plausibly* sound, because Premise 1 is at least plausible and, on the presupposition that Premise 1 is true, Premise 2 is non-vacuously true.

The hedged nature of Carruthers' discussion forces his critics into an awkward position. We can't just point to continuing debate about overflow, because Carruthers is well aware of this and has already priced it in. Our challenge is to show that the above argument *isn't even plausibly sound*. We need to show that, no matter how the science of consciousness develops, this argument won't be vindicated. This is the challenge I aim to meet.

3. Motivating the surprising conditional

What motivates the surprising conditional? It is motivated, Carruthers argues, by "a mismatch between our concept of phenomenal consciousness (which is all-or-nothing) and arguably our best theory of the property that the concept picks out (which admits of degrees across species)" (Carruthers 2018a, p. 54).

The idea is that GWT is our best theory of phenomenal consciousness, and non-human animals will approximate the human global broadcast mechanism to varying degrees. Even in other great apes, there will be some differences, because not *all* the consumer systems of the human global workspace will be in place: mechanisms supporting language, and perhaps (more controversially) mechanisms supporting theory-of-mind and normative cognition will be absent. As we look at more phylogenetically distant taxa (rodents, birds, reptiles, amphibians, fish, invertebrates…), we will find further varieties of broadcast mechanism, no doubt differing in many ways from the human global broadcast mechanism, with substantially fewer consumer systems. GWT tells us that entry to the human global broadcast mechanism is sufficient for phenomenal consciousness, but it is silent on which (if any) of these different forms of broadcast mechanism will also confer phenomenal consciousness on their entrants. It is not that GWT *denies* that these mechanisms are sufficient: it is just completely silent on the issue.

So much the worse for GWT, one might think. There must be some fact of the matter about which forms of non-human broadcast mechanism confer phenomenal consciousness on their entrants and which do not. If GWT remains silent on this issue, it cannot be a complete theory of phenomenal consciousness. Wrong, says Carruthers: our intuition that there is some fact of the matter here is not unassailable. If GWT provides a correct and *fully reductive* theory of human consciousness, despite remaining entirely silent on questions of non-human consciousness, then we should be prepared to give up our initial intuition that there is a fact of the matter in the non-human case.

We see here the weight being carried by the phrase "fully reductive". The mere correctness of GWT, as a theory of the mechanisms supporting human consciousness, is no reason at all to

take its silence on questions of animal consciousness as implying there is no fact of the matter about those questions. A theory can be correct yet partial, limited in its scope to a particular domain (in this case, a particular species). But Carruthers' thought is this: if the theory is also *fully reductive*, in the sense that it fully dissolves the appearance of a hard problem or explanatory gap, it becomes harder to insist that the theory is incomplete and that some further theory is needed. The theory has fully satisfied the explanatory need a theory of phenomenal consciousness must answer. We can use our theory to "explain everything that stands in need of an explanation—including the appearance of an explanatory gap" (Carruthers 2018a, p. 58).

How is GWT supposed to do that? Its main proponents, Baars and Dehaene, have never made such claims on its behalf. But Carruthers sees a close and harmonious relationship between GWT and the Phenomenal Concept Strategy (PCS)—the strategy that he and others have long seen as the right way for a materialist to respond to the explanatory gap (Carruthers 2000; Papineau 2002; Carruthers and Veillet 2007; Balog 2012). The aim of the PCS is to explain the existence of an explanatory gap and its attendant intuitions about zombies, Mary and inverted qualia (and I will assume some familiarity with these thought experiments here, rather than explaining them) by appeal to phenomenal concepts. Phenomenal concepts (such as the concepts of phenomenal redness or phenomenal blueness) are hypothesized to be inferentially sealed off from functional and physical concepts (such as the concepts of a neuron or a brain), in such a way as to lead to the conceivability of physical/functional duplicates without conscious experiences.

An anti-materialist will happily grant the existence of such concepts, but will take them to refer to phenomenal properties that are themselves non-physical and non-functional—to

qualia, in one sense of that term. But the key insight of the PCS is that we can posit phenomenal concepts without having to posit special phenomenal properties to which they refer. We can be, to use Carruthers' own term, *qualia irrealists*. We can maintain that phenomenal concepts, despite generating various dualist intuitions by virtue of their inferential isolation from other concepts, nonetheless refer to physical properties.

This is where, for Carruthers, GWT enters the scene: it gives us a theory of the nature of the physical properties in question. For Carruthers, phenomenal concepts are indexical, acquaintance-based concepts of globally broadcast non-conceptual contents. When a content enters the global workspace, we are able to form a concept of it in a special way. We form an indexical concept of *this-R*—this nonconceptual content in the workspace now. A group of globally broadcast nonconceptual contents comes to be conceptualized as phenomenal redness, a different group is conceptualized as phenomenal blueness, and so on. These concepts are inferentially isolated from physical/functional concepts, so we end up prone to dualist intuitions, even though the referents of these concepts are not qualia but physical states of the global workspace.

So, when Carruthers describes GWT as a *fully reductive* theory, he means that, when combined with the PCS, it has the resources to explain the existence of an explanatory gap or a hard problem, and to do so without positing special phenomenal properties. With this done, he argues, nothing else can reasonably be demanded of a theory of consciousness. The story is finished. He is aware that some readers will continue to demand that a complete theory should yield a determinate fact of the matter about the distribution of phenomenal consciousness in the natural world, but he thinks this is an unreasonable demand that flows from a tacit dualism. For Carruthers, if the conjunction of GWT and PCS (henceforth:

GWT+PCS) leaves our questions about non-human animals unsettled, then that is because these questions do not have determinate answers.[4]

4. Carruthers' cat

At the core of Carruthers's argument is the assumption that phenomenal consciousness is all-or-nothing, or *non-gradable*. It does not come in degrees: a state is either fully phenomenally conscious or it is not phenomenally conscious at all. Carruthers needs this assumption for two reasons. The first is that, according to GWT, entry to the human global broadcast mechanism is all-or-nothing. There are no degrees of entry: a content is either fully inside the global workspace or fully outside. So, Carruthers' identification of globally broadcast nonconceptual contents with phenomenally conscious contents could not be the whole story if phenomenal consciousness came in degrees. Some further theory would be needed about how the degree of phenomenal consciousness relates to the properties of the global broadcast network. Moreover (and this is the second reason), if we had such a theory, it might imply determinate facts of the matter about consciousness in non-human animals. For example, if we had a well-confirmed theory according to which broadcast to a minimal set of systems (e.g. perception, motor control and memory) was sufficient for a minimal degree of

[4] Carruthers is not the first to defend such a view. David Papineau (2002, 2020), another prominent defender of the PCS, also maintains that, once we fully reject dualism and fully embrace what Chalmers (2010) has called "Type-B materialism", we will abandon the idea that there are determinate facts of the matter about consciousness in non-human animals. He too thinks the demand for facts about the distribution of consciousness in nature flows from tacit dualism (though see Balog 2020 for criticism).

phenomenal consciousness, we could then infer that an animal with the minimal network was at least minimally phenomenally conscious. Carruthers' argument for indeterminacy would unravel.

I understand, then, why Carruthers makes this assumption. What's harder to understand is the combination of an unwavering insistence on the non-gradable, all-or-nothing nature of phenomenal consciousness with a relaxed attitude to the possibility that it might be indeterminate. When it comes to the question "Does this animal instantiate *zero* phenomenally conscious states during its lifetime, or *at least one* phenomenally conscious state during its lifetime?", Carruthers believes that there is no fact of the matter. One might well ask: How can the answer be indeterminate if, for any given state of an animal, that state must be either fully conscious or fully non-conscious?

Carruthers' position is that there are determinate answers to questions about phenomenal consciousness only for humans, because entry to the global workspace confers phenomenal consciousness and entry to the global workspace is all-or-nothing. In other animals, entry to the broadcast mechanism that most closely approximates human global broadcast is still all-or-nothing, but it is indeterminate whether or not this mechanism confers phenomenal consciousness on its entrants, because GWT+PCS fails to yield a determinate fact of the matter.

To see the intuitive strangeness of the view, consider the representations that enter, say, the closest approximation of the human global broadcast mechanism in a cat's brain. Are these representations consciously experienced by the cat or not? It is indeterminate, on Carruthers'

view, whether these representations are phenomenally conscious or not. Carruthers' cat neither determinately feels nor determinately does not feel.

Moreover, while it is natural to choose a cat as the example here, any creature with broadcasting capabilities that fall short of those of a healthy adult human with a fully functioning global broadcast network will do just as well. Infants, if they have a partial global broadcast network in which some of the consumer systems (e.g. theory-of-mind, planning, reasoning, verbal report) have not yet developed, neither determinately feel nor determinately do not feel. Patients in a minimally conscious state, if they have a partial global broadcast network with some of the consumer systems incapacitated, neither determinately feel nor determinately do not feel.

Carruthers, then, takes the concept of phenomenal consciousness to be like this: It does not admit of *gradations* (states cannot be more or less conscious) but it does admit of *indeterminacy* (states can be indeterminate between fully conscious and fully non-conscious). And this indeterminacy runs rampant, with wildly counterintuitive results, once we look beyond the core case of a healthy adult human with a fully functioning global workspace.

Speaking for myself, I think my intuitive concept of phenomenal consciousness is of something that is not just non-gradable but also determinately present or absent, on or off, in a given animal at a given time. The cat feels or it does not, as surely as it is either alive or dead. So does the infant or the minimally conscious patient. Of course, the content experienced by such a patient may be much less rich, along various dimensions, than that of a healthy adult (Bayne et al. 2016), but there will intuitively be a fact of the matter about

whether the patient has some conscious experiences or no conscious experiences. Carruthers must hold that I am mistaken about my own concept: I mistakenly assume that the determinacy which holds for states of healthy adult humans also holds for states of non-human animals, infants, and patients in a minimally conscious state. But if I can be wrong about this, couldn't I also be wrong about gradability? Couldn't I also be mistaken in thinking that the non-gradability that holds for states of humans also holds for states of non-human animals?

I am unsure about the rules of this game. Carruthers takes one widely but not universally held view about the ordinary concept of phenomenal consciousness (its non-gradability in all cases) to be indisputable and guaranteed to withstand the progress of the science of consciousness, and another widely but not universally held view about the ordinary concept of phenomenal consciousness (its determinacy in all cases) to be in error and likely to fall by the wayside. The double-standard is puzzling. I do not myself see a strong reason for thinking that the all-or-nothing assumption will survive future scientific progress.[5]

If we did reach a point at which GWT stood triumphant as a well-confirmed, correct and fully reductive theory of phenomenal consciousness in healthy adult humans, it seems to me that a sensible reaction would be to reconsider the non-gradability of the concept of phenomenal consciousness, so as to make sense of how there could still be facts of the matter about consciousness in other cases. It would make sense to explore ways of quantifying grades of global broadcasting across species and grades of phenomenal consciousness, so that we could (perhaps) empirically identify the property of being phenomenally conscious to grade *n* with the property of entering a global workspace of grade *n*. This would be counterintuitive, but

[5] See, for example., Godfrey-Smith's (2020) contribution to this special issue.

rampant indeterminacy is at least as counterintuitive. I don't see what would recommend Carruthers' preferred reaction, whereby we simply resign ourselves to there being no facts of the matter about consciousness outside the paradigm case of a healthy adult human.

5. The core mechanism response

Carruthers maintains that phenomenal consciousness is all-or-nothing—in the sense that it is non-gradable—whereas global broadcast mechanisms vary continuously across species. The last section considered one reaction to this: If we were to accept GWT as a correct, fully reductive theory of phenomenal consciousness, would it not make sense to revise our concept of phenomenal consciousness to make room for gradability? Carruthers must already accept that the concept is revisable, since he urges us to revise its intuitive determinacy, yet he takes the intuitive non-gradability of the concept to be unalterable.

I now want to consider a different reaction: Is it really true that global broadcasting varies continuously across species, with no sharp boundary marking the point at which it goes from determinately present to determinately absent? Could it not be the case that there is a *core global broadcast mechanism* that is determinately present in many animals and determinately absent in others?

This putative core mechanism integrates information from perceptual, evaluative and mnemonic systems and broadcasts the integrated representations back to those systems and onwards to systems of motor control. It ensures that the brain's critical behaviour-guiding systems are all on the same page, working with the same integrated representations of body and world. The identity conditions of the core mechanism are not sensitive to the precise

input systems or consumer systems. The inputs and consumers will vary across species, but the same core mechanism is still there, whether or not (for example) a language module or a planning module is receiving the broadcast.

If there is such a core mechanism, with fairly coarse-grained identity conditions, then perhaps there will be no need, in the hypothetical future in which GWT is strongly empirically supported, to revise our ordinary concept of phenomenal consciousness. Perhaps we will simply come to empirically identify the property of phenomenal consciousness with the property of being broadcast by the core mechanism.

It is worth emphasizing that the idea of a core broadcast mechanism is not an ad hoc amendment to the global workspace theory. Dehaene and Changeux state the basic commitment of the global workspace theory as follows:

> The global neuronal workspace (GNW) hypothesis … proposes that associative perceptual, motor, attention, memory, and value areas interconnect to form a higher-level unified space where information is broadly shared and broadcasted back to lower-level processors. (Dehaene and Changeux 2011, p. 209).

This idea is a captured in a widely reproduced figure in Dehaene et al. (1998), a figure that, for many consciousness researchers, encapsulates the idea of the global workspace. In that figure, there is no significant role for broadcast to a wide range of consumer systems, including verbal report, reasoning and planning. That was clearly important for Baars (1989), but it appears to be less important for Dehaene and Changeux, for whom the essential feature of a global workspace is that brain areas associated with perceptual, attentional, mnemonic,

evaluative and motoric processing interconnect, so that integrated representations are broadcast back to the input systems and onwards to motor systems. It makes sense to ask which non-human animals have a mechanism of this general type and which do not. This is not an ill-formed question, or an unproductive question to ask. Could it be that all and only those animals that possess such a mechanism have phenomenally conscious experiences? Dehaene (2014, p. 246) appears sympathetic to this idea, writing that "I would not be surprised if we discovered that all mammals, and probably many species of birds and fish, show evidence of a convergent evolution to the same sort of conscious workspace."[6]

[6] Thanks to Henry Shevlin for pointing me to this quotation.

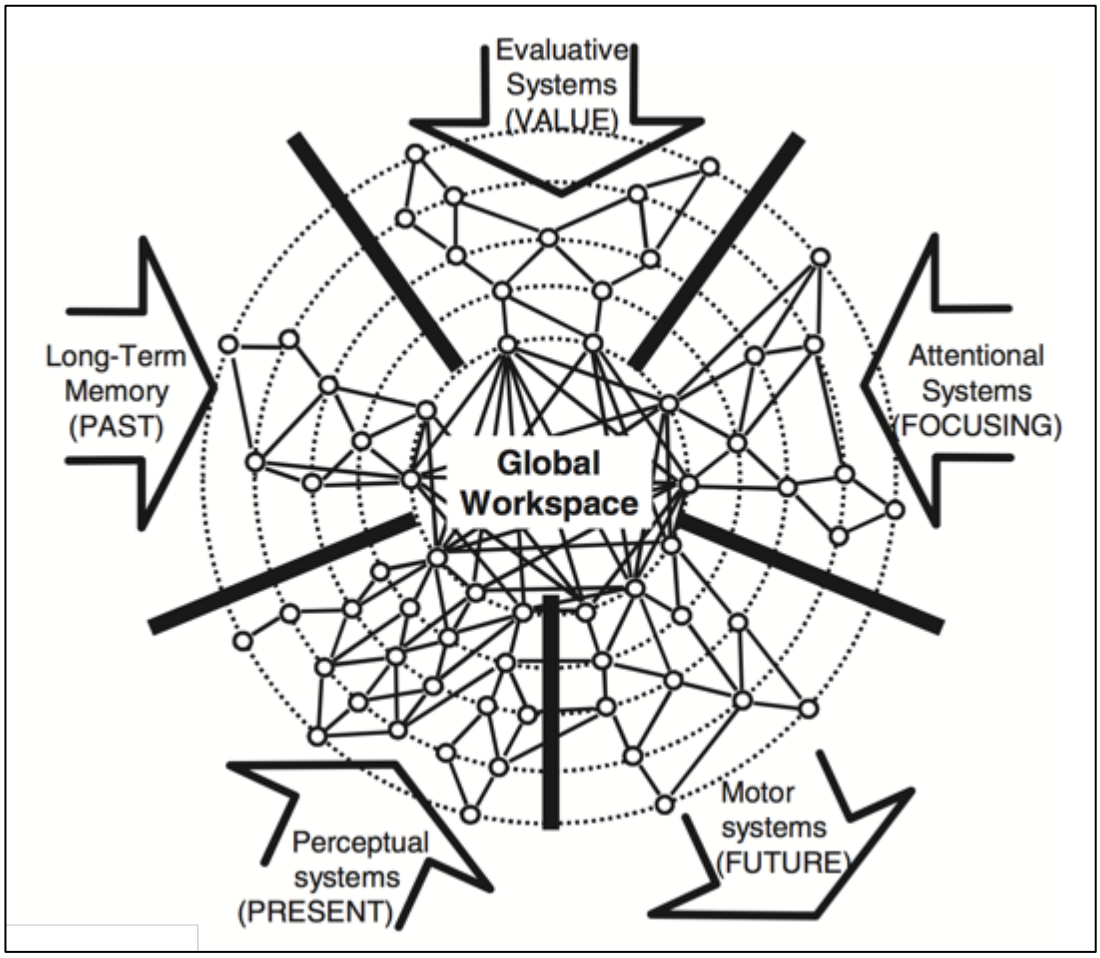

Figure 1: The figure used by Dehaene et al. (1998) to illustrate the basic idea of a global workspace. Note that broadcast to a wide range of consumer systems such as planning, reasoning and verbal report does not feature in the figure. © (1998) National Academy of Sciences. The copyright holder permits non-commercial reuse.

Carruthers (2017, p. 84) too appears to have been briefly tempted by this idea, writing that "what constitutes phenomenal consciousness is being a globally broadcast non-conceptual state. And there is plenty of reason to think that many species of animal (perhaps all vertebrates) have states of that general kind." Carruthers (2018a,b; 2019, Section 7.2) raises the same possibility again in more recent work, but this time to dismiss it. He floats the idea that the "global workspace can be identified with those attention-like networks across species that play the 'centering' role in the species in question" (2019, p. 147). He then gives two

main reasons for rejecting the idea that the property of phenomenal consciousness could be identified with the property of being broadcast by the core mechanism. But neither reason is convincing.

Here is the first reason: the core mechanism alone is insufficient to explain our explanatory-gap-generating intuitions about consciousness. To explain these intuitions, we need the whole human global broadcast network, including the consumer systems that are involved in creating and applying phenomenal concepts. The underlying assumption seems to be that, for a mechanism to be determinately sufficient for phenomenal consciousness, it must be sufficient for the generation and application of phenomenal concepts, and thereby sufficient to explain the explanatory gap.

But why accept this assumption? It is an unwelcome echo of the HOT theory. A defender of GWT+PCS is free to accept that the referents of our phenomenal concepts—globally broadcast nonconceptual contents—can still exist in the absence of such concepts. Animals which possess the core mechanism will have contents of the same kind, despite (presumably) lacking the ability to reflect upon them. The same kind of broadcasting is still happening, and the same kind of content is being broadcast, but the receivers are less sophisticated. GWT+PCS gives us no reason to think the presence or absence of phenomenal consciousness depends on the sophistication of the receivers.

Carruthers' second reason resurfaces on several occasions in his (2019) book, as well in his (2018a, b) articles. In short, it is this: for phenomenal consciousness to be identified with the contents of a core global broadcast mechanism, phenomenal concepts (including the concept

of phenomenal consciousness itself) would need to be natural kind concepts, but they are not. As Carruthers puts it:

> [O]ne thing that everyone has agreed on since Kripke (1980) is that terms referring to phenomenally conscious mental states aren't used as natural-kind terms. In contrast, it is generally agreed that our concepts for substances like water *are* natural-kind ones. Even before we knew anything about chemistry, we used the concept WATER to refer to the underlying nature or essence of the recognizable stuff that fills our lakes and rivers (H_2O); and it turned out that it was that very same stuff that presents as ice in some circumstances (frozen water) and as mist in others (evaporated water). But our phenomenal concepts aren't like that. We don't use them with the intention of referring to whatever natural kind underlies those experiences, whatever that might turn out to be, and however that kind might be presented in other creatures. On the contrary, we mean to refer just to the qualities we are aware of in ourselves. (Carruthers 2019, pp. 147-8)

Carruthers, in this passage, appears to assume that the reference of a concept depends on referential intentions. We intend phenomenal concepts to refer to "qualities", so they can't turn out to refer to natural kinds. This is deeply puzzling in the context of Carruthers' overall package of views. He is, after all, a qualia irrealist who holds that phenomenal concepts refer to globally broadcast nonconceptual contents. We don't *intend* them to refer to globally broadcast nonconceptual contents. It just turns out, given GWT+PCS, that these are the properties they track. Carruthers' deployment of the PCS crucially relies on the idea that phenomenal concepts refer to the physical properties that they in fact track, and need not refer to what we *think* they pick out or *intend* them to pick out. We think they refer to qualia,

hence our vulnerability to hard-problem-generating thought experiments, but their real referents are not qualia.

Carruthers could consistently hold on the views expressed in the above passage, accept that phenomenal concepts are non-referring (because there are no "qualities" that fit our referential intentions), and endorse eliminativism about phenomenal consciousness. Or, to avoid eliminativism, he could consistently hold on to the view that phenomenal concepts pick out globally broadcast nonconceptual contents despite our referential intentions, grant that referential intentions do not fix reference, and allow that the concept of phenomenal consciousness itself (i.e. the very concept of there being "something it's like" to be me) might refer to a property quite different from the sort of property we think it picks out. But he can't have it both ways.

If we grant that referential intentions do not fix reference, we leave open the possibility that the concept of phenomenal consciousness does indeed refer to a neurobiological or cognitive natural kind, such as a core global broadcast mechanism. A version of the second option is defended elsewhere in this volume by Tim Bayne and Nicholas Shea (2020), who also discuss the above passage from Carruthers, though they remain neutral as to whether the natural kind in question is a form of global broadcast or something else. I conclude that Carruthers has not given us good reasons to reject the core mechanism response.

6. Crystal balls

It is strange to be debating the implications for animal consciousness of possible future developments in the science of human consciousness. If strong evidence *against* GWT as a

theory of phenomenal consciousness comes along in the near future, Carruthers' surprising conditional will be vacuously true but uninteresting. This is one plausible resolution to this debate. But I also see two other plausible possibilities, suggested by the discussions of Section 5 and 4 respectively:

(1) As GWT becomes increasingly well-confirmed as a theory of consciousness in humans, animal consciousness researchers study the global broadcast networks of non-human animals and find a reasonably crisp boundary between those animals that possess a core global broadcast mechanism and those that do not. This core mechanism is the best candidate for the property that attracts the reference of the ordinary concept of phenomenal consciousness. So, researchers make an empirical identification: the property of phenomenal consciousness is the property of being broadcast by the core mechanism.

(2) As GWT becomes increasingly well-confirmed as a theory of consciousness in humans, animal consciousness researchers study the global broadcast networks of non-human animals and find *no* crisp boundary between those animals that possess a core global broadcast mechanism and those that do not. There is simply a continuum of variation, right down to the simplest attentional networks in the simplest nervous systems, such as that of the nematode worm Caenorhabditis elegans. Researchers conclude that the concept of phenomenal consciousness itself requires revision. It intuitively refers to an all-or-nothing property, but we need to reconstruct it as referring to a graded property. A state is phenomenally conscious to degree *n* when it is globally broadcast to degree *n*.

What I don't find at all likely is a scenario in which the rise of GWT leads us to abandon the idea that there are determinate facts of the matter about animal consciousness. Admittedly, this too is possible, but I suspect that idea of a core global broadcast mechanism that is determinately present in a wide range of animals will survive empirical scrutiny. If it does not, I suspect that the concept of phenomenal consciousness will prove to be more labile on the question of gradability and less labile on the question of determinacy than Carruthers predicts.

7. Epilogue

This paper has focussed on the arguments of Carruthers' (2018a, b) articles. In *Human and Animal Minds,* these considerations are collectively described as "the negative semantic argument" (Carruthers 2019, pp. 140-154). A different, partially independent set of considerations is then presented for the view that there are no facts of the matter about animal consciousness, which Carruthers calls "the positive semantic argument" (Carruthers 2019, pp. 155-160). This positive semantic argument is also presented in Carruthers' contribution to this volume (Carruthers 2020). The positive semantic argument does not logically depend on the global workspace theory, or on the assumption that phenomenal consciousness is non-gradable, so at first glance it seems to escape the criticisms presented here. The argument instead rests on a positive account of the semantics of ascriptions of consciousness to other beings.

The positive account in question is the following:

> The truth-conditions of a judgement like, "Creature C has perceptual states that are *this-E*" [where *this-E* is a phenomenal concept] might be this: "If I were to be aware of creature C's perceptual states, then I would judge them to be *this-E*." (Carruthers 2019, p. 157)

In other words, the truth-conditions of ascriptions of consciousness to other beings are posited to involve counterfactuals about how I would (or would not) first-personally apply my own phenomenal concepts to their states, given the chance. It is then argued that, when the target of the ascription is a non-human animal, these counterfactuals are unevaluable—there is no fact of the matter about whether they are true or not. For "there is no fact of the matter about how human first-person concepts of experience would or would not apply if instantiated in the mind of the animal" (Carruthers 2019, p. 158).

I'm inclined to agree that these counterfactuals are unevaluable. However, this argument is only as compelling as the semantic theory on which it rests. What motivates that theory? Why should we believe that the truth conditions of ascriptions of consciousness to non-human animals involve counterfactuals about how I would first-personally apply my own concepts to their states, given the chance? Crucial to motivating this theory is the rejection of an alternative already discussed, on which our phenomenal concepts (including the concept of phenomenal consciousness itself) refer to natural kinds (Carruthers 2019, p. 156; Bayne and Shea 2020). We have seen that Carruthers has not given us compelling reasons to reject that alternative theory (see the end of Section 5). He repeats the consideration that we don't *intend* to use these concepts to pick out natural kinds (2019, pp. 156), but referential intentions are not decisive. If we insist that our referential intentions decisively fix the reference of phenomenal concepts, then they either refer to intrinsic qualities of experiences

or to nothing—not to globally broadcast nonconceptual contents. So, Carruthers' positive semantic argument is unconvincing for the same reason that his attempt to block the core mechanism response is unconvincing. In both cases, referential intentions are called upon to carry more weight than they can bear.

I do not want to downplay the difficulty of constructing a positive semantic account of ascriptions of phenomenal consciousness—an account that explains how the concept of phenomenal consciousness can refer determinately to a single material property, such as entry to a core global broadcast mechanism, despite differing significantly from a canonical natural kind concept. The difficulty of this challenge has been rightly emphasized by David Papineau (2002, Ch. 7). Katalin Balog (2020) even goes so far as to call it the "hardest problem of consciousness". This is an unsolved problem, but not a good reason to "stop caring about consciousness". If one possible semantic account fails to yield determinate truth conditions, that is a reason to look for a different semantic account, not a reason to abandon the search.

Acknowledgements

I thank Tim Bayne, Peter Carruthers, Stevan Harnad, Eva Jablonka, Nick Shea, Henry Shevlin and an anonymous reviewer for their comments and advice. This work was funded

by the European Research Council (ERC) under the European Union's (EU) Horizon 2020 research and innovation programme (Grant agreement No. 851145).

References

Baars, B. J. (1989) *A Cognitive Theory of Consciousness*. Cambridge: Cambridge University Press.

Baars, B. J. (2005) Global workspace theory of consciousness: toward a cognitive neuroscience of human experience. *Progress in Brain Research* 150:45-53.

Baars, B. J. (2017). The global workspace theory of consciousness: Predictions and results. In Schneider, S. & Velmans, M. (Eds.), *The Blackwell Companion to Consciousness*. 2nd edition (pp. 227-242). Hobokon, NJ: Wiley-Blackwell.

Bayne, T., Hohwy, J. and Owen, A. M. (2016) Are there levels of consciousness? *Trends in Cognitive Sciences* 20:405-413.

Bayne, T. and Shea, N. (2020) Consciousness, concepts and natural kinds. *Philosophical Topics*.

Balog, K. (2012) In defense of the phenomenal concept strategy. *Philosophy and Phenomenological Research* 84:1-23.

Balog, K. (2020) Hard, harder, hardest. In Arthur Sullivan (ed.), *Sensations, Thoughts, and Language: Essays in Honor of Brian Loar*. New York, USA: Routledge. pp. 265-289.

Bronfman, Z. Z., Brezis, N., Jacobson, H. & Usher, M. (2014). We see more than we can report: "Cost free" color phenomenality outside focal attention. *Psychological Science* 25:1394-1403.

Bronfman, Z. Z., Jacobsen, H. and Usher, M. (2019) Impoverished or rich consciousness outside attentional focus: Recent data tip the balance for Overflow. *Mind and Language* 34:423-444.

Carruthers, P. (1989) Brute experience. *Journal of Philosophy* 86:256-269.

Carruthers, P. (1992) *The Animals Issue: Moral Theory in Practice*. Cambridge: Cambridge University Press.

Carruthers, P. (1999) Sympathy and subjectivity. *Australasian Journal of Philosophy* 77:465-482.

Carruthers, P. (2000) *Phenomenal Consciousness: A Naturalistic Theory*. Cambridge: Cambridge University Press.

Carruthers, P. (2004) Suffering without subjectivity. *Philosophical Studies* 121:99-125.

Carruthers, P. (2005) *Consciousness: Essays from a Higher-Order Perspective*. Oxford: Oxford University Press.

Carruthers, P. (2013) Evolution of working memory. *Proceedings of the National Academy of Sciences USA* 110:10371-10378.

Carruthers, P. (2017) In defence of first-order representationalism. *Journal of Consciousness Studies* 24:74-87.

Carruthers, P. (2018a) Comparative psychology without consciousness. *Consciousness and Cognition* 63:47-60.

Carruthers, P. (2018b) The problem of animal consciousness. *Proceedings and Addresses of the APA* 92:179-205.

Carruthers, P. (2019) *Human and Animal Minds: The Consciousness Questions Laid to Rest*. Oxford: Oxford University Press.

Carruthers, P. (2020) Stop caring about consciousness. *Philosophical Topics*.

Carruthers, P. and Veillet, B. (2007) The phenomenal concept strategy. *Journal of Consciousness Studies* 14:212-236.

Chalmers, D. J. (2010) *The Character of Consciousness*. New York: Oxford University Press.

Dehaene, S. (2014) *Consciousness and the Brain: Deciphering How the Brain Encodes Our Thoughts*. New York: Viking Press.

Dehaene, S. and Changeux, J-P. (2011) Experimental and theoretical approaches to conscious processing. *Neuron* 70:200-227.

Dehaene, S. and Naccache, L. (2001) Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework. *Cognition* 79:1-37.

Dehaene, S., Kerszberg, M. and Changeux, J-P. (1998) A neuronal model of a global workspace in effortful cognitive tasks. *Proceedings of the National Academy of Sciences USA* 95:14529-14534.

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

Godfrey-Smith, P. (2020) Gradualism and the evolution of experience. *Philosophical Topics*.

Jamieson, D. and Bekoff, M. (1992) Carruthers on non-conscious experience. *Analysis* 52:23-28.

Overgaard, M. (2018) Phenomenal consciousness and cognitive access. *Philosophical Transactions of the Royal Society B: Biological Sciences* 373:20170353. https://doi.org/10.1098/rstb.2017.0353

Papineau, D. (2002) *Thinking about Consciousness*. Oxford: Oxford University Press.

Papineau, D. (2020) The problem of consciousness. In Kriegel, U. (Ed.), *The Oxford Handbook of the Philosophy of Consciousness* (pp. 14-38). New York: Oxford University Press.

Phillips, I. (2018) The methodological puzzle of phenomenal consciousness. *Philosophical Transactions of the Royal Society B: Biological Sciences* 373:20170347. https://doi.org/10.1098/rstb.2017.0347

Schwitzgebel, E. (2016). Phenomenal consciousness, defined and defended as innocently as I can manage. *Journal of Consciousness Studies*, *23*, 224-235.