Phenomenology and the Cognitive Sciences (2020) 19:803–819
https://doi.org/10.1007/s11097-019-09633-6

Phenomenality, conscious states, and consciousness inessentialism

Mikio Akagi[1]

Published online: 12 August 2019

Abstract
I draw attention to an ambiguity of the expression 'phenomenal consciousness' that is an avoidable yet persistent source of conceptual confusion among consciousness scientists. The ambiguity is between what I call *phenomenality* and what I call *conscious states*, where the former denotes an abstract property and the latter denotes a phenomenon or class of its instances. Since sentences featuring these two terms have different semantic properties, it is possible to equivocate over the term 'consciousness'. It is also possible to fail to distinguish between statements that are true about conscious states in virtue of their phenomenality, and statements that are true in virtue of other properties of conscious states. I review empirically informed arguments by scientists Bernard Balleine and Anthony Dickinson, Stevan Harnad, and Jeffrey Alan Gray, arguing that each of them makes errors based on the ambiguity. I conclude with some tentative suggestions for avoiding further confusion about the ambiguity.

Keywords Phenomenal consciousness · Consciousness science · Consciousness inessentialism · The hard problem of consciousness

1 Introduction

Phenomenal consciousness is an unusually recalcitrant object of investigation. It is difficult to operationalize or quantify, and resists converging characterizations (see e.g. Irvine 2013). PHENOMENAL CONSCIOUSNESS as a concept is not much more yielding, and there is little that can be said about it that is uncontroversial. Because we have such a frail grip on either the facts or the idea, it can be difficult to think clearly about consciousness. In this paper I will draw attention to a logical feature of consciousness talk, and discuss some errors that arise in consciousness scholarship in connection with

✉ Mikio Akagi
m.akagi@tcu.edu

[1] John V. Roach Honors College, Texas Christian University, TCU Box 297022, Fort Worth, TX 76129, USA

this feature. The errors that concern me are not the result of the unique structure of consciousness as a concept or as a natural phenomenon; they arise from a general feature of abstract concepts, and are made, I expect, not because consciousness is an especially idiosyncratic concept, but because our understanding of consciousness is so frail that some of our normal rational mechanisms for detecting confused thinking get little grip. The result, anyway, is that even well-respected consciousness researchers sometimes engage in confused thinking.

The focal point of my discussion is the ambiguity of the expression 'phenomenal consciousness' between what I will call *phenomenality* and what I will call *conscious states*. This ambiguity is not a special one for consciousness; it is not a substantive ambiguity between natural phenomena that we might conflate, like creature consciousness and consciousness of contents, or P-consciousness and A-consciousness (Block 1995). The ambiguity that concerns me is a semantic ambiguity between related but logically distinct uses of certain abstract expressions. Since the distinction is a logical one, and not about interpretations of empirical results or recommendations about scientific kinds, my claims should not be empirically controversial. Nor will they be empirically illuminating. My goal is merely to promote conceptual clarity as a supplement to the empirical study of consciousness, not as a substitute for it.

After describing the distinction between phenomenality and conscious states more precisely, I will discuss the role that the distinction plays in three discussions by consciousness scientists concerning consciousness inessentialism. First, I will describe a paper by Bernard Balleine and Anthony Dickinson in which the conflation between phenomenality and conscious states features quite baldly. Second, I will review an argument by Stevan Harnad about consciousness inessentialism that features more subtle confusions that arise from the ambiguity. Third, I will examine Jeffrey Alan Gray's discussion of the hard problem of consciousness, arguing that although he is more careful than Harnad about the distinction he succumbs to an equivocation much like Balleine and Dickinson do. My hope is that the consideration of examples and the clear distinguishing of explananda will serve to make this equivocation less tempting.

2 A logical distinction

On the one hand, the expression 'phenomenal consciousness' can refer to an abstract property of states or processes (i.e. of psychological, cognitive, or neural states or processes, henceforth *mental states* or *states*[1]). Understood in this way, it can refer to the *what-it's-like-ness*, *first-person accessibility*, or *first-person self-evidence* (etc.) of states. These are all distinct properties, worth carefully distinguishing in the study of consciousness, but they can be considered together for the sake of the present discussion. On the other hand, 'phenomenal consciousness' might refer to the natural

[1] These various categories all have distinct logics of use, which it is usually bad practice to conflate. For example, psychological states are states of persons, whereas neural states or processes are typically understood as states of or processes in brains or organisms. My present discussion, however, requires no assumptions about which of these states can be said to be conscious. In the end, it is plausible that different senses of consciousness are appropriate to different cross-sections of these states. Although the term 'mental state' is normally a contentious one, I stipulate that for the present paper it is used in the above sense, as a relatively theory-free expression.

phenomena that exhibit the properties mentioned above. The expression in this second sense names not an abstract property, but designates states or processes that bear this property among others. I will refer to phenomenal consciousness in the first sense as *phenomenality*, and phenomenal consciousness in the second sense as *conscious states*, and I will refer to the cleavage between abstract properties and their instances as the *logical ambiguity* or the *logical distinction*.[2]

The logical distinction cannot do any empirical work, since it does not distinguish kinds; it merely distinguishes different aspects of a single property or phenomenon. It is simply the distinction between properties and the class of their bearers, as between roundness and round things, though it is easier to overlook for several reasons (perhaps because 'roundness' and 'round things' are not homonymous in English, or because *roundness* is rarely thought of as a cohesive and interesting phenomenon[3]). However, a term like 'homeostasis' bears the ambiguity: homeostasis can be understood either as an abstract property, or as a phenomenon concerning certain natural and artificial systems. That is, there is a logical distinction between homeostasis in the abstract and as homeostatic systems, just as there is a distinction between roundness and round things.

There are two semantic observations I would like to make. First, sentences featuring a property rather than its instances differ in their semantic properties. For example:

L1 '*Round things* can generally be rolled' is true.
L2 '*Roundness* can generally be rolled' is false or meaningless.
L3 '*Homeostatic systems* are self-regulating' is true.
L4 '*Homeostasis* is self-regulating' is false or meaningless.

Likewise:

L5 '*Conscious states* have qualitative properties' is true.
L6 '*Phenomenality* has qualitative properties' is false or meaningless.[4]

(L6) is not obviously problematic to many, but it is not true. Phenomenality is a property of states of persons, not a state of persons. We generally consider people to be phenomenally conscious, but one does not say of any conscious person that they are phenomenal. That the confusion is tempting here is a further reason to make this distinction explicit. Because sentences featuring these two senses of 'consciousness' are semantically distinguishable in this way, it is possible to equivocate over 'consciousness'.

My second semantic observation is somewhat more subtle. Phenomenality, I have said, is a property of some states, and conscious states are just the states that have that property. However, when one talks about conscious states one can make claims in

[2] The distinction is akin to Sellars' (1962) between a property and a distributive singular term.
[3] Perhaps by some geometers, or perhaps roundness with specific etiologies is considered an interesting phenomenon by e.g. geologists.
[4] Bear in mind that by 'conscious states' I mean phenomenally conscious states or processes, and by 'phenomenality' I mean the property in virtue of which some states or processes are phenomenally conscious. So the contrast between (L5) and (L6) cannot be that (L6) is more specific. Thanks to Julian Kiverstein for pressing me on this point.

which the phenomenality of the states features essentially, or claims in which the phenomenality of conscious states is accidental to the truth value of the claim. For instance, when one claims that round things make good wheels, the roundness of things is not incidental to the truth of the sentence. Things that are not round do not make good wheels, for just the reason that they are not round. However, when one claims that round things make good plates, the roundness is not essential. Things that aren't round make fine plates (if unconventional ones, at least in the West). Likewise for consciousness. Conscious states have many kinds of properties—perhaps physical or neurological properties, for example. In discussing consciousness inessentialism and the hard problem, we must distinguish claims that are true of conscious states in virtue of their phenomenality, and claims that are true of conscious states in virtue of their other properties.

3 The interface model

The first of the three cases I will discuss is a relatively straightforward case of equivocation over phenomenality and conscious states. Balleine and Dickinson (1998) are interested in consciousness inessentialism and the hard problem of consciousness. *Consciousness inessentialism* is the claim that consciousness is never essential to cognition or behavior (Flanagan 1992, 129). It is frequently associated with the view that consciousness could not have been the object of natural selection, that 'consciousness could... be a useless by-product of something that is useful' (Balleine and Dickinson 1998, 58). The *hard problem of consciousness* is of course the problem of explaining why there is such a property as phenomenality, and what its nature is (Chalmers 1995). The hard problem is distinguished from so-called *easy problems*, e.g. which neural phenomena correlate with reportability, or what the computational structure of visual edge-detection is. The easy problems are 'easy' because they can be investigated with known scientific methodologies, whereas when it comes to the hard problem we have no idea what kind of thing could count as solution at all, let alone what the correct solution is. Balleine and Dickinson aim both to provide evidence against consciousness inessentialism and to make progress on the hard problem. They claim, with respect to extant accounts of consciousness, that

> where these approaches have fallen is in their failure to suggest what advantage the *conscious* expression of these emotional states confers over and above their *unconscious* biological correlates. (Balleine and Dickinson 1998, 60; emphasis added)

That is, previous consciousness research falls short in that it does not explain the function of *phenomenality* rather than of conscious states, which of course have properties aside from phenomenality.

Balleine and Dickinson propose that consciousness is a functional adaptation to a natural problem faced by animals. Animals are capable of perception and action, which are governed by cognitive processes (1998, 60–61, 71). However, in order to act so as to realize their survival and reproductive goals, animals must have a way for physiological information (e.g. about digestive satiation) to become accessible to cognition (1998, 62–76). Their hypothesis is

> that the need to ground the goals of cognitively mediated actions in biologically relevant processes was solved by the evolution of an interface between these biological processes and cognition, namely consciousness. (1998, 76)

Balleine and Dickinson report two studies to support their claim. The studies provide evidence that food-seeking behavior in mice is more influenced by food-specific associations than by the general physiological state of the mice at the time of testing. One study manipulated hunger (Balleine 1992), and the other manipulated nausea using lithium chloride and ondansetron (Balleine et al. 1995). There are a number of methodological worries one might raise about these studies, but the objection that concerns me presently is that Balleine and Dickinson's methods do not successfully dissociate phenomenality from conscious states.

Balleine and Dickinson argue that the first study shows that mice require 'consummatory contact' with objects—e.g. they must eat food pellets rather than just see them—in order to establish behavior-guiding associations about the desirability of the pellets (Balleine 1992; Balleine and Dickinson 1998, 66–70). However, there is no evidence that the *phenomenal* properties of consummatory states contribute causally to the behavioral differences. The consumption of food has many non-phenomenal consequences neurologically and chemically that could be responsible for the observed change in behavior, either instead of or in addition to the phenomenal consequences of food-consumption. Balleine and Dickinson's results are consistent with the possibility that consciousness is, in their words, a 'useless by-product' of the learning mechanism.

In the second study, mice were given food pellets along with lithium chloride (LiCl), which makes them ill. Balleine and colleagues prevented the formation of negative associations in some mice by administering ondansetron, an anti-emetic chemical that counteracted the LiCl (Balleine et al. 1995; Balleine and Dickinson 1998, 78–80). Mice given ondansetron did not experience nausea, and did not form taste aversions.[5] However, ondansetron functions neurochemically, by blocking serotonin receptors and influencing the activity of the vagus nerve. The study confounds the manipulation of phenomenal properties of the mice's mental states with the manipulation of other (e.g. neurochemical) properties. Although both of Balleine and Dickinson's studies plausibly concern *conscious states* and behavior, they do not isolate the role that *phenomenality* plays in mediating behavior.

I am not arguing that Balleine and Dickinson's hypothesis about the function of consciousness is false, only that they fail to provide evidence against consciousness inessentialism, or to engage with the hard problem of consciousness. The hard problem of consciousness is precisely about understanding phenomenality independently of conscious states, but Balleine and Dickinson's evidence does not distinguish between truths about conscious states in which their phenomenality is essential, and truths about conscious states in which their phenomenality is incidental.

[5] More precisely, in Balleine et al. (1995) experiment 1, mice injected with ondansetron were less likely to form taste aversions than mice injected with a saline solution, and in experiment 2 mice injected with ondansetron exhibited the extinction of a conditioned behavior less precipitously than mice injected with saline.

4 Consciousness inessentialism and the function of consciousness

Stevan Harnad argues that since consciousness is not necessary for realizing any adaptive advantage, it has no function. Harnad's argument turns on observing the logical distinction, but does so too crudely. Harnad's argument is flawed in several respects, but as with Balleine and Dickinson's discussion I am not considering it in order to reject its conclusion. Rather, I review it as an illustration of the kind of problematic reasoning that is endorsed and discussed by some consciousness scientists.

Harnad expresses the core of his argument like this:

> Tell me whatever you think the adaptive advantage of doing something consciously is, including the internal, causal mechanism that generates the capacity to do it, and then explain to me how that advantage would be lost in doing exactly the same thing unconsciously, *with exactly the same causal mechanism*. (2002, 4)

That is, for any putative advantage of consciousness, Harnad claims to be able to undermine it by showing that the same advantage could be had without consciousness. Harnad illustrates this challenge by taking the case of pain. One might suppose, he offers, that phenomenal pain serves the function of detecting tissue damage and motivating its avoidance. However, if we had the same nervous mechanism, with the same neural and behavioral consequences, but we were to feel no pain when this occurred, then we would have precisely the same adaptive advantage. Therefore, pain cannot have the function of detecting tissue damage (Harnad 2002, 4–5). More formally, Harnad's argument can be reconstructed as a schema for indirect argument, along the following lines:

H1 Suppose a process X is conscious, and the adaptive advantage of X is to accomplish Y.

H2 For any X, there is another process X' such that X' also accomplishes Y [i.e. X and X' have identical causal mechanisms] and X' manifests without consciousness.

HC So accomplishing Y is not a function of consciousness, after all.

Of course, put this way Harnad's argument sketch is clearly elliptical, for these premises aren't quite sufficient to derive (HC). In order to make the argument valid, Harnad must rely on a tacit premise of the form:

H3 Accomplishing Y can only be a function of consciousness if there is no process X', such that X' accomplishes Y and X' manifests without consciousness.

(H2) and (H3) always entail a conclusion (HC) that contradicts the supposition (H1).[6] However, both (H2) and (H3) are false. (H2) presumes that if two processes accomplish

[6] Of course, since the supposition is one term of the contradiction, the strategy of indirect proof is not necessary; the argument works just as well without the initial premise. However, this structure is somewhat truer to Harnad's prose expression of the argument, and allows the introduction of the terms rendered here as X and Y.

the same effect, they have the same causal mechanism, and (H3) expresses the principle that if there is a non-conscious version of a process, then the phenomenality of one instance of the process cannot make a causal difference to its function. Block (1980) has articulated cases that serve as counterexamples to both claims.

Block's counterexamples are directed primarily against the principle expressed by (H3). The thrust of Block's argument is that there may be properties or components of a functional system[7] that are ineliminable in some implementations of the system, but eliminable in others. For example, suppose[8] there is a hydraulic computer that is functionally identical to my electronic calculator. The properties of the fluid in the hydraulic computer are an ineliminable component of its functioning, although a functionally identical machine—the calculator—may fulfill the same functions with no fluid at all.

> The hydraulic device will presumably have states whose causal consequences depend crucially on the properties of the fluid, *even though* there could be functionally identical states that did not depend on any properties of fluid (but on electrical properties instead). (Block 1980, 262)

Similarly, electrical properties of the hydraulic computer may play no role in its functioning, although they are essential to proper functioning of the electronic calculator. Block's suggestion is that consciousness may be analogous to the fluid or the fluid pressure in the hydraulic computer. Harnad may claim, rightly, that any function performed consciously in humans could in principle be performed unconsciously. However, that claim does not imply that any function performed consciously in humans could in principle be performed unconsciously *in humans*, or that such functions could be performed unconsciously *through the same causal mechanisms*, since the causal implementation of those functions in humans may depend crucially on consciousness although it does not depend on consciousness in every implementation. So, contra (H3), it is not the case that if something is a function of consciousness, there must be no way of implementing the system that leaves consciousness out.

Now consider (H2). Harnad slides between the claim that two processes engender the same adaptive advantage and the claim that two processes have the same causal mechanism. If he further assumes (and he does) that two such processes can differ with respect to whether they are conscious, he has begged the question in favor of consciousness inessentialism. That is, he has ignored the possibility, articulated by Block, that even though an adaptive advantage may be achieved by an unconscious process in

[7] I adopt functionalist language here, but this should not create controversy. Functionalism, as a controversial thesis in the philosophy of mind, is the claim that mental phenomena can be exhaustively characterized in functional terms. It is much less controversial that functionalist apparatus may be used to describe various aspects of phenomena, perhaps not exhaustively. Moreover, the authors I am considering are all tolerant of functional characterizations. Block assumes the cogency of functionalist language in his discussion. Harnad endorses an implausibly strong functionalism, which is vulnerable to Block's objections and on which indistinguishability in a Turing test implies indistinguishability with respect to adaptive fitness. Gray has objections to what he calls 'functionalism' (2004, 123–147), but the view to which he objects is an extremely strong thesis with unclear relations to more traditional functionalisms (cf. 2004, 132–133: traditional functionalists are committed only to what Gray calls the 'primary inference', not the 'complementary inference' or the third, 'further' inference. None of Gray's objections are addressed to the primary inference).

[8] I take some liberties with Block's examples.

some causal implementation of a functional system, consciousness may be essential to a different causal implementation of a functionally identical system. For instance, although a computer can solve quadratic equations without conscious attention, that does not mean that humans can do so. Harnad has not shown, merely because such computers can exist, that the phenomenality of attention is inessential to the way humans perform such tasks.

5 Cardiac periodicity as a test phenomenon

These objections to Harnad's argument will be made clearer if they are illustrated on an argument parallel to Harnad's, but about something better understood than consciousness. Consider the periodicity of the mammalian heartbeat. Since we understand the nature of cardiac periodicity much better than we understand consciousness, consideration of this parallel argument may reveal semantic anomalies of Harnad's argument. If the structure of Harnad's argument can be used to suggest that cardiac periodicity is a mysterious phenomenon, or that it confers no adaptive advantage on certain organisms (like mammals), then we should see that extension as a reductio of the structure of Harnad's argument. Consider Harnad's argument again, but with selective substitutions:

PH1	Suppose that the mammalian heartbeat is periodic, and the adaptive advantage of the heartbeat is to accomplish the circulation of blood.
PH2	There is another process *X′* such that *X′* also accomplishes circulation [i.e. *X′* and the mammalian heartbeat have identical causal mechanisms] and *X′* manifests without cardiac periodicity.
PH3	Accomplishing the circulation of blood can only be a function of cardiac periodicity if there is no process *X′*, such that *X′* accomplishes the circulation of blood and *X′* is not periodic.
PHC	So promoting circulation is not a function of cardiac periodicity, after all.

As with Harnad's argument, the second and third premises are false. Mammalian hearts do have the function of circulating blood, and their periodic rhythm is essential to the way they accomplish that function. There are indeed kinds of pumps, e.g. Archimedes' screw pumps, that function continuously instead of periodically, but contra (PH2) the fact that there are such pumps does not imply that pumping can be achieved without periodicity *and* by the same means as the mammalian heart. Periodicity is an essential property of normal mammalian cardiac function. Furthermore, contra (PH3), the possibility of continuous pumping processes does not entail that blood circulation is not the function of cardiac periodicity.

However, consideration of this parallel argument does reveal why it might be tempting to make the claims that Harnad does. There is something semantically odd about the claim that the function of periodicity is to circulate blood. It is awkward to say that the *periodicity itself* circulates blood, although it is natural and true to claim that in the mammalian heart *a periodic process* circulates blood. That is, whereas a statement featuring the first term in the logical ambiguity about periodicity has a controversial semantic value, a statement featuring the second term is clearly true. I profess no opinion on whether it is true or false or meaningless that 'periodicity itself circulates blood'. What I do insist upon is that many

fluent speakers of English will resist endorsing such a statement. If we take the periodicity case as a model, then it seems that it would be possible to produce subjective puzzlement by asking 'what is the function of phenomenality?' even if there were no hard problem.

To be clear, I am not claiming that consciousness is not puzzling. There are properties of functional systems that are neither essential for accomplishing a function in general, nor essential for a particular implementation of a function. Consider another example from Block: the coloration of wire insulation inside a computer.

> The wires are (say) red, though since their color makes no difference to the working of the computer, they could have been some other color without making any functional difference. [...] The color is 'epiphenomenal.' (Block 1980, 262–263)

The truth in the vicinity of Harnad's view is that we do not know what kind of role is played by consciousness. For all we know, consciousness may not be like periodicity in the heart or fluid pressure in the hydraulic computer, but like the color of wire insulation in a computer. Nevertheless, Harnad's argument fails to establish that this is the case.

Consideration of the flaws in Harnad's argument should narrow our focus if we are interested in epiphenomenalism or the hard problem. Harnad is correct that establishing a function for conscious states is insufficient to establish a function for phenomenality. That red wires play a role in the functioning of my computer does not entail that the redness of the wires plays a role. However, it is also improper to demand that in order for phenomenality to play a function, it must do so independently of the other properties of conscious states and the particular system in which they function. It is possible to circulate blood with a pump that lacks a periodic character, but that does not mean that cardiac periodicity is epiphenomenal to mammalian circulation. Rather, if we want to evaluate consciousness inessentialism, we must ask what role conscious states manage to play for us in virtue of their phenomenality.

6 Gray's consciousness inessentialism

Despite the flaws of Harnad's argument, Jeffrey Alan Gray is impressed and troubled by it. He presents a version of the argument in order to motivate his own discussion. Gray does not quite endorse the conclusion of the argument, but takes the argument to suggest we are faced with a dilemma between anti-naturalist epiphenomenalism and a methodological overhaul of the natural sciences (2004, 73–74).[9] Gray manages to avoid some Harnad's errors. Gray's version of the argument has a distinct structure, and although it ultimately has faults similar to Harnad's version, it makes the role of the logical equivocation clearer. However, although Gray is more careful than Harnad in his discussion of Harnad's argument about consciousness inessentialism, when he discusses his tentative solution to the hard problem he commits the logical equivocation more clearly than Harnad does.

[9] I qualify Gray's support for the argument and the dilemma because the lacunae in his discussion raise interpretive puzzles, about which I will say more below.

Gray's argument, reconstructed from excerpts of his discussion, goes something like this:

G0 'Survival value is enhanced if a characteristic increases the chances of an individual's (1) staying alive and/or (2) producing more offspring.'
G1 'The increase in survival value [conferred by consciousness] requires changes in behaviour.'
G2 'These changes (argues Harnad) can be fully accounted for by the brain processes and their output in behaviour.'
GC 'Therefore, the accompanying conscious experiences do not contribute *in their own right* to the enhanced survival value. They just come along for the ride.' (2004, 72)

This argument follows the structure of Gray's prose, but requires some explanation and tidying up to look good with numbered premises. (G0) is supposed to be a consequence of the theory of natural selection. One could object to the details, but I'll accept the spirit of it here. (G1) is plausibly true, based on (G0) and some extra premises. That is, one might try to make a naïve distinction between traits that *directly* affect an organism's interaction with its environment, and those that do so *indirectly*. It is not clear to me that there is a principled distinction to uphold in this vicinity, but the essential point is that the brain, as an internal organ, does not directly affect an organism's interactions with its environment in the way that e.g. coloration serves to camouflage or to signal, or claws serve as natural weapons or tools, etc. Rather, the main relevance of the brain to an organism's survival value is manifested in the ways it modulates the organism's direct relations with its environment. Such a justification of (G1) is, as I said, naïve, but something in the vicinity is plausibly correct and it can be accepted for the present discussion. (G2) is where the heavy lifting occurs in Gray's argument. We can understand it to be motivated by Harnad's argument, and in particular by Harnad's (H2) and (H3), which I argued earlier are false. However, (G2) is somewhat different from any of the statements in Harnad's argument, and Gray's formulation of Harnad's argument is worth making explicit because he seems to be more sensitive than Harnad to the logical ambiguity.

Tidied up for formal presentation, Gray's argument goes like this:

G*1 In order to enhance survival value, consciousness must have behavioral consequences.
G*2 Consciousness (independently of brain processes) has no behavioral consequences.
G*C Therefore, consciousness (in its own right) does not enhance survival value.

We can grant (G*1) for our present purposes. (G*2) and (G*C), however, bear interesting qualifications—consciousness, considered independently of 'brain processes', and consciousness 'in its own right'. There are at least two ways to interpret these qualifications. We may read them both as free expressions of the same qualification on consciousness, in which case the argument is valid as it stands, or we may interpret them as expressing different qualifications, in which case there must be a tacit premise or principle expressing the relation between independence from brain processes and

own-right enhancement of survival value. I think it is plausible that Gray thought of the qualifications as distinct. Given Gray's enthusiasm for Harnad's argument, he might follow Harnad in thinking that in order for phenomenality to contribute to survival value (the way cardiac periodicity does), phenomenality must have behavioral consequences that brain states without phenomenality do not have. However, as I argued above, this is an improper demand. Consider the case of periodicity again: periodicity does not contribute to circulation *independently of cardiac processes*. Rather, certain cardiac processes (involving e.g. coordinated rhythmic contraction of cardiac muscle) give rise to cardiac periodicity while accomplishing their function. Since Gray's argument is less interesting if read this way, I prefer for the moment to read 'independently of brain processes' and 'in its own right' as expressing the same qualification. That is, both qualifications indicate that consciousness is to be understood, for the purposes of the argument, *in its own right*.

If we read Gray this way, it seems that what he is being sensitive to when he qualifies consciousness in this way is precisely the logical ambiguity. His qualifications seem to encourage us *not* to understand his argument like this:

G*2′ Conscious states have no behavioral consequences.
G*C′ Conscious states do not enhance survival value.

After all, these statements are probably false. Conscious states, which have both phenomenal and other kinds of properties, plausibly including causal ones, probably *do* have behavioral consequences. If that is true then it is plausible that conscious states do somehow enhance survival value. But Gray seems to phrase his argument so as not to deny these facts. What he seems to mean is this:

G*2″ Phenomenality has no behavioral consequences.
G*C″ Phenomenality does not enhance survival value.

Since we do not want to succumb to the semantic and metaphysical troubles that can confuse discussion of Harnad's version of the argument, it is worth stipulating that if (G*2″) and (G*C″) are true, they are true for humans and some other organisms in the way that the statements

PG2 Cardiac periodicity has physiological consequences.

and

PGC Cardiac periodicity enhances survival value.

are true for mammals and some other organisms. That is, whether (G*2″) and (G*C″) are true should depend upon facts about the role played by phenomenality in the functioning of humans and other organisms, not upon the metaphysics of properties. Of course, it is not clear whether (G*2″) and (G*C″) are in fact true or not. As I see it, one cannot become entitled to the conclusion (G*C″) by virtue of simply assuming the equally controversial (G*2″), and no remarks by either Harnad or Gray successfully motivate either statement. Gray curiously does not make it clear what he thinks of this

argument. On the one hand, Gray takes this argument very seriously; he uses it to structure his search for a solution to the hard problem for the remainder of his book. On the other hand, he resists the argument's conclusion. He does not believe that it has been shown to be impossible that there might be a selected function for phenomenality, and he proposes hypotheses about what those functions might be. However, on the first hand again, Gray articulates no explicit objections to the argument.

What was true in the vicinity of Harnad's argument is still true here. Concerns about natural selection aside, an account of the function of phenomenality, as distinct from the function of conscious states, must give us reasons to think of consciousness on the analogy of periodicity rather than wire coloration. It must specify a role that conscious states play *in virtue of their phenomenality* and not merely in virtue of their other properties. Importantly, Gray does not need to find a role that phenomenality *necessarily* plays, or even one that phenomenality actually plays. Gray claims only to be 'creeping up on the hard problem,' not solving it. What Gray requires in order to accomplish that goal is merely a promising hypothesis. He needs to identify a particular role that could *possibly* be played by the phenomenality of mental states. However, even this is a tall order, and although Gray would seem to be sensitive to the logical ambiguity here, this sensitivity gets muddled later on. Gray's positive suggestions about the role of conscious states either concern conscious states independently of their phenomenality, or they do not make it clear how the phenomenality of those states is supposed to accomplish the function he suggests.

7 Consciousness as a medium of display

Gray's suggestion comes in two stages. On the first pass, Gray suggests that the adaptive advantage conferred by consciousness is that of 'late error detection', but acknowledges that the advantage is secured through various causal properties of conscious states, not the phenomenality of those states. On his second pass, Gray tries to single out the specific role that phenomenality plays in the mechanisms that serve perception and late error detection. Gray's proposal is that phenomenal consciousness serves as a 'medium of display'. However, he fails to make it clear how phenomenality contributes to the accomplishment of this role. I will discuss these two suggestions in turn.

Gray claims that the adaptive function of consciousness is as a part of 'late error detection', the intrinsic reinforcement or punishment structure that results in the promotion or inhibition of behaviors. Gray insists that the function of consciousness must be 'late' because he is impressed by claims like Libet's (1983) that consciousness it too slow to co-ordinate our actions on-line.[10] Gray gives the general idea of late error detection by illustrating how it would work in a case of pain:

> Your hand goes too close to the flame, you withdraw it (unconsciously), *you then feel the pain and, in consequence, you review the action* that just led you to approach the flame too close and too incautiously. (2004, 76)

[10] But for a strong summary of reasons not to take the Libet studies seriously, see e.g. Flanagan (1992, 136–138).

Of course, this is a function that could be accomplished by states that are only incidentally conscious. Gray has given us no reason to think that the phenomenality of the states plays an essential role in the implementation of the function, and Gray confesses this shortcoming. To stop here in a solution to the hard problem, he admits, would be to 'slip it back under the carpet':

> But, when we try this on, there is likely to be a Stevan Harnad who will—quite rightly—drag it out again. For we don't at present understand how conscious experience, whether or not seemingly equipped with the right kind of survival value, can have causal effects in its own right, as distinct from those of the brain processes it accompanies. (2004, 76–77)

Gray observes that we don't know enough about the phenomenon of consciousness to be able to say with confidence how phenomenality could play a role in survival or cognition, and this is precisely the difficulty of the hard problem. It is unclear how late error detection is a role that overcomes the epiphenomenalist worries with which Gray frames his discussion. What Gray requires is a role that can be served by the phenomenality of conscious states, rather than the other features of conscious states.

Gray improves upon this initial suggestion by trying to articulate such a role. His grand suggestion is that 'consciousness acts as a medium of display' (2004, 108). Consciousness contributes to late error detection by enabling 'the juxtaposition and comparison of variables controlled by different unconscious servomechanisms, especially when these are in different sensory modalities' (2004, 104–105). That is, phenomenality somehow enables there to be a general (not modality-specific) representation—the 'display'—that is an essential component of the implementation of late error detection in humans and related organisms. The suggestion is akin to Fodor's that there is a 'central system' in human cognition (Fodor 1985). Gray illustrates his suggestion by means of an analogy:

> Suppose I am in St Marks Square in Venice and have sufficient artistic talent... to make a passable sketch of it. Later, I use the sketch as an aid to recall St Marks. Thus the sketch expands my capacity to remember—a causal effect. But the sketch clearly doesn't have this causal effect in its own right. The sketch is made by the brain... and later used by the brain. All the causal mechanisms lie in the brain. Still, any full description of the causal chain that leads to my recall of St Marks, must include an account of the role played by the sketch. (2004, 109)

Gray's suggestion is that consciousness' role in cognition is analogous to the role played by the sketch in this instance of recall. Since Gray finds dualism unacceptable, the display cannot be an immaterial analogue of the sketch. Rather, it must be implemented by brain states.[11]

[11] See e.g.: 'And this percept is in my brain, made up of activity in whatever circuits have been activated in its construction, in the visual cortex and the other parts of the brain with which the visual cortex interacts' (Gray 2004, 109).

> Like my sketch, the visual percept of St Marks is constructed by my (unconscious) brain. Like the sketch, the percept expands the powers of my brain... Just how does conscious vision exert these causal effects? The sketch analogy suggests that conscious perception, like the sketch of St Marks, deploys no causal mechanisms of its own. The percept is there, like the sketch, merely as a display, one used by the unconscious brain. (2004, 109)

Gray's view seems to be that the display overcomes Harnad's concerns because it 'expands the powers of the brain'—it has its own 'causal effects'—but is naturalistically acceptable because it involves no causal mechanisms apart from neurochemical mechanisms (2004, 109–110). In light of this picture of the role of phenomenal consciousness, Gray recasts the hard question as the question of 'how does the unconscious brain create and inspect the display medium (qualia) of conscious perception?' (2004, 121).

However, this is a very mysterious suggestion. Gray envisions the display as an amodal representation that enables both information-transfer between modality-specific processes and comparison of perceptions with expectations, and he suggests that unconscious mechanisms in the brain interact with the display by generating and examining it. Assuming for the sake of argument that this is in fact how the brain works, I fail to see how Gray thinks *phenomenality* is involved in display's accomplishment of its functions. One possibility is that the display is the result of the second transduction in what Dennett calls 'double transduction'. That is, sensory organs transduce physical stimuli into neural signals, and at some point in the brain the neural signals are transduced into qualia that bear special, super-representational properties that resemble their representanda in some privileged way. But belief in double transduction of this sort is generally agreed to be an error (Dennett 1998, 2018).[12] The functions that Gray suggests are accomplished by the display all concern information storage and manipulation, and can be accomplished by information-processing and computational features of mental states. The objection I am raising is *not* Harnad's argument, which holds that since there could be unconscious processes that function like the display does, consciousness can play no role in the display's function (although the antecedent is true, and if Gray is really worried by Harnad's argument he should be unsatisfied with his own suggestion). Rather, my objection is that the function Gray ascribes to consciousness seems to be, even in the human case, accomplished *through the information-processing and computational properties* of conscious states, and not through their phenomenality. Unless Gray has an account of how phenomenality plays a role in the realization information-processing capacities in humans, there is no solution to the hard problem in the offing.

A further line of argument Gray pursues against the epiphenomenalist is that the display medium of consciousness plays an essential role in the way language-acquisition, aesthetic appreciation and scientific inquiry are accomplished in humans. However, even if this were to be true, it would only be the display *as conscious states* that plays a role. Its information-processing properties and its causal connections are what explain the capacities to share reference, feel perceptual pleasure, or construct theories.

[12] Nevertheless, the double transduction view is probably consistent with Uriah Kriegel's (2005) interpretation of Gray as a dualist with a philosophically unconventional lexicon.

So although Gray's version of Harnad's argument was canny and avoided equivocating over phenomenality and conscious states, and although he recognized that his 'late error detection' function explained the role played merely by conscious states and not by their phenomenality, his ultimate suggestion also falls short. The functions accomplished by the display may involve conscious states in humans and related organisms, but Gray does not explain how phenomenality accomplishes those functions. Even if Gray's display hypothesis is correct, he offers no convincing argument that the display achieves its function in virtue of the phenomenality of the states that implement it.

8 Conclusion

I have discussed three cases in which consciousness researchers interested in the hard problem and consciousness inessentialism make critical errors about the logical distinction between properties and their instances. Balleine and Dickinson claim to contribute to a solution of the hard problem, to have isolated the role of phenomenality, distinct from other properties of conscious states. However, their manipulations do not separate the role of phenomenality from the role of e.g. serotonin receptor blocking. So they are entitled to claim only that they have investigated the role of conscious states, not of phenomenality. Harnad improves on Balleine and Dickinson by observing that finding a role for conscious states is not equivalent to finding a role for the phenomenality of conscious states, however his argument is also flawed. The most significant flaw is probably his failure to acknowledge cases like those that Block points out, which involve properties that are essential to some implementations of a function but not others. Gray, in his presentation of Harnad's argument, may avoid the error to which Block objects. He identifies, correctly, that in order to make progress on the hard problem he needs to explain not the role of conscious states (as Balleine and Dickinson attempt), or the role of phenomenality in abstraction from a particular system (as Harnad demands), but the role of the phenomenality of conscious states in a particular system. However, his positive suggestion succumbs to the error that Balleine and Dickinson commit, and does not explain the contribution of phenomenality to the function of conscious states.

Being careful to observe the logical distinction and to avoid related errors can help us avoid the argumentative troubles in these cases. I have tried to promote clarity in this paper by marshaling examples like Block's, and using cardiac periodicity as a test for equivocation about the logical distinction. While I have focused on authors that run afoul of the logical distinction, there others who do not. For example, 'mental paint' is ex hypothesi a phenomenal property, distinct from representational and other properties of states (Harman 1990; Block 2010). I am not sure whether to be convinced of Block's (2010) defense of mental paint, but it seems to be a notion cleverly designed to avoid the errors I describe above. However, merely observing the distinction does not bring us any closer to positive knowledge about consciousness. At best, it allows us to put the question at the heart of the hard problem more clearly, and to distinguish it from similar but distinct questions. There are at least four distinct explananda in the vicinity of these discussions. First, it is one thing to know or ask about which of the mental states are the conscious ones, the ones that bear the property of phenomenality. Second, it is a

different thing to know or ask what (say, adaptive or cognitive) role the conscious states play. Third, it is yet another thing to know or ask what it is in virtue of which some states have the property of phenomenality, and what having phenomenality consists in. Finally, it is still another thing to know or ask what role is played by the *phenomenality* of conscious states, as opposed to their other properties. Only the third and fourth explananda concern the hard problem of consciousness, and consciousness inessentialism is the claim that the fourth explanandum admits of no positive answer.

It is not clear to me, from the fact that we have no clear idea how phenomenality could contribute in its own right to fitness or function, that it could not. For my part I do not see what adaptive consequences could follow from the phenomenality of a state, or from there being a structured system of such states. However, to argue from such an autobiographical premise to the conclusion that there cannot therefore be such consequences is an execrable argument from lack of imagination. While such arguments have their place in human life, on good days I do not endorse such arguments in philosophical or scientific inquiry.

Acknowledgments I am grateful to the members of Mazviita Chirimuuta's 2011 seminar on 'A Science of Consciousness' for feedback on these ideas, in particular that of Mazviita Chirimuuta, Trey Boone, Joseph McCaffrey, and Lisa Lederer.

Compliance with ethical standards I declare that neither I nor any member of my immediate family have any affiliation with or involvement in any organization or entity with a financial or non-financial interest in the subject matter of this manuscript.

I have not fabricated or falsified any data. The research represented in this manuscript poses no threat to national security or public safety.

References

Balleine, B. W. (1992). The role of incentive learning in instrumental performance following shifts in primary motivation. *Journal of Experimental Psychology: Animal Behavior Processes, 18*, 236–250.

Balleine, B. W., & Dickinson, A. (1998). Consciousness–the interface between affect and cognition. In J. Cornwell (Ed.), *Consciousness and human identity* (pp. 57–85). Oxford: Oxford University Press.

Balleine, B. W., Garner, C., & Dickinson, A. (1995). Instrumental outcome devaluation is attenuated by ondansetron. *Quarterly Journal of Experimental Psychology, 48B*, 235–251.

Block, N. (1980). Are absent qualia possible? *Philosophical Review, 89*, 257–274.

Block, N. (1995). On a confusion about a function of consciousness. *Behavioral and Brain Sciences, 18*, 227–287.

Block, N. (2010). Attention and mental paint. *Philosophical Issues, 20*, 23–63.

Chalmers, D. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies, 2*, 200–219.

Dennett, D. C. (1998). The myth of double transduction. In S. R. Hameroff, A. W. Kaszniak, & A. C. Scott (Eds.), *Toward a science of consciousness II: The second Tucson discussions and debates* (pp. 97–107). Cambridge: MIT Press.

Dennett, D. C. (2018). Facing up to the hard question of consciousness. *Philosophical Transactions of the Royal Society B, 373*, 20170342. https://doi.org/10.1098/rstb.2017.0342.

Flanagan, O. (1992). *Consciousness reconsidered*. Cambridge: MIT Press.

Fodor, J. A. (1985). *Modularity of mind*. Cambridge: MIT Press.

Gray, J. A. (2004). *Consciousness: Creeping up on the hard problem*. Oxford: Oxford University Press.

Harman, G. (1990). The intrinsic quality of experience. *Philosophical Perspectives, 4*, 31–52.

Harnad, S. (2002). Turing indistinguishability and the blind watchmaker. In J. H. Fetzer (Ed.), *Consciousness evolving: Advances in consciousness research* (pp. 3–18). Amsterdam: John Benjamins.

Irvine, E. (2013). Measures of consciousness. *Philosophy Compass, 8*, 285–297.
Kriegel, U. (2005). Review of *Consciousness: Creeping up on the Hard Problem*, by Jeffrey Alan Gray. *Mind, 114*, 417–421.
Libet, B., Gleason, C. A., Wright, E. W., & Pearl, D. K. (1983). Time of conscious intention to act in relation to onset of cerebral activity (readiness potential): The unconscious initiation of a freely voluntary act. *Brain, 106*, 623–642.
Sellars, W. (1962). Naming and saying. *Philosophy of Science, 29*, 7–26.