



# Phenomenal consciousness and moral status: taking the moral option

Joseph Gough

To cite this article: Joseph Gough (2025) Phenomenal consciousness and moral status: taking the moral option, *Philosophical Psychology*, 38:6, 2489-2516, DOI: [10.1080/09515089.2024.2339353](https://doi.org/10.1080/09515089.2024.2339353)

To link to this article: <https://doi.org/10.1080/09515089.2024.2339353>



© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 08 Apr 2024.



Submit your article to this journal [↗](#)



Article views: 874



View related articles [↗](#)



View Crossmark data [↗](#)



# Phenomenal consciousness and moral status: taking the moral option

Joseph Gough 

Philosophy, University of Oxford, Oxford, Oxfordshire

## ABSTRACT

Intuitively, there is a close link between moral status and phenomenal consciousness. Taking the link seriously can serve as the basis of a proposal that appears to have a surprising number of theoretical benefits. This proposal is the moral option, according to which moral status is partly determinative of phenomenal consciousness, and phenomenal consciousness is sufficient for possession of a moral property I refer to as “moral status.” I argue for this view on the basis of its ability to shed light on the distribution problem of determining which systems are phenomenally conscious. I explicate it by drawing on a theory of attributions of phenomenal consciousness, and considering its relationship to positions on the metaphysics of consciousness. I defend it against a series of objections, including three based on previous theories of the link between phenomenal consciousness and moral status. The moral option should, in my view, be a serious contender among theories of phenomenal consciousness.

## ARTICLE HISTORY

Received 26 February 2021  
Accepted 1 April 2024

## KEYWORDS

Phenomenal consciousness;  
phenomenal stance;  
empathy; moral status;  
distribution problem

## 1. Introduction

This paper argues that the presence or absence of phenomenal system-consciousness, the property of a system such that there is “something it is like” to be that system, is partly determined by the presence or absence of a moral property which I cautiously refer to as “moral status” – I clarify my reasons for caution in response to an objection in §3.3. I call this position “the moral option”, the view that there is some significant moral property partly determinative of phenomenal consciousness.

Such a view requires that there is a strong link between moral status and phenomenal system-consciousness. In particular it requires that that moral status is *necessary* for phenomenal consciousness, and thereby phenomenal consciousness is *sufficient* for moral status – a contentious claim that I will defend further below. It does not require, and I do not defend, the claim that

---

**CONTACT** Joseph Gough  [Joe.gough@philosophy.ox.ac.uk](mailto:Joe.gough@philosophy.ox.ac.uk)  Philosophy, University of Oxford, Oxford, Oxfordshire

© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.  
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

that moral status is *sufficient* for phenomenal consciousness, nor therefore, the claim phenomenal consciousness is *necessary* for moral status (see also Bradford, 2023; Shepherd, 2023, 2024).

In support of the moral option, I argue that explain facts about whether systems are phenomenally conscious or not, we will need to draw directly on moral considerations. This is an issue that arises frequently in scientific work that speaks to the “distribution problem”, the problem of determining which systems possess phenomenal consciousness. Different groups of scientists use different criteria, and criteria are frequently debated and changed; often, clashes between different criteria appear to reach an impasse. These debates, changes, and impasses can look groundless when moral considerations are left out of the picture, but their rationale becomes clear and their apparent tractability drastically improves when moral considerations are drawn on directly.

The moral option does not fit into Carruthers (2000) influential taxonomy of positions on phenomenal consciousness. The first division Carruthers considers is between theories where phenomenal consciousness is explicable, and theories where it is not. All those theories where phenomenal consciousness is explicable, on his taxonomy, are naturalistic theories, which are further divided into neurological vs cognitive/functional (the latter contains many more subdivisions, including nonrepresentational vs representational, first-order vs higher-order, and so on). The moral option is not a purely cognitive/functional theory, nor a purely neurological theory of consciousness. However, it does view phenomenal consciousness as wholly explicable, an explanation I develop over §2, and expand upon across a series of objections and replies in §3.<sup>1</sup>

## 2. The moral option

The moral option, in its most basic form, is the view that there is some moral property that is partly determinative of phenomenal consciousness. This leaves many open questions. My aim in this section is to elaborate this core claim into a theory of phenomenal consciousness. First, I lay out some necessary technical machinery. In particular, I look at two distinctions, one between two kinds of determinative relation that might hold between facts known as “anchoring” and “grounding”, and one between the phenomenal consciousness of systems (henceforth, “system-consciousness”) and the phenomenal consciousness of states (henceforth, “state-consciousness”). I then offer my core argument for the theory, which draws on the realities of the debate over the “distribution problem”, the problem of determining which systems are or are not conscious.

Having laid out the core argument for the theory, I explain a theory of consciousness attributions known as the “phenomenal stance” in the interests of intellectual honesty (since this theory, despite offering only weak

support for the moral option, played a significant role in inspiring it) and clarification of this theory of consciousness (§2.2). I then go on to consider the role and status of non-moral components of phenomenal consciousness (§2.3), and the position of the moral option in debates over the metaphysics of phenomenal consciousness (§2.4).

## 2.1. *The moral option and the distribution problem*

### 2.1.1. *Kinds of determination*

Epstein (2015, 2019) defends a distinction between two kinds of determinative relation that might hold between facts, anchoring and grounding. “Anchors” are facts that fix the boundaries of a kind or category; “grounds” are facts about a particular individual that determine whether that individual is a member of that category. For example, one might imagine that an object’s *being a teapot* is grounded entirely in facts about the shape, size, heat-resistance, etc, of an object, *but* that a different set of facts, bearing on human practices around tea, intentions, and health-requirements, fix the physical requirements for an object to be a teapot. A particular teapot, on such a view, is a teapot because it meets a set of criteria specified in entirely physical terms, but these criteria are determined largely by facts about human beings and how they make tea.

The moral option, as I have explained it, is neutral between whether moral facts are anchors or grounds for system-consciousness. It is entirely consistent with the moral option to claim that phenomenal consciousness is morally-grounded: that certain non-moral facts about a system and its states, in conjunction with the fact that it has moral status, ground the fact that it is phenomenally conscious.

However, the moral option does not need to grant the view that moral facts ground facts about phenomenal consciousness. It is equally well secured by the claim that moral facts anchor facts about phenomenal consciousness. On such a view, the correct criteria for judging whether or not a system is phenomenally conscious are entirely non-moral. Even so, this version of the moral option claims, moral facts are part of how these criteria are determined in the first place.

It could be a necessary truth that all phenomenally conscious systems have moral status, even if moral facts are anchors, but not grounds, for facts about whether or not a system is phenomenally conscious. Such a view is likely to require that there is a hugely complex disjunctive physical property as part of the grounds of phenomenal consciousness. This physical property would be a disjunction of all the configurations of physical properties that might bestow a system with moral status. This disjunction would be a morally-anchored category, and hence phenomenal consciousness would be a morally-anchored category,

even though neither would be morally-grounded. The grounds of phenomenal consciousness would be specifiable in entirely non-moral terms.

I wish to keep my defense of the moral option neutral between moral anchors and moral grounds, at least for the time being. In §2.1.3, I argue that the current state of the scientific debate over the distribution problem reflects an implicit agreement that it is the potentially morally-salient features of creatures that are candidates for system-consciousness, and of their states, which should be used to judge whether or not they are phenomenally conscious. This is consistent with either moral anchors or moral grounds: the way these debates are held does not distinguish between whether different parties in these debates view the relevant moral features are grounds for the phenomenal consciousness of particular systems, or anchors of the category as a whole.<sup>2</sup>

### 2.1.2. *Phenomenal consciousness: State and system*

An important distinction for clarifying the moral option, for clarifying my argument in the next section, and for preempting an objection, is between state-consciousness and system-consciousness. Quite plausibly, it makes no sense to think of a phenomenally conscious system with no capacity for phenomenally conscious states. As such, one might think that for a system to be phenomenally conscious is just for it to have the capacity for phenomenally conscious states (cf., eg, Bayne, 2007; Manson, 2000; McBride, 1999). The way to investigate system-consciousness, on this view, is to start by explaining state-consciousness: first, one should work out why, in (eg.) a human being, some states are conscious while others are not; then, for any given system, one can simply check whether it is the sort of system whose states could have whatever feature makes some states conscious in humans (see also §3.4).

I am happy to grant that it makes no sense to imagine a system with system-consciousness but no capacity for state-conscious states. Let us also imagine that every state-conscious state of a human being shares some feature, F (for example, widespread influence across neural systems, or being poised for control of moral behavior). Even if this were true, it would be fallacious to conclude that looking for instances of F is all one needs to do to look for system-consciousness and state-consciousness in other systems than human beings. The problem with this line of reasoning is that even if F explains why a state is conscious *when it is a state of a system-conscious system*, there may still be further independent constraints on being a system-conscious system in the first place (saliently, for example, moral ones). Imagine again that state-consciousness in a phenomenally conscious system can be identified with *being poised for control of molar behaviour*; this would *not* mean that, eg, a computer with states poised for control of its

overall behavior was phenomenally conscious, or that its states were phenomenally conscious.

The capacity for states with F, whatever F turns out to be, may be a necessary condition on phenomenal system-consciousness. However, *being a state of a phenomenally conscious system* may also be a non-trivial necessary condition on *being state-conscious*. That is, there may be some systemic property of the overall system, G, that is also required. So it may well be that for phenomenal consciousness to arise, you need the co-instantiation of a capacity for states with F, and systemic feature G – that is for a system to be phenomenally conscious, it may need systemic property G and the capacity for states with F, while for a state to be phenomenally conscious, it may need feature F, and to be instantiated in a system with G. Looking at the features that make a difference between conscious and nonconscious states in humans may reveal F to us, but it will not reveal G, the systemic features that allow states with F to be phenomenally state-conscious.

As such, I am happy to grant that there may be some feature of a system's states, F, which makes those states conscious, and explains why they and not its others states are conscious. I am also happy to grant that facts about which states possess F may be neither morally-grounded nor morally-anchored. These claims are consistent with the moral option. The reason for this is that even if the capacity for states with non-moral feature F partially grounds whether or not a system possesses phenomenal consciousness, this does not rule out the possibility that there are also moral grounds, nor the possibility that there are moral anchors.

This preempts an objection to the moral option: there appears to be no moral beliefs involved in determining whether one of one's own states is phenomenally conscious. The moral option can grant this, on either version. Even if one's own moral status is a partial ground of one's being phenomenally conscious and having the capacity for phenomenally conscious states, it does not mean that in determining which of one's states are conscious one would need to draw on moral considerations. The reason for this is that in such considerations, it is entirely taken for granted that one has moral status, and indeed that one is system-conscious. As such, one would be licensed in judging whether or not one's own states are phenomenally conscious to focus exclusively on non-moral facts about those states.

### 2.1.3. *The debate over the distribution problem*

In discussing the distribution problem, I want to focus on the debate over a particular case in order to be more fine-grained about how such debates proceed. I will focus on the debate over fish pain. In the debate over fish pain, those who accept that there is evidence of fish pain and those who do not appear to share agreement on the following claims:

- (1) Functional profile: there is a certain state, *f-pain\**, with a certain functional profile, which the relevant fish instantiate. *F-pain\** is involved in damage avoidance, plays a (quasi-)motivational role which can compete with other sources of (quasi-)motivation, is responsive to analgesics as reflected in the behavior of fish (who may pay significant costs to access analgesia), results in long-term alterations to behavior, and so on (Broom, 2014, 2016; Browman et al., 2019; Diggles et al., 2017; Key, 2015, 2016; Key et al., 2017; Meyers-Manor, 2018; Rose, 2002; Rose et al., 2014; Sneddon et al., 2014, 2018);
- (2) *If fish as a system have the capacity for phenomenal consciousness, then f-pain\* is phenomenally conscious;*
- (3) *If f-pain\* is phenomenally conscious, then fish as a system have the capacity for phenomenal consciousness;*
- (4) *If f-pain\* is phenomenally conscious, then f-pain\* is pain;*
- (5) *If fish experience pain or are phenomenally conscious, then our treatment of fish should change.*

Point 5 is important, because while it does not offer any direct support to the moral option, it does at least suggest that the debate proceeds on the assumption that phenomenal consciousness and pain have moral significance. Points 1 and 4 is important because they appear together to suggest that the debate over fish pain does not primarily reflect a difference in views of the functional and biological workings of fish, but instead a difference in views over what it takes for a system to instantiate pain; taken together with points 2, this appears to further reflect a disagreement over what it takes for a system to instantiate phenomenal consciousness. Point 3, like point 2, reflects the first relevance of the debate over fish pain to the distribution problem.

The separation of points 2 and 3 reflects a tacit agreement within the debate with my view of the distinction between state- and system-consciousness. Two separate, independently discussed and debated issues are (a) whether fish are phenomenally system-conscious, the denial of which is used as an objection to the claim that *f-pain\** is phenomenally state-conscious, *and* (b) whether *f-pain\** is phenomenally state-conscious, the denial of which is used to deny that there is any reason to believe that fish are phenomenally system-conscious.

The evidence offered bearing on (a) consists in evidence about the global functional and structural organization of and constraints on fish cognition and biology. The evidence offered bearing on (b) consists in evidence about the functional role and material implementation of *f-pain\** in particular. Not only do those party to that debate discuss whether relevant features of pain-in-humans are instantiated by *f-pain\**, they *also* discuss whether

relevant systemic features of human beings and especially their nervous systems are instantiated by fish and their nervous systems. Those who deny that f-pain\* is state-conscious do so in part *by* denying that fish are creature-conscious, eg, by denying that fish have the requisite neural basis for phenomenal consciousness (eg, Rose, 2002; Rose et al., 2014). Those who assert that f-pain\* is state-conscious do so in part *by* affirming that fish are creature-conscious, eg, by asserting that they have sufficiently complex “mental lives” to count as phenomenally conscious systems (eg, Broom, 2014, 2016).

Scientific work on the distribution problem in general, and on fish pain in particular, provides good abductive evidence for the moral option. Given the level of empirical agreement, the debate over fish pain looks intractable if consciousness is not at all a morally-determined property. The disagreement over fish pain turns primarily on how exacting the criteria for pain and phenomenal consciousness ought to be. Over, how “exacting” the criteria are turns on how similar fish are required to be to humans. In this debate, human beings are treated, sometimes implicitly and sometimes explicitly, as something of a paradigm case: the states and systemic properties of fish are compared with those of human beings in determining whether or not they count as phenomenally conscious (*contra* Figdor, 2017, 2018). However, clearly, for f-pain\* to count as phenomenally conscious pain and for fish to count as phenomenally conscious, they do not need to share every feature with human pain and human beings. Fish are not, after all, human beings.

What is needed is an account of how the criteria are and ought to be generated, given the myriad possible dimensions of comparison between fish and human beings (whether the comparison being made is between them or their states). There does not seem to be much promise in looking to conceptual analyzes of “phenomenal consciousness” or mental state terms, in looking to general scientific principles, or in holding out for direct access to the inner lives (or lack thereof) of fish. One answer has been explicitly offered within the debate, and it is not in agreement with the moral option – but nor, in my view, is it consistent with how the debate proceeds. Sneddon et al. (2014), claim that the criteria chosen are those that reflect the *function* of pain – damage avoidance.

For some criteria, this seems eminently plausible – for example, regarding fish, it is asked whether they are capable of avoidance behavior, and regarding f-pain\*, it is asked whether it plays a role in damage avoidance. However, even if this were true, not *all* the features of human cognition that reflect the damage-avoidance function of pain make the list – for example, it is not required that fish be able to reflect on which things are likely to be painful and which actions are likely to lead to pain.

Moreover, they clearly want *more* from fish pain than that it is a state whose function is damage avoidance – as they explicitly say, the resultant

behavior must be more than a mere reflex. Regarding fish, this means asking whether they are capable of long-term, motivated behavioral changes involving learning, alongside fear, selective attention, and trade-offs and balancing between motivations, and more (eg, Sneddon et al., 2014, 2018). Regarding f-pain\*, it is particularly telling that researchers have asked whether it can motivate behaviors such as “self-administration of analgesia” and “[paying] a cost to access analgesia” (Sneddon et al., 2014, p. 203).

What is the rationale behind these criteria? Clearly, they reflect more than the idea that pain is just some function involved in damage avoidance. The requirement that f-pain\* motivate paying a significant cost to access analgesics actually appears to count directly against achieving the function of damage-avoidance; instead, it seems to be based on the idea that f-pain\* should be a *bad* state to be in over and above the disvalue of any damage involved, one that fish are therefore motivated to avoid over and above their motivation to avoid damage. Likewise, the systemic features of fish that are required seem to be motivated by a requirement that fish should be capable of *recognizing* their state as a bad one to be in, one worth avoiding and worrying about, and that this recognition should be reflected in their trade-offs, learning, and long-term behavioral changes.

Endorsing the moral option helps to give a principled way of identifying which the salient dimensions of similarity are – they are the morally salient dimensions of similarity, those reflecting the moral significance of phenomenal consciousness and phenomenally conscious states like “pain”. The appeal to the function of pain from within the debate does not succeed. In fact, this appeal is belied by the criteria used by the very scientists who make the appeal. These criteria reflect an implicit agreement with the idea that it is the potentially morally-salient features of fish and their states which should be used to judge whether or not they are phenomenally conscious.

## 2.2. *The phenomenal stance*

The moral option is closely related to and partly inspired by a view of consciousness attributions called the “phenomenal stance” put forward by Robbins and Jack (Jack & Robbins, 2012; Robbins & Jack, 2006). They argue that there is compelling evidence from the cognitive sciences, with a special focus on cognitive neuroscience, that the capacity to attribute *intentional* mental states to a system (mental states that are *about* things or *directed toward* things) is distinct from the capacity to treat a system as *phenomenally conscious* (as being such that there is something it is like to be that system). They call this latter capacity the “phenomenal stance”. They do so in a nod to Dennett’s work, in which he puts forward and motivates a taxonomy of three “stances” one can take toward other individuals (Dennett, 1978, 1987, 1991).

Dennett calls his three stances the “physical”, “design”, and “intentional” stances – the first involves seeing an entity as an arrangement of physical entities, governed by physical laws; the second involves seeing an entity as a designed system, one with functional properties and parts; the last involves attributing intentional states to an entity, and seeing it as an agent. Robbins and Jack use the term “phenomenal stance” because, according to their view, Dennett’s taxonomy is importantly incomplete, and another “stance” is required to complete it.

Their 2006 paper puts forward the idea of the phenomenal stance and argues that is neurally and functionally distinct from the intentional stance and from the physical stance (which they also conceive as cognitive modules of some kind). The bulk of the philosophical work in their paper is to relate this picture to the meta-problem (and therefore the hard problem) of consciousness, treating the difficulty humans have explaining consciousness scientifically as primarily resulting from “cognitive Balkanization”, ie, as “result[ing] from more or less hard-wired restrictions on information flow across competing neural network” (*ibid.*, p.78).

Taking the phenomenal stance toward a system means treating it as a “phenomenal system”, “a locus on phenomenal experience” (Robbins & Jack, 2006, p. 69). While the phenomenal stance *involves* attributing phenomenally conscious states to systems, it “involves more than mere rote ascription” of these states, requiring in addition ‘a felt appreciation of their qualitative character’ (*ibid.*, p. 69–70). This “felt appreciation” is driven by “instinctive empathy”, directed toward the system to which the phenomenal stance is taken. Because of this, the phenomenal stance is not meant to be a “cold”, intellectual affair; the (involuntary) autonomic arousal associated with “hot”, “instinctive” empathy is treated as a core part of the stance (Robbins & Jack, 2006, pp. 68–69).

They flesh this out a little further by saying that the phenomenal stance essentially “involves a kind of emotional sensitivity” which consists in reacting appropriately to “the hedonic value for the bearer” of phenomenal states – “it is usually pleasant to observe another’s pleasure, and distressing to observe their distress” (Robbins & Jack, 2006, p. 70). As they put it, “if you can’t understand what it is to feel sad, you can’t regard something as feeling sad – at least, not in the full-blooded way that the phenomenal stance requires” (*ibid.*). Instinctive empathy, on their view, is therefore a morally-engaged form of imaginative empathy: it involves imaginatively reconstructing and occupying the perspective of another, in a way that engages moral motivation and emotional response.

This theory of consciousness attribution weakly supports the moral option. The move here is akin to Dennett’s move from the intentional stance to the nature of intentional states. In Dennett’s view, the function of the intentional stance is to enable us to better predict (and therefore

strategize about and manipulate) others' behavior. Dennett adopts what might be called "realist instrumentalism" about intentional states. It is instrumental in that it supposes that the reason for us to talk about intentional states and so forth is because it is instrumentally useful for us to do so. It is realist in that there is some way that systems themselves must be to warrant application of the intentional stance, and therefore to count as having propositional attitudes and so forth. In particular, any system that warrants the intentional stance must conform to a pattern of behavior such that the intentional stance can issue in non-compressible descriptions of that behavior – a "real pattern" (Dennett, 1991; for an attempt to fashion a sufficient condition out of this necessary condition, see; Ladyman & Ross, 2007).

The moral option might be seen as taking the analogous move with respect to the phenomenal stance: warranting instinctive empathy by having moral status is thereby cast as a constitutive necessary condition on being phenomenally conscious. Unfortunately, while this support may be genuine, it is indeed weak: it is not clear that Robbins and Jack operationalize phenomenal consciousness in a way consistent with the philosophical debate; it is not clear that Dennett's appeal to "function" would generalize to this context; and finally, it is not clear that Dennett's move from attributions of a class of state to the nature of that class of state is that compelling to begin with.

There are, however, two core insights that I want to emphasize from their work: first, that Dennett's theory of mental state (or intentional state) attributions does not generalize to phenomenal states, which appear to be more closely tied to moral cognition than strategic social cognition; secondly, that a certain form of empathy is plausibly closely tied to both moral status *and* attributions of phenomenal consciousness.

This relates to a key desideratum on any theory of consciousness: fleshing out the often-elusive idea of *what it is like*. One of the key anchors of the concept of phenomenal consciousness is the idea of *something it is like*: something it is like to be that creature, for creature-consciousness, and something it is like to be in that state, for state-consciousness. Fleshing this out is especially important because it is unclear what phenomenal consciousness is supposed to be in the first place, and that it is a widespread folk notion rather than a theoreticians' invention (Díaz, 2021; Mandik, 2016; Sytsma & Machery, 2010; Sytsma & Ozdemir, 2019).<sup>3</sup> The moral option, accepting this link between moral status, imaginative empathy, and phenomenal consciousness, can claim the following: what it means to claim that there is something it is like to be a particular creature is to claim it is (at least in principle) an appropriate object for this kind of morally-engaged, imaginative empathy; what it means to claim that there is something it is like to be in a particular state is to claim that, if we

imaginatively empathize a creature in such a state, being in such a state is part of what we ought to imagine.

### 2.3. *Non-moral components on the moral option*

As such, my preferred version of the moral option claims that phenomenally conscious systems are those that are appropriate objects of this particular kind of empathy, morally-engaged imaginative empathy, where they are appropriate objects of this particular kind of empathy partly because they have moral status. Phenomenally conscious states of a system, in turn, are those that one would have to imagine when engaging empathetically with that system.

The moral option, as I am defending it, does not claim that moral status is *all* that is required for phenomenal consciousness. Partly, this is dialectical: my argument supports the idea that when it comes to phenomenal consciousness and moral status, the order of explanatory priority runs from moral status to phenomenal consciousness. However, I also do not *believe* that moral status is sufficient for phenomenal consciousness. For example, I believe that people in comas have moral status. More generally, I believe that a larger class of systems plausibly has interests, goals, or conditions of welfare and wellbeing (henceforth, just “interests”) than is conscious, and that any of these things alone may well be enough to secure some kind of moral status (Bradford, 2023) – for example, rainforests may well have interests, as well as conditions of welfare and wellbeing, and this may be enough to secure them some kind of moral status even in the absence of phenomenal consciousness.

This is not a problem for the empathy-based version of the moral option I am defending. Having interests does not guarantee that a system is an appropriate object of any kind of empathy. There are at least three ways that one might go about defending this position, all of which are compatible with one another. One is to claim that, eg, a rainforest is not an appropriate object of empathy because, despite its moral status, it is simply far beyond the limits of our empathetic and imaginative capacities. This is not my preferred way to go, since I would want to grant that there might be phenomenally conscious systems that we cannot empathize with only due to such limitations – as such, I think when it comes to attributions of phenomenal consciousness, the moral option should idealize away from limits on such capacities in humans.

Another way to go is to claim that the systemic features that might secure interests for a system still do not suffice to secure that it is an appropriate object of empathy. Of course, one could not just leave it there. However, there are many plausible candidates for systemic features that rainforests apparently lack and that make it very hard to see how they might

legitimately be objects of empathy. One proposal along these lines that I find appealing is that they lack anything resembling beliefs, a perspective, and other key prerequisites of any broadly epistemic notion of “subjectivity” or “a point of view” (cf. Davidson, 1984, 2001; McDowell, 1996a, 1996b; Sellars, 1963)

The final way to go is to claim that rainforests lack the capacity to have any states that are plausibly state-conscious. They have no states that might be, eg, globally broadcast to control systems, poised for control of their molar behavior, or indeed any of the other candidates for what might give rise to a phenomenally-conscious state of a phenomenally-conscious system. I find both of these latter two claims plausible in the case of rainforests, and this is I hope suggestive of a more general way of thinking about how a system might have moral status and yet not be phenomenally conscious: it might have interests, and yet lack the capacity for any plausibly conscious states, and lack the kinds of systemic features that plausibly bestow a system with a point of view on the world.

#### 2.4. *The metaphysics of the moral option*

One might worry that the moral option is metaphysically quite extravagant: that it requires both a dualism about mental properties, and a realism about moral properties. I do not believe that this is true. Even if phenomenal consciousness is *not* a physical property, the moral option still does not require us to posit a set of fundamental mental properties. For a realist, it requires us to acknowledge *moral* properties, of which phenomenal consciousness is just one. If moral properties are not physical properties, this may in a sense be dualistic, but it is more parsimonious for any moral realist who is already forced to acknowledge moral properties.

I, however, am not a moral realist in this sense. I hold a position indicated by Price (1997, 2011, 2015, 2019), building on Blackburn (1993), p. I am a naturalist, and arguably a “quasi-realist” in that I do not take myself to be committed to a domain of fundamental, mind-independent moral properties, and I believe that human moral talk is scientifically explicable. However, at the same time, I believe that moral talk is literally true (and that there are in this sense moral facts), in just the same way as other kinds of talk including the most technical discussions of scientific physics and ordinary talk of tables and chairs.

Additionally, I believe that the moral option is consistent with physicalism, at least if it opts for the anchoring version. On such a view, *being phenomenally conscious* consists in instantiating an extremely complex conjunctive and disjunctive physical property. Of course, instantiating this physical property counts as instantiating phenomenal consciousness due to

moral facts, but for reasons adumbrated above, I do not take these moral facts to be metaphysically extravagant in the first place.

The moral option, as I am proposing it, therefore claims that a system's being phenomenally conscious is a matter of being an appropriate object of "instinctive empathy", morally-engaged imaginative empathy. All such systems have moral status. Facts about moral status either partially ground or anchor facts about which systems are phenomenal consciousness, and this is already implicit in debates over the distribution problem, where it should be made explicit. Additionally, I have suggested that two further features are partial grounds for system-consciousness: first, the capacity for states of the sort that would be state-conscious is a system-conscious system; secondly, a "perspective", perhaps constituted by the kind of epistemic states that are required for a system to be an appropriate object of any form empathy, morally-engaged or not. This is a position consistent with physicalism, and which does not require commitment to moral realism.

### 3. Problems with the moral option

#### 3.1. *The counterintuitiveness of the moral option*

To begin, I want to try to undercut some of the counterintuitiveness of the moral option. I am not sure I can fully get rid of its apparent counterintuitiveness, but I do not think this is especially damning or anything close to conclusive. One way to undercut its counterintuitiveness is to highlight that there are many things that people believe are wholly non-moral, but are in fact probably partly morally constituted. For example, many people believe that a state's being a democracy is entirely a matter of its procedures and organization, but it is extremely plausible that it is in fact partly a moral matter whether or not a state ought to be counted as a democracy.

Another way to undercut the counter-intuitiveness of the moral option is to highlight that similar arguments have been made regarding other psychological capacities, and regarding psychological capacities in general. For example, Crary (2016) argues attributing many psychological predicates constitutively involves attributing morally-salient properties to other individuals, and that moral properties are partly constitutive of the psychological capacities thereby ascribed; if this is correct, then it is no surprise that the same applies to phenomenal consciousness.

It is also worth stressing that many significant arguments surrounding psychological language plausibly make tacit appeals to moral considerations in making their case. For example, Putnam's (1967) hugely influential argument for the multiple realizability of mental states considers a Martian whose inner workings were entirely different than our own, but who nevertheless behaved exactly as if they were in pain. He

believes that we would naturally say, and should say, that they experience pain. However, *ex hypothesi*, their physical make-up is so different than ours that their pain could not plausibly be said to have the same physical realization base. In my view, part of the intuitive appeal of this argument is that it would clearly be *wrong* to ignore the apparent pain of such a creature, and that their version of pain would be saliently similar to ours in making the same kinds of moral demands on us. Such considerations about how we would feel obliged to treat and interact with systems with radically different constitutions than us are drawn on quite explicitly in Schwitzgebel (2015), who is directly discussing consciousness rather than merely the phenomenally conscious state, pain.

It may also be that the view that phenomenal consciousness is morally grounded is more counterintuitive than the view that it is morally anchored. The view that it is partly morally grounded would entail that some moral property of a conscious system directly (partly) explains why it is system-conscious. However, on the view that it is partly morally anchored, the only direct involvement of moral facts is in determining the criteria for which systems are and are not system-conscious. For me at least, this is beyond the realms of anything I have intuitions about one way or another.

A final way of trying to undermine the counterintuitiveness of the moral option is to ask how else phenomenal consciousness could have so much moral significance? One particularly significant attempt to overcome Hume's argument for an "is-ought gap" between descriptive and moral concepts is to claim that many key concepts are "thick" with both descriptive and evaluative content, defying a clear division between moral and non-moral language (Väyrynen, 2016). If this is the right line of response, then it would be no surprise if a concept with as much moral significance as phenomenal consciousness turns out to be thick; if thick concepts refer to partly morally-constituted phenomena, then it would be no surprise if phenomenal consciousness turns out to be partly morally-constituted.

While not exactly a form of counter-intuitiveness, one might worry that the moral options, especially a version that emphasizes the role of the phenomenal stance, renders it scientifically intractable to study phenomenal consciousness. I do not believe that this is the case, and indeed, hope that it might be helpful in guiding appropriate forms of scientific inquiry. It admittedly suggests that some methods might need to be engaged which are not widely accepted in the science of consciousness as it stands, but these are not methods which find no home in scientific practice. In particular, this theory suggests that explicit moral debate and disagreement, and that rigorous use of empathy, might be necessary for fixing the target of consciousness research, and for determining answers to the distribution question.

This is not, in my view, a particularly dangerous conclusion. Research on many topics, such as health, wellbeing, democracy, and illness, requires explicit moral and ethical consideration in setting the target for scientific study (Alexandrova, 2017; Gough, 2022; Kingma, 2007, 2010, 2017; Krauss, 2016; Nagel, 1961). Work already exists on making rigorous use of empathy informed by careful observation and empirical data in scientific research, including in anthropology and ethnography (eg, Watts, 2008), psychiatry (eg, Gough, 2021, 2023), and indeed in animal psychology (eg, Steward, 2018). Such work argues that empathy can be informed at least by information about the inner workings of a particular system (Gough, 2021, 2023), but also by careful observation of a system in its environment (Crary, 2016; Steward, 2018). It may well be possible to make informed, rigorous, and falsifiable judgments about whether they have the kinds of inner lives and possess the kinds of moral status that warrant the relevant kind of empathy, and are jointly constitutive of phenomenal consciousness.

### 3.2. *Zombies*

The story thus far is open to a possible objection based on the much-discussed philosophical zombie, a system functionally and physically just like us, but without any phenomenal consciousness (D. Chalmers, 1996). Zombies arguable constitute a counter-example to the theory for two reasons. First, because given the functional and physical similarity, it appears to be totally groundless to say that it is not an appropriate object of instinctive empathy. Secondly, because despite the functional and physical similarity, many view the philosophical zombie as at least morally different than us, if it has any moral status at all. In both cases – whether in explaining why it is not an appropriate object of empathy, and or explaining why it is morally different than us – it seems natural to draw on its lack of phenomenal consciousness, a move which is not obviously available to a defender of the moral option.

I do not find this objection particularly compelling. In my view, in stipulating that the zombie is not phenomenally conscious, the thought experiment simply stipulates that it is not an appropriate object of the relevant kind of empathy. We are simply told not to imagine its inner life, and not to project ourselves into its shoes. This is conceivable, because it is possible to conceive of a system, even one which if we were to encounter it in the actual world we would feel forced to empathize with, and yet not empathize with it. The apparent change in moral status turns on our refusal to empathize in a way that it is essential to engaging certain kinds of moral cognition. As such, while I admit the conceivability of philosophical zombies, on these grounds, I am inclined to deny their metaphysical possibility: it is possible to conceive of a creature just like us but which we *ought* to treat

differently in any number of ways, including that we should not extend our empathy to it; it does not mean that such a creature is metaphysically possible.

The zombie problem is also used to illustrate the hard problem of consciousness. The hard problem of consciousness is that of explaining *why* a physical system or state is conscious rather than not – whatever the “correlates” of phenomenal consciousness might be, it is the problem of explaining *why* they are the correlates of phenomenal consciousness. The closely related “meta-problem” of consciousness is the problem of explaining why the hard problem seems so hard in the first place. An answer only available to forms of the moral option is that that the hard problem is hard because it is an instance of the general difficulty of the “is-ought gap”, the problem of deriving moral facts from non-moral facts. This also helps to answer the meta-hard problem: the hard problem *seems* hard because it *is* hard, in turn because it is, to reiterate, hard to derive moral facts from non-moral facts.

### 3.3. Explaining moral status in terms of phenomenal consciousness

It is extremely widely-believed that there is *some* important link between phenomenal properties and moral properties. The dominant option in current attempts to explain the link between phenomenal consciousness and moral status is to explain moral status in terms of phenomenal consciousness, not *vice versa*. There are several possible explanations of the link between phenomenal consciousness and moral status on the table.

Shepherd (2024) considers a range of possible explanations for how phenomenal consciousness might explain moral status. In particular, he considers how phenomenal consciousness might explain moral status on a class of positions, “sentientism” on which either phenomenal consciousness, or some species of phenomenal consciousness, is both necessary and sufficient for moral status. He divides sentientism into valence sentientism, and broad sentientism. Valence sentientism, according to which the capacity for valenced phenomenally conscious states is the full grounds of moral status, he thinks has an easy time explaining the link between phenomenal consciousness and moral status:

The valence sentientist can say that valenced experiences are valuable because the experiences are good or bad. Valenced experiences wear their value on their sleeve, so to speak— the value is there in the phenomenal character, in what it is like to have the experiences. (p.3)

Broad sentientism, according to which phenomenal consciousness *in general* is the full grounds of moral status, he thinks has a harder time explaining this link. However, he sees there as being several possible options.

D. J. Chalmers (2022, p. 419) argues that phenomenal consciousness, even in the absence of valenced experiences, can ground “serious intellectual and moral goals.” Chalmers, in support of this claim, offers the example of Vulcans, sophisticated creatures with rich inner lives but without any valenced experiences. Shepherd considers several possible ways of explaining this intuition; the ones he considers most promising build on Birch (2022) and consider the possibility that conscious registration of one’s interests might somehow ground non-derivatively valuable or disvaluable experiences. Lee (2022) adumbrates toward a similar position; he argues that sentientism is not in an analogous dialectical position to speciesism because it is clear that many conscious entities have interests, but not clear that any non-conscious entities have interests, and suggests that this may be because conscious entities are capable of knowledge and desire while non-conscious entities are not.

Even if phenomenal consciousness is not *necessary* for moral status, as Shepherd (2023, 2024) in fact concludes, there may still be explanatory links from phenomenal consciousness to moral status. For example, Bradford (2023, p. 914) argues against the claim that phenomenal consciousness is necessary for moral status, but for the claim, “CCC” that any “state involving conscious awareness of a [welfare-enhancing state] *has greater value* that a [welfare-enhancing state] obtaining but *not* entering a subject’s awareness.” On such a view, phenomenal consciousness is not necessary for moral status, but is nevertheless an important part of the moral status and level of moral considerability of particular systems. She explains this view, according to which phenomenally conscious awareness grounds *greater value* or disvalue, as follows:

conscious awareness of a W-state exemplifies a *unifying* relation. Consciousness ties the [welfare-enhancing state] to its possessor by bringing it into his experience. The whole consisting of a [welfare-enhancing state] and [the relevant conscious state] is unified by consciousness. When a subject S is aware of S’s [welfare-enhancing state] *w*, S’s [relevant conscious state] *unites* S with *w*. Since a unifying relation is instantiated, we have our explanation for why the whole is an instance of organic unity. CCC is, in other words, a specific instance of a more general value theoretic principle, the principle of organic unity. Consciousness, one might say, therefore enhances or activates the latent value of the. [welfare-enhancing state] (pp.916–917)

These positions differ in extremely significant ways, but they all share a similar form. Phenomenal consciousness is importantly linked to moral status because it involves a capacity for a phenomenally conscious states (whether valenced or not, and whether part of an organic unity or not), and this kind of state grounds important moral properties bearing on the significance and the very existence of a system’s interests.

Two objections might be levied against my position on the basis of the state of the debate. The first is that the moral option is rendered implausible

because it rules out these plausible theories of the link between phenomenal consciousness and moral status. However, I believe that this is only the case on the basis of several further assumptions which are not central to the moral option. These positions all hold that (some facts about) the phenomenal consciousness of a particular system explain the moral status of that system, and offer theories of which facts and why. This is only ruled by the moral option if it entails that (a) the moral status of a particular system explains all the facts about its phenomenal consciousness, whether partially or wholly, and (b) there is no possibility of reciprocal or mutual explanation, that is, of two facts which are each part of the explanation of the other. Claim (b) is contentious, and clearly no part of the moral option. Claim (a) is only plausibly part of the moral option if the determination relation involved is specified as grounding, but is not part of the moral option if it instead opts for anchoring.

The second objection is based on the passive perceiver, a phenomenologically conscious system with a rich perceptual experience of the world, but without interests, goals, valenced states, or any other properties that might bestow it with moral status (Lee, 2018; Shepherd, 2024). The passive perceiver is used as an objection to broad sentientism, and in support of valenced sentientism or other forms of sentientism that claim that only species of consciousness suffice for moral status.

The passive perceiver, constitutes a counter example because it is clearly phenomenally conscious, and yet apparently has no moral status. It is this latter point that I am inclined to deny, and here I want to clarify what I mean by “moral status” as I promised that I would in the introduction. The passive perceiver is a system with no morally-salient state-conscious states, and no interests. It does not, therefore, appear to make any moral demands on us. However, I think there is still a sense in which it has a moral status different than any system that is *not* phenomenally conscious.

I have already suggested that being phenomenally creature-conscious, having the capacity for phenomenally conscious states, and having interests each depend on their own set of constraints. I will refer again to the feature that make the difference between conscious and unconscious states in a phenomenally conscious system as “F”. The passive perceiver is unlike an unconscious system in at least one key way, and possibly two. Most significantly, if the passive perceiver were to instantiate a negatively valenced state with feature F, it would be of direct moral significance. Conversely, were an unconscious system to token a negatively valenced state with feature F, it would nevertheless not be a conscious negatively valenced state, and nevertheless not be of direct moral significance.

More controversially, it seems plausible to me that if it could be induced to have interests, these interests would make a different kind of demand on us than those of any unconscious system – in just the same way as the

interests of a cat and the interests of a rainforest make different kinds of demand on us. This is perhaps controversial, at least if, following Lee (2022), the key difference between conscious and non-conscious systems may well be whether they have interests.

Put a different way, the passive perceiver is still an appropriate object of morally-engaged and morally-informative empathy: it is just that what we learn when we empathize with the passive perceiver is that there are no demands on us. However, being an appropriate object of this kind of empathy still bestows a potential significance on many of its states and interests which the states of unconscious systems lack: *were* it to have any valenced states or interests, *then* they ought to engage our moral motivation and emotions in the special way that instinctive empathy allows.

### 3.4. *Decoupling moral status and phenomenal consciousness*

A quite different form of objection to the moral option relies on the denial that phenomenal consciousness has *anything* to do with moral status. Such a position has been elaborated and defended by Carruthers (1999, 2000, 2004). Carruthers defends a higher-order thought theory of consciousness, according to which second-order representations (representations of representations) are the basis for phenomenal consciousness. He also views it as implausible that (the relevant kind of) second-order representations are widespread throughout the animal kingdom.

In defending this theory, the first two objections that Carruthers (2000, ch.7) considers are that it is clearly counterintuitive to deny phenomenal consciousness to many animals, and that it is morally unacceptable to do so because of the link between moral status and phenomenal consciousness. To the first objection, he argues that our intuitions about the distribution of phenomenal consciousness are worth very little, and moreover, that the intuition “can easily be explained away as a mere by-product of imaginative identification with [animals]’ states’ (p.183).

To the second objection, he claims that phenomenal consciousness, once properly understood, has very little to do with moral status. In particular, he claims that while phenomenal consciousness relies on higher-order representations, morally significant states rely only on first-order representations. These morally salient, first-order representational states can exist in the absence of higher-order representation and hence of phenomenal consciousness, in his view. Such first-order representational states include desires and the frustration thereof (Carruthers, 2000) and suffering (Carruthers, 2004). Carruthers’ (Carruthers, 1999, 2000, 2004) key move in defending the higher-order thought theory is to suggest that perhaps “*sympathy* and *imagination* can, and should, be pulled apart” (1999, p.476). He suggests that

“[s]ympathy can surely be grounded in a purely third-person understanding of someone’s situation and/or mental states; and certainly, a *desire to help* need not be preceded by any sort of imaginative identification” (*ibid.*).

There are two aspects of Carruthers’ argument that I want to highlight, since I think that they actually go some way to casting the moral option as the default option in the debate. First, that Carruthers thinks that the intuitive, default position, which stands in need of rejection by his lights, is that phenomenal consciousness *is* tightly linked to moral status. Secondly, that Carruthers thinks that the standard way people come to view other systems as by phenomenally conscious is by “imaginative identification” with their states.

I am not a believer in higher-order theories of consciousness, nor am I convinced by Carruthers’ notion of full-blooded sympathy without imaginative identification – although certainly, I think desires to help and the recognition of moral duties may proceed without imaginative identification. However, my aim here is not to reject the majority of Carruthers’ position, since I believe that it is in fact largely compatible with the moral option.

The higher-order thought theory of consciousness proceeds quite explicitly in the way I rejected in §2.1.2. Carruthers (2000, ch.6) motivates the higher-order thought theory of consciousness primarily by surveying a wide range of unconscious, first-order representational states in humans, and then suggesting that the feature that plausibly makes the difference between unconscious first-order representational states, and conscious ones, is whether there is an associated higher-order representation. Higher-order representations are cast as F, the feature that makes the difference between unconscious and conscious states. The presence or absence of higher-order representations is, at least in Carruthers’ eyes, an entirely non-moral fact. Carruthers then assumes that in working out whether animals are conscious *all* one needs to do is work out whether or not they have the capacity for states with this non-moral feature – and concludes that they generally do not, and that they are therefore not phenomenally conscious.

The *only* way that this is not consistent with the moral option is the assumption that *all* one needs to do to work out whether a system is phenomenally conscious is to look at whether a system has the capacity for higher-order representations. This is, in my view, patently false. One could easily imagine any number of systems with a small repertoire of first-order representational states, and an equally small repertoire of higher-order representational states corresponding to each first-order representational state – for example, a digital thermometer with representations of temperature, that also generated second-order representations of its representations of temperature, perhaps in the form of a backlog or record so they could be checked later or used to generate data. It seems highly implausible

to claim that such a system would be conscious, especially while claiming that, eg, cats and dogs are not.

What is lacking from the slightly-sophisticated thermometer? While it has the capacity for higher-order representational states, which might be exactly the sort of state that is state-conscious in a system-conscious system, it does *not* have the features that make it a system-conscious system. It is this latter set of features which I claim are morally significant, and whose presence or absence is either partially morally-grounded or morally-anchored. I have already explained in §2.3 why I do not believe that it is a problem for the moral option if there are systems with moral status that lack phenomenal conscious – as Carruthers construes many animals to be. As such, there is no inconsistency between the moral option and the view that the higher-order thought theory identifies a necessary condition on phenomenal consciousness, and fully explains state-consciousness (or the lack thereof) *among states of phenomenally conscious systems*.

#### 4. Sharpness and panpsychism

Goff (2014, 2017) offers an argument for panpsychism from the non-vagueness of phenomenal consciousness. He claims that no non-reductive theory of phenomenal consciousness (that admits its existence) other than panpsychism can meet the following criterion: *non-vagueness*. According to Goff, “it is not a vague matter whether or not something is conscious” – there must be a fact of the matter, one way or another, about whether any given system is phenomenally conscious (2017, p.239). A term is vague if and only if it admits of *penumbra: borderline cases* in which it is not clear, and not rationally decidable, whether or not the term can be properly applied.

On a worldly theory of vagueness, it is the world itself that is vague (eg, Williams, 2008). On a linguistic theory of vagueness, descriptions are vague because their proper application to borderline cases is not fully determined by their use (eg, Fine, 1975). Goff’s argument, in brief, claims that the only way to get a perfectly precise condition on which systems are phenomenally conscious (other than denying that *anything* is conscious) is to claim that *everything* is conscious – ie, to endorse panpsychism. Otherwise, an implausibly small difference between two highly similar systems will render one phenomenally conscious and one not.<sup>4</sup>

Goff (2017) puts the issue most evocatively by considering the growth of a human from embryo to baby. If the embryo is unconscious, and the baby is conscious, there must be some precise moment where the developing system goes from unconscious to conscious. But this seems implausible – any one of the changes along the way seems unfit to be *the* change that makes the difference. A similar problem arises in considering the

distribution problem – there exists a multidimensional scale of complexity with complex life-forms at one extreme, and the most simple life-forms at the other (eg, Godfrey-Smith, 2016, 2019). Where, precisely, on this scale is the line between the phenomenally conscious and phenomenally unconscious systems supposed to lie? There seems to be no rational way to get to a perfectly precise answer to this question using a purely functional characterization of phenomenal consciousness.<sup>5</sup>

The moral option has tools here available to it alone. Many share the intuition that moral concepts are “sharp” in just the same way as Goff thinks the concept of phenomenal consciousness is; since the moral option can say that moral facts are part of the determination of which systems are and are not phenomenally conscious, it can claim that phenomenal consciousness inherits its “sharpness” from the moral properties involved (cf. Cutter, 2017).

This is not a knock-down response, since some have argued that moral concepts can be vague (Dougherty, 2014; Dunaway, 2017; Schoenfield, 2016). However, there are good reasons not to want to believe that moral concepts are vague. If they are, it results in all sorts of troublesome consequences – either one must abandon moral realism, or endorse ontic vagueness (Schoenfield, 2016); even if one abandons moral realism, is not entirely clear that opponent metaethical theories are much better placed to handle the vagueness of moral concepts (Baima, 2014). As such, and given the possibility of responding to many of the arguments that moral concepts are vague, Kultgen (2008) argues that we should adopt a general strategy of treating it as an absolute last resort to assume that moral concepts as vague.

Others, such as Constantinescu (2014) convey the vagueness of moral concepts as unproblematic, “reflect[ing] a wholesome state of undecidedness concerning certain moral issues, indicating that we have left the door open for further debate and negotiation” (p. 152) – although Constantinescu conveys this as analogous to concepts like *bald* and *heap*, paradigmatically vague terms, it seems to me that there is a significant difference. In those cases, there is simply no rational way to decide whether or not the term applies; in moral cases, the decision has simply yet to be made. If this is right, then moral concepts are *not* vague any more than it is “vague” precisely how tall a mountain is before we measure it – we simply do not *know* yet.

More generally, many of the arguments for moral vagueness seem to rely on a kind of *subjectivism* – the idea that moral properties and facts constitutively depend on our moral beliefs, and so that prior to our decision about whether something is, eg, permissible, there is no fact of the matter about whether it is permissible (eg, Dougherty, 2014; Dunaway, 2017; Pravato, 2022). Since I reject this kind of subjectivism, I reject most of these arguments for the vagueness of moral concepts.

Even if moral concepts *can* be vague, the moral option has tools available to it alone. The reason for this is that even if moral concepts can *in fact* be vague many appear to share the *intuition* that they are not, and this may explain away the *intuition* that phenomenal consciousness is not vague – and it is this that needs explaining, since it is at least equally controversial to claim that “phenomenal consciousness” does not admit of borderline cases.

## 5. Conclusion

Many of us become interested in the questions of phenomenal consciousness because we are interested in working out which other creatures and systems we should care about and why. The moral option speaks directly to this concern. The moral option is the view that a system’s *being phenomenally conscious* is partly explained by moral properties of that system. It reflects the state of the debate over the distribution problem of which systems are phenomenally conscious, fits well with a theory of consciousness attributions as essentially relying on morally-engaged imaginative empathy, provides positive guidance to the science of consciousness, and has unique tools at its disposal for resisting arguments for panpsychism. It provides a novel diagnosis of the zombie problem, hard problem, and meta-problem of consciousness. It is consistent with there being a scientific study of consciousness, and consistent with nonreductive physicalism. It should therefore, in my view, be a serious contender in debates over the nature of phenomenal consciousness.

## Notes

1. I am somewhat ambivalent about whether the moral option counts as a naturalistic option. As I argue below, it is neither antiscientific, nor incompatible with physicalism, at least on some construals.
2. This may perhaps put some pressure on the significance of the distinction – see also Schaffer (2019), who argues that “anchoring” is just a kind of grounding.
3. D. Chalmers (2018) contends in response to Sytsma and Machery (2010) that mental state terms have both a “phenomenal” and “functional” reading, and that experimental evidence for the concept of phenomenal consciousness is not easily acquired because some mental state terms, like “see”, tend to receive a functional reading while others, like “pain”, tend to receive a phenomenal reading. The moral option has some tools to explain this imbalance – it might be that “pain” is more likely to receive a phenomenal reading because it is more morally significant, and attributions of pain are more closely tied to morally-engaged imaginative empathy.
4. On an epistemic theory of vagueness, use *does* fully determine whether or not the term applies, but it is a simple brute fact, and we cannot know one way or the other whether the term applies (eg, Williamson, 1994). According to Goff, phenomenal consciousness is simply not vague in a worldly sense; phenomenal consciousness cannot be vague in the second sense because it is impossible to “sharpen” our use of

the term (see 2014 for his argument). Goff does not consider the epistemic theory of vagueness; however, his argument for panpsychism works claims to work from the premise that phenomenal consciousness is not vague in a linguistic or worldly sense, so ought to be compatible with epistemic vagueness.

5. Goff is not the first to offer an argument from vagueness about phenomenal consciousness. However, many of these arguments run from vagueness to *eliminativism* about phenomenal consciousness (Antony, 2008; Everett, 1996).

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## Funding

This work was supported by the British Academy.

## ORCID

Joseph Gough  <http://orcid.org/0000-0003-2591-5744>

## References

- Alexandrova, A. (2017). *A philosophy for the science of well-being*. Oxford University Press.
- Antony, M. V. (2008). Are our concepts conscious state and conscious creature vague? *Erkenntnis*, 68(2), 239–263. <https://doi.org/10.1007/s10670-007-9061-2>
- Baima, N. (2014). The problem of ethical vagueness for expressivism. *Ethical Theory and Moral Practice*, 17(4), 593–605. <https://doi.org/10.1007/s10677-014-9493-0>
- Bayne, T. (2007). CONSCIOUS STATES AND CONSCIOUS CREATURES: EXPLANATION IN THE SCIENTIFIC STUDY OF CONSCIOUSNESS. *Philosophical Perspectives*, 21(1), 1–22. <https://doi.org/10.1111/j.1520-8583.2007.00118.x>
- Birch, J. (2022). Materialism and the moral status of animals. *The Philosophical Quarterly*, 72(4), 795–815. <https://doi.org/10.1093/pq/pqab072>
- Blackburn, S. (1993). *Essays in Quasi-Realism* (Vol. 1). Oxford University Press.
- Bradford, G. (2023). Consciousness and welfare subjectivity. *Noûs*, 57(4), 905–921. <https://doi.org/10.1111/nous.12434>
- Broom, D. M. (2014). *Sentience and animal welfare*. CABI.
- Broom, D. M. (2016). Fish brains and behaviour indicate capacity for feeling pain. *Animal Sentience*, 3(4). <https://doi.org/10.51291/2377-7478.1031>
- Browman, H. I., Cooke, S. J., Cowx, I. G., Derbyshire, S. W. G., Kasumyan, A., Key, B., Rose, J. D., Schwab, A., Skitfesvik, A. B., Stevens, E. D., Watson, C. A., & Arlinghaus, R. (2019). Welfare of aquatic animals: Where things are, where they are going, and what it means for research, aquaculture, recreational angling, and commercial fishing. *ICES Journal of Marine Science*, 76(1), 82–92. <https://doi.org/10.1093/icesjms/fsy067>
- Carruthers, P. (1999). Sympathy and subjectivity. *Australasian Journal of Philosophy*, 77(4), 465–482. <https://doi.org/10.1080/00048409912349231>
- Carruthers, P. (2000). *Phenomenal consciousness: A naturalistic theory*. Cambridge University Press.

- Carruthers, P. (2004). Suffering without subjectivity. *Philosophical Studies*, 121(2), 99–125. <https://doi.org/10.1007/s11098-004-3635-5>
- Chalmers, D. (1996). *The conscious mind: In search of a fundamental theory*. Oxford University Press.
- Chalmers, D. (2018). The meta-problem of consciousness. *Journal of Consciousness Studies*, 25(9–10), 6–61.
- Chalmers, D. J. (2022). *Reality+: Virtual worlds and the problems of philosophy*. W. W. Norton.
- Constantinescu, C. (2014). Moral vagueness: A dilemma for non-naturalism. In R. Shafer-Landau (Ed.), *Oxford studies in metaethics: Volume 9* (pp. 152–185). Oxford University Press.
- Crary, A. (2016). *Inside ethics: On the demands of moral thought*. Harvard University Press.
- Cutter, B. (2017). The METAPHYSICAL IMPLICATIONS of the MORAL SIGNIFICANCE of CONSCIOUSNESS. *Philosophical Perspectives*, 31(1), 103–130. <https://doi.org/10.1111/phpe.12092>
- Davidson, D. (1984). First person authority. *Dialectica*, 38(2–3), 101–111. <https://doi.org/10.1111/j.1746-8361.1984.tb01238.x>
- Davidson, D. (2001). *Subjective, intersubjective, objective* (Vol. 3). OUP Oxford.
- Dennett, D. C. (1978). *Brainstorms: Philosophical essays on mind and psychology*. (4155184) 1st ed. Bradford Books.
- Dennett, D. C. (1987). *The intentional stance*. MIT Press.
- Dennett, D. C. (1991). Real patterns. *The Journal of Philosophy*, 88(1), 27–51. <https://doi.org/10.2307/2027085>
- Díaz, R. (2021). Do people think consciousness poses a hard problem?: Empirical evidence on the meta-problem of consciousness. *Journal of Consciousness Studies*, 28(3–4), 55–75.
- Diggles, B. K., Arlinghaus, R., Browman, H. I., Cooke, S. J., Cowx, I. G., Kasumyan, A. O., Key, B., Rose, J. D., Sawynok, W., Schwab, A., Skiftesvik, A. B., Stevens, E. D., Watson, C. A., & Wynne, C. D. L. (2017). Responses of larval zebrafish to low pH immersion assay. Comment on Lopez-luna et al. *The Journal of Experimental Biology*, 220(Pt 17), 3191–3192. <https://doi.org/10.1242/jeb.162834>
- Dougherty, T. (2014). Vague Value. *Philosophy and Phenomenological Research*, 89(2), 352–372. <https://doi.org/10.1111/phpr.12026>
- Dunaway, B. (2017). Ethical vagueness and practical reasoning. *The Philosophical Quarterly*, 67(266), 38–60. <https://doi.org/10.1093/pq/pqw038>
- Epstein, B. (2015). *The ant trap: Rebuilding the foundations of the social sciences*. Oxford University Press. (p. 18265216).
- Epstein, B. (2019). Anchoring versus grounding: Reply to schaffer. *Philosophy and Phenomenological Research*, 99(3), 768–781. <https://doi.org/10.1111/phpr.12644>
- Everett, A. (1996). Qualia and vagueness. *Synthese*, 106(2), 205–226. <https://doi.org/10.1007/BF00413700>
- Figdor, C. (2017). On the proper domain of psychological predicates. *Synthese*, 194(11), 4289–4310. <https://doi.org/10.1007/s11229-014-0603-2>
- Figdor, C. (2018). *Pieces of mind: The proper domain of psychological predicates*. Oxford University Press.
- Fine, K. (1975). Vagueness, truth and logic. *Synthese*, 30(3–4), 265–300. <https://doi.org/10.1007/BF00485047>
- Godfrey-Smith, P. (2016). Mind, matter, and metabolism. *The Journal of Philosophy*, 113(10), 481–506.
- Goff, P. (2014). Orthodox property dualism+the linguistic theory of vagueness=Panpsychism. In R. Brown (Ed.), *Consciousness inside and out*:

- Phenomenology, neuroscience, and the nature of experience* (pp. 75–91). Springer Netherlands. [https://doi.org/10.1007/978-94-007-6001-1\\_7](https://doi.org/10.1007/978-94-007-6001-1_7).
- Goff, P. (2017). *Consciousness and fundamental reality*. Oxford University Press.
- Gough, J. (2021). On the proper epistemology of the mental for psychiatry: What's the point of understanding and explaining? *The British Journal for the Philosophy of Science*, 000–000. <https://doi.org/10.1086/715106>
- Gough, J. (2022). The many theories of mind: Eliminativism and pluralism in context. *Synthese*, 200(4), 325. <https://doi.org/10.1007/s11229-022-03804-w>
- Gough, J. (2023). Understanding understanding in psychiatry. *History of Psychiatry*, 34(3), 249–261. <https://doi.org/10.1177/0957154X231163275>
- Jack, A. I., & Robbins, P. (2012). The phenomenal stance revisited. *Review of Philosophy and Psychology*, 3(3), 383–403. <https://doi.org/10.1007/s13164-012-0104-5>
- Key, B. (2015). Fish do not feel pain and its implications for understanding phenomenal consciousness. *Biology and Philosophy*, 30(2), 149–165. <https://doi.org/10.1007/s10539-014-9469-4>
- Key, B. (2016). Why fish do not feel pain. *Animal Sentience*, 3(1). <https://doi.org/10.51291/2377-7478.1011>
- Key, B., Arlinghaus, R., Browman, H. I., Cooke, S. J., Cowx, I. G., Diggles, B. K., Rose, J. D., Sawynok, W., Schwab, A., Skiftesvik, A. B., Stevens, E. D., & Watson, C. A. (2017). Problems with equating thermal preference with 'emotional fever' and sentience: Comment on 'Fish can show emotional fever: Stress-induced hyperthermia in zebrafish' by rey et al (2015). *Proceedings of the Royal Society of London B*. (2015). 284(1847), 20160681. 20160681. <https://doi.org/10.1098/rspb.2016.0681>
- Kingma, E. (2007). What is it to be healthy? *Analysis*, 67(2), 128–133. <https://doi.org/10.1111/j.1467-8284.2007.00662.x>
- Kingma, E. (2010). Paracetamol, poison, and polio: Why boorse's account of function fails to distinguish health and disease. *British Journal for the Philosophy of Science*, 61(2), 241–264. <https://doi.org/10.1093/bjps/axp034>
- Kingma, E. (2017). Health, disease, and naturalism: Hausman on the public value of health. *Public Health Ethics*, 10(2), 109–121. <https://doi.org/10.1093/phe/phx001>
- Krauss, A. (2016). The scientific limits of understanding the (potential) relationship between complex social phenomena: The case of democracy and inequality. *Journal of Economic Methodology*, 23(1), 97–109. <https://doi.org/10.1080/1350178X.2015.1069372>
- Kultgen, B. (2008). *Vagueness and ethics*. [PhD dissertation]. University of Colorado.
- Ladyman, J., & Ross, D. (2007). *Every thing must go: Metaphysics naturalized*. Oxford University Press.
- Lee, A. Y. (2018). Is consciousness intrinsically valuable? *Philosophical Studies*, 175(1), 1–17. <https://doi.org/10.1007/s11098-018-1032-8>
- Lee, A. Y. (2022). Speciesism and sentientism. *Journal of Consciousness Studies*, 29(3), 205–228. <https://doi.org/10.53765/20512201.29.3.205>
- Mandik, P. (2016). Meta-Illusionism and Qualia Quietism. *Journal of Consciousness Studies*, 23(11–12), 140–148.
- Manson, N. (2000). State consciousness and creature consciousness: A real distinction. *Philosophical Psychology*, 13(3), 405–410. <https://doi.org/10.1080/09515080050128196>
- McBride, R. (1999). Consciousness and the state/transitive/creature distinction. *Philosophical Psychology*, 12(2), 181–196. <https://doi.org/10.1080/095150899105864>
- McDowell, J. (1996a). Action, meaning, and the self. In *Mind and world* (pp. 87–107). Harvard University Press; JSTOR. <https://doi.org/10.2307/j.ctvjghtzj.9>
- McDowell, J. (1996b). Rational and other animals. In *Mind and world* (pp. 108–126). Harvard University Press; JSTOR. <https://doi.org/10.2307/j.ctvjghtzj.10>

- Meyers-Manor, J. E. (2018). If it looks like a duck: Fish fit the criteria for pain perception. *Animal Sentience*, 21(7). <https://doi.org/10.51291/2377-7478.1328>
- Nagel, E. (1961). *The structure of science: Problems in the logic of scientific explanation*. Harcourt, Brace & World.
- Pravato, G. (2022). Indeterminacy and Normativity. *Erkenntnis*, 87(5), 2119–2141. <https://doi.org/10.1007/s10670-020-00293-6>
- Price, H. (1997). Naturalism and the fate of the m-worlds: Huw price. *Aristotelian Society Supplementary Volume*, 71(1), 247–268. <https://doi.org/10.1111/1467-8349.00029>
- Price, H. (2011). *Naturalism without mirrors*. Oxford University Press.
- Price, H. (2015). 8 from quasi-realism to global expressivism—and back again? In *Passions and projections: Themes from the philosophy of Simon Blackburn* (p. 0). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780198723172.003.0008>.
- Price, H. (2019). Global expressivism by the method of differences. In *Royal institute of philosophy supplement* (Vol. 86, pp. 133–154). Cambridge Core. <https://doi.org/10.1017/S1358246119000109>.
- Putnam, H. (1967). Psychological predicates. In W. H. Capitan & D. D. Merrill (Eds.), *Art, Mind, and Religion* (Vol. 1, pp. 37–48). Pittsburgh: University of Pittsburgh Press.
- Robbins, P., & Jack, A. I. (2006). The phenomenal stance. *Philosophical Studies*, 127(1), 59–85. <https://doi.org/10.1007/s11098-005-1730-x>
- Rose, J. D. (2002). The neurobehavioral nature of fishes and the question of awareness and pain. *Reviews in Fisheries Science*, 10(1), 1–38. <https://doi.org/10.1080/20026491051668>
- Rose, J. D., Arlinghaus, R., Cooke, S. J., Diggles, B. K., Sawynok, W., Stevens, E. D., & Wynne, C. D. L. (2014). Can fish really feel pain? *Fish and Fisheries*, 15(1), 97–133. <https://doi.org/10.1111/faf.12010>
- Schaffer, J. (2019). Anchoring as grounding: On epstein's the ant trap. *Philosophy and Phenomenological Research*, 99(3), 749–767. <https://doi.org/10.1111/phpr.12645>
- Schoenfield, M. (2016). Moral vagueness is ontic vagueness. *Ethics*, 126(2), 257–282. <https://doi.org/10.1086/683541>
- Schwitzgebel, E. (2015). If materialism is true, the United States is probably conscious. *Philosophical Studies*, 172(7), 1697–1721. <https://doi.org/10.1007/s11098-014-0387-8>
- Sellars, W. (1963). *Empiricism and the philosophy of mind*. Routledge Kegan and Paul.
- Shepherd, J. (2023). Non-human moral status: Problems with phenomenal consciousness. *AJOB Neuroscience*, 14(2), 148–157. <https://doi.org/10.1080/21507740.2022.2148770>
- Shepherd, J. (2024). Sentience, Vulcans, and zombies: The value of phenomenal consciousness. *AI & SOCIETY*. <https://doi.org/10.1007/s00146-023-01835-6>
- Sneddon, L. U., Elwood, R. W., Adamo, S. A., & Leach, M. C. (2014). Defining and assessing animal pain. *Animal Behaviour*, 97, 201–212. <https://doi.org/10.1016/j.anbehav.2014.09.007>
- Sneddon, L. U., Lopez-Luna, J., Wolfenden, D. C. C., Leach, M. C., Valentim, A. M., Steenbergen, P. J., Bardine, N., Currie, A. D., Broom, D. M., & Brown, C. (2018). Fish sentience denial: Muddying the waters. *Animal Sentience*, 21(1), 115. <https://doi.org/10.51291/2377-7478.1317>
- Steward, H. (2018). Morgan's canon: Animal psychology in the twentieth century and beyond. In *Animals*. Oxford University Press. <https://doi.org/10.1093/oso/9780199375967.003.0018>.
- Sytsma, J., & Machery, E. (2010). Two conceptions of subjective experience. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, 151(2), 299–327. <https://doi.org/10.1007/s11098-009-9439-x>
- Sytsma, J., & Ozdemir, E. (2019). No problem: Evidence that the concept of phenomenal consciousness is not widespread. *Journal of Consciousness Studies*, 26(9–10), 241–256.

- Väyrynen, P. (2016). Thick ethical concepts. In E. N. Zalta (Ed.), *Stanford encyclopedia of philosophy*.
- Watts, J. H. (2008). Emotion, empathy and exit: Reflections on doing ethnographic qualitative research on sensitive topics. *Medical Sociology Online*, 3(2), 3–14.
- Williams, J. R. G. (2008). Ontic vagueness and metaphysical indeterminacy. *Philosophy Compass*, 3(4), 763–788. <https://doi.org/10.1111/j.1747-9991.2008.00151.x>
- Williamson, T. (1994). *Vagueness* (Vol. 81). Routledge.