**Consciousness is not the Key to Moral Standing**

**David Papineau**

**1. Introduction**

Which creatures have moral standing? Precisely those that are conscious, says nearly everyone[[1]](#footnote-1). In this paper I shall argue that this is wrong. The concept *conscious* is ill-suited to delimit the class of *moral patients*—that is, creatures with moral standing, creatures with moral interests.

The concept *conscious* is ill-suited to define this moral category not because it focuses on the wrong thing, but because it focuses so badly. It is a loose categorization that serves our purposes well enough when we are dealing with human beings. But it fails to draw clear lines once we move beyond the human realm to consider various non-human creatures. In particular, it fails to draw clear lines in cases where the moral lines are themselves clear. I take this to show that the moral facts must be grounded in something other than facts about consciousness.

In this paper I shall take materialism about the mind as given. As I see it, my overall argument simply draws out some of the consequences which follow from materialism. Still, while I take my conclusions to be reasonably straightforward corollaries of materialism, they are by no means widely accepted, even among those who profess to be materialists. In my view, this is because many of those who identify as materialists do not fully embrace materialism. They harbour implicit dualist thoughts, and this stops them seeing clearly what materialism requires. I shall return to this issue below.

I shall not attempt to defend materialism here. This is a large subject on which I and others have written extensively elsewhere, and it would take us too far afield to repeat the arguments.[[2]](#footnote-2) So the analysis of this paper will be of most direct concern to those who are already committed to materialism. But even those who are not so committed might still be interested in the conditional thesis that, if you embrace materialism, then you should not think of moral standing as hinging on questions of consciousness.

Some seek to address the relevance of consciousness to morality by considering philosophical “zombies”, beings who are posited as physically identical to human beings but who lack any conscious experiences. Given that zombies lack any conscious feelings, these philosophers ask, should they have less moral standing than their normal human counterparts, or even no moral standing?

For materialists, zombies are metaphysically impossible, even if they are coherently conceivable. Because of this, the significance of moral intuitions about zombies is not straightforward. Still, we need not get bogged down in these dialectical niceties at this stage. Rather than start with zombies, I shall focus on real cases from the animal world, and shall also consider possible artificial intelligences. Clarity about these by no means fanciful cases will help us to see what to say about zombies. I shall return to zombies in section 13 below.

**2. Theories of Consciousness**

My target is the view that non-human creatures have moral standing if and only if they are conscious. As a preliminary to addressing this view, let us first consider how we might decide whether or not some non-human creature is conscious. (I shall say that a *creature* is conscious if it is *sometimes* in a state that renders it conscious. So a conscious creature needn’t be conscious all the time.)

Well, plenty of theories of consciousness are on offer: global workspace[[3]](#footnote-3), higher-order[[4]](#footnote-4), integrated information[[5]](#footnote-5), perceptual reality monitoring[[6]](#footnote-6), unlimited associative learning[[7]](#footnote-7), and so forth. And, in line with this, plenty of studies examine how far different animal species[[8]](#footnote-8), and indeed different artificial intelligence systems[[9]](#footnote-9), satisfy these theories’ accounts of consciousness.

As might be expected, however, this kind of analysis produces mixed results. Theories that set the bar for consciousness high, such as higher-order theories, are relatively demanding, and so generally deem creatures who are less sophisticated than humans not to be conscious. By contrast, less demanding theories of consciousness, like those requiring only integrated information or unlimited associative learning, are far more inclusive and happy to grant consciousness to a range of non-human animals and even some existing artificial intelligence systems.

This kind of divergence presents a challenge to the programme of deciding moral standing on the basis of consciousness. If we can’t agree which creatures are conscious, how can we tell which qualify as moral patients? In response, defenders of the programme are likely to object (indeed Ned Block and others have objected[[10]](#footnote-10)) that measuring non-human creatures against different theories of consciousness puts the cart before the horse. The important issue, they will say, is not which creatures are conscious *according to* different theories of consciousness, but which creatures are *phenomenally conscious*. Whether any of the above-mentioned theories of consciousness captures the nature of phenomenal consciousness is an empirical matter. In effect, say the objectors, different theories of consciousness propose different a posteriori accounts of the material basis of phenomenal consciousness. These proposals then need to be compared with the evidence to see which if any has successfully identified the material “correlates of consciousness”. Only then will be in a position to use the successful theory to tell us which non-human creatures enjoy phenomenal consciousness.

**3. Phenomenal Consciousness**

What is phenomenal consciousness and how can we find out which a posteriori theory gets it right? There are two questions here. Let me take them in turn.

In his influential “On a Confusion about a Function of Consciousness” (1995), Ned Block drew a distinction between phenomenal consciousness and “access” consciousness. Mental states possess access consciousness, Block specified, if they are generally available for reasoning, action control and (in creatures with language) speech control. As to phenomenal consciousness, Block admitted that he could not define it “in any remotely non-circular way” and that “really all one can do it *point* to the phenomenon”. By way of further elucidation, Block invoked Thomas Nagel’s characterization of conscious states as those that “it is like something” to be in.

The idea, then, is that we get the idea of phenomenal consciousness directly, from our familiarity with our own phenomenally conscious states. This direct first-personal way of thinking about conscious states then leaves it open what further physical or functional processes might constitute consciousness. It is now standard to say that this way of thinking deploys “phenomenal concepts”. There is a large literature on the nature and significance of phenomenal concepts[[11]](#footnote-11), but for present purposes all that will matter is that phenomenal concepts afford a way of thinking about conscious states that leaves it conceptually open what further physical or functional features such states might possess.

A point of clarification before proceeding. We can distinguish a determinable phenomenal concept of consciousness-as-such—a concept of the determinable property involved in all conscious states—from determinate phenomenal concepts of more specific phenomenal properties, such as being in pain, hearing a loud noise, and so on. My focus in what follows will be on the determinable concept of consciousness-as-such, though many of my points will apply to the more determinate phenomenal concepts as well.

And a point of terminology. When I use “consciousness” without qualification in what follows, I should be understood to mean phenomenal consciousness. I myself doubt that that this concept succeeds in picking out some fully determinate feature of reality, for reasons that will become clear as proceed. But I do not dispute that it is the right way to frame the philosophical issues. Those who hold that moral standing is tied to consciousness are generally to be understood as deploying the concept of phenomenal consciousness that Block has articulated, and so I shall follow them in this.

**4. Identifying the Neural Correlates**

Now for the second question, about empirical investigation of which material processes constitute consciousness. At first pass, an obvious methodology recommends itself. Since our concept of consciousness derives from our familiarity with our own conscious states, why not use that familiarity to pin down the material correlates of consciousness? Thus we might observe, or experimentally arrange, that people are in such-and-such brain states, and then ask them to tell us whether or not those states are conscious. By this means we might hope to identity some brain property that is always present whenever humans are conscious and always absent when they aren’t.

This methodology can produce some striking results. For example, it suggests that the visual information (the “where” information) that guides out immediate bodily movements is not conscious.[[12]](#footnote-12) Conversely, it indicates that subjects whose painful suffering is relieved by morphine remain conscious of their pains even after they cease to trouble them.[[13]](#footnote-13)

Still, questions can be raised about this methodology. In particular, the reliance on subjects’ reports threatens to weigh the scales illegitimately in favour of an identification of consciousness with some species of access consciousness. For example, when some subject reports that some cognitive state is not conscious—say the visual information guiding their hand movements—is it legitimate therewith to conclude that the state really isn’t phenomenally conscious? Might not the negative report be due, not to a lack of phenomenal consciousness per se, but to some barrier to the subject accessing that consciousness? After all, the whole idea of phenomenal consciousness was introduced to avoid assuming any automatic tie between general cognitive accessibility and phenomenal consciousness itself.

Some researchers aim to overcome this worry with a “no-report” paradigm that relies, not on explicit reports of consciousness, but on other symptoms that have been found to be associated with phenomenal consciousness in uncontroversial cases.[[14]](#footnote-14) It is less than clear, however, that this fully addresses the worry, given that subjects’ reports would still seem to play an essential role in picking out the associated symptoms in the uncontroversial cases.[[15]](#footnote-15)

A different way of driving a methodological wedge between phenomenal consciousness and cognitive access appeals to the idea of *natural kinds*. Natural kinds are categories whose instances share many properties. For example, all samples of gold share colour, lustre, density, melting and boiling points, electrical and heat conductivity, and many other properties. Again, all snakes are alike in having slender bodies, no limbs, external ears, or eyelids, and only one functional lung. Now, in nearly all such cases, one central property will typically explain why all the other kind-typical properties cluster together in the kind’s instances. (Godman, Mallozzi and Papineau 2020.) Sometimes this central property will be an intrinsic property, as with chemical substances like gold, where molecular constitution explains the shared properties. In other cases the central property will be extrinsic, as with biological taxa like snakes, where shared ancestry arguably plays the central explanatory role.

This structure of natural kinds creates room for some of our initial ways of identifying kinds to prove mistaken. We might initially classify samples of iron pyrite as gold, on account of their colour and lustre, but these identifications would be in error given that iron pyrite does not share the molecular constitution of genuine gold. Similarly, we might initially consider slowworms to be snakes, based on their slender bodies and lack of limbs, but this too would be a mistake given that they have a different ancestry from snakes.

Nicholas Shea (2012) has suggested that we can appeal to natural kinds to add structure to the study of consciousness. (See also Bayne and Shea 2022.) In the first instance, we identify conscious states via their accessibility to reasoning and action control in humans. But then further investigation reveals that these states share a cluster of other distinctive properties. Shea suggests that, in addition to their connection to cognitive access, phenomenally conscious states turn out all to interact with associative conditioning in particular ways, to permit sensitivity to novel features of complex stimuli, to be susceptible to certain kinds of illusion, to be related to specific neural structures, and so on.

As with natural kinds in general, some central explanatory property might then turn out to be responsible for this clustering of properties in conscious human states. Thus Shea conjectures that local cortical resonance sustained by cortico-thalamic loops might play this role. That is, perhaps this kind of local resonance underlies all the further characteristic features of consciousness. If so, then this would then create room for certain states that are not themselves cognitively accessible to qualify as conscious. For example, if some locally resonant states were not so accessible, we would then have reason, given Shea’s conjecture, to conclude that cognitive accessibility per se is an initial but by no means infallible guide to consciousness. Of course, it might also turn out that general cognitive accessibility itself offers the best account of the clustering of features in conscious states, in which case we would have reason to conclude a posteriori that such availability is indeed constitutive of consciousness. But at least Shea’s methodology does not predetermine this outcome a priori, but rather makes it answer to the empirical facts.

**5. Problems with Animals**

Let us then suppose, for the sake of the argument, that empirical research allows us to draw a line around the category of neural states that are conscious in humans. Perhaps use of the “no-report” paradigm, or Shea’s appeal to natural kinds, will imply that these states do not entirely coincide with those which are cognitively accessible. Or it might turn out, on the other hand, that consciousness in humans always involves cognitively accessibility after all.

Still, however that turns out, far worse problems arise when we seek to find out about consciousness in non-human animals. The methodology of consciousness research quickly starts to unravel once we move away from the human case.[[16]](#footnote-16)

The underlying problem is the clustering of a range of different and dissociable features in conscious human states. As we have seen, in addition to the initial criterion of cognitive accessibility, these states will also all display interaction with specific kinds of associative conditioning, sensitivity to novel features of complex stimuli, certain distinctive neural processes, and so on. Now, these features can, and do, come apart in other creatures. Other animals will share some, but not all, of the features that are characteristic of consciousness in humans, and those that they do share will generally be shared to a limited degree. This raises questions about how high to set the bar for consciousness. For examples, does it suffice for consciousness to match humans in associative conditioning, even if not in sensitivity to novel stimuli, and if so are the full range of human associative powers required?

At first pass, empirical research into humans would seem impotent to answer such questions. Since conscious human states fully display all the relevant features, we won’t be able to decide the issue by asking humans which of their states are conscious. This will just return us to the class of human states that fully share the many different features, without telling us which of these features, and to what degree, are crucial for consciousness. What we need are rather test cases in which the suite of features which characterise human consciousness are only partially present, plus some way of checking whether that still suffices for consciousness. Well, a range of animals might provide test cases all right—but of course we will have no independent way of checking whether they are conscious, since they can’t tell us.

It might seem that Shea’s methodology could help here. What if some underlying neural feature turns out to explain the markers of consciousness in humans, and so clarifies which of them are really diagnostic of consciousness? Well, this would be relevant, but it will scarcely remove the puzzle, since any such underlying neural feature will itself come in degrees. To illustrate, consider Shea’s own suggestion: consciousness depends on local resonance sustained by cortico-thalamic loops. But how much resonance, and how much of a cortico-thalamic loop? Even if we stick to mammals, brain structures vary significantly. Should we then require the specific cortical structures that are present in humans, or is it enough that local neural resonance is delivered by some more basic brain process? Once more, research into humans would seem impotent to decide this, since human conscious states will satisfy both the demanding and the permissive requirement. And again it will be no good looking to other animals to provide test cases, since we will have no independent hold on whether they are conscious.

It gets even worse once we bring in non-mammals as well. Many birds, arthropods and cephalopods arguably display similar mental powers to intelligent mammals. But in them these powers are underpinned by very different brain structures which have evolved independently of the mammalian cerebral cortex. In birds, learning and memory are facilitated, not by a developed cortex, but by the differently structured avian pallium. In arthropods the analogous powers depend on “mushroom bodies”. And in cephalopods they seem to derive from systems of brain lobes. Does this make a difference to whether or not these non-mammals are conscious?

To focus the issue, suppose that we have somehow settled, despite the arguments above, on the level of mental ability required for consciousness in mammals—certain powers of associative conditioning and stimulus recognition, say, together with the cortical structures that sustain them. Now, what about the birds, arthropods and cephalopods who display the same levels of cognitive sophistication but lack the cortical structures? Do their cognitive powers suffice for consciousness, or are they disqualified as candidates for consciousness because of their qualitatively different brains? Once more, empirical research seems to be stymied. Research with humans cannot decide the issue. Conscious human states will involve both the cognitive powers and the mammalian brains. So when humans confirm these states as conscious, that doesn’t tell us whether the consciousness depends on the cognitive powers alone or whether it needs the mammalian cortical structures as well. The relevant non-mammals, by contrast, have the cognitive powers without the cortical brains. But that doesn’t help with our question either, given that we have no independent way of checking whether or not they are conscious, since once more they can’t tell us.[[17]](#footnote-17) [[18]](#footnote-18)

**6. The Indeterminacy of Animal Consciousness**

What should we make of these barriers to finding out which non-human animals are conscious? A natural first thought is that the difficulty is purely epistemological. Surely reality itself draws a clear line between those animals who enjoy consciousness from those who do not. After all, how could there not be such a line? On one side the light of consciousness is on, on the other it is not. True, our current research strategies seem unable to reveal where this line lies, for the reasons outlined in the last section. Still, we might hope that further research and reflection will eventually succeed in overcoming the epistemological challenge. Alternatively, perhaps, we might be condemned to permanent ignorance about the conscious status of our non-human relatives. But in either case, so this first thought does, the fault lies with our state of knowledge, not with the real distinction between conscious and non-conscious creatures.

I have a quite different response. I don’t accept there is any definite fact of the matter as to whether non-human animals are conscious. In my view, it is objectively indeterminate whether cephalopods, or crabs, or crows, or even cats are conscious.

I say this not because of some general verificationist principle that we can’t lay claim to facts that outstrip our ability to find out about them. I hold no brief for the view that our representational powers are limited to matters that are epistemically accessible to us. Rather my denial of determinacy is specific to the concept of consciousness itself. I don’t think this concept is sharply enough constituted to determine whether non-human creatures are conscious or not.

Recall how the concept was articulated above. It laid down no a priori conditions on conscious states. Rather it secured its reference ostensively. We were supposed to get the idea directly, from familiarity with our own conscious states. As Block put it, “really all one can do it *point* to the phenomenon”.

Given this, it should scarcely be surprising that the concept becomes indeterminate once we move away from its ostensive human base. It covers states . . . like *these*—and then we point to our own paradigm conscious human states. But “likethese” in which respects exactly? Similar in their involvement in distinctively human mental powers, or only in the kind of mental powers other animals share with us? Similar in their fine-grained cortical features or their more abstract structural ones? No pointing at human states is going to tell us, given that the states we are pointing to manifest all these features.

**7. The Intuition of Distinctness**

So I say that there is no facts of the matter about consciousness once we move beyond the familiar human cases. Some readers will no doubt find this claim incredible. How can it be indeterminate whether crabs or crows are conscious? Think of it from the animal’s point of view. Is it like something to be them, or not? How can there not be a fact about that, even if it is hard for us to find out? Either the light is on, or it isn’t. Maybe in the most primitive conscious creatures the light is dim. But a dim light is still a light.

I grant that this is a strong intuition. But I say it is driven by a mistaken metaphysical picture. In my view, the idea that consciousness must be determinate stems from an implicit commitment to dualism. We find it very natural to think that conscious phenomena are somehow metaphysically extra to brain processes, some kind of additional element of reality that is present in conscious beings. And this thought then of course implies that consciousness will always be determinate, that there will be a definite line between cases whether this extra element is present—where the light is on—and cases where it isn’t.

In previous writings I have argued at length that this dualist “intuition of distinctness” is well-nigh inescapable and responsible for much confusion about consciousness (Papineau 1993, 2002, 2007b, 2020). Rather than repeat these arguments here, let me simply observe that nearly everybody who studies consciousness, including many of those would count themselves as physicalists, are very ready to talk about brain processes “generating”, or “yielding”, or “giving rise to”, or “causing” conscious states. But this is already dualist talk. Fire “generates”, “causes”, “yields” or “gives rise to” smoke. But H2O doesn’t “generate”, “cause”, “yield” or “give rise to” water. It *is* water. To speak of brain processes as “generating” conscious states, and so on, already betrays an implicit commitment a metaphysical gap between the brain states and the conscious states.

One question which arises at this point is the source of this dualist intuition. If dualism is false, why does it force itself on us in this intuitive way? However, let us not pursue this issue here. There is plenty of relevant literature. (See Papineau 2011, 2020 for references.) The more important point for present purposes is simply that we do all feel this intuition, even those of us who are theoretically opposed to it, and so find ourselves thinking that it must be determinate which creatures are conscious.

Here is a parable that will help make my position clear. Imagine that an indigenous Arctic tribe traditionally makes a substance from whale blubber that they burnt in lamps and used for lubrication. They call it “oil”. Then some petrochemical company develops a structurally indistinguishable product but with a different chemical composition and market it as “new oil”. A heated debate follows. Is this new stuff really *oil*? The more sensible members of the tribe argue that there is no fact of the matter here, only an occasion for terminological decision. When the term “oil” was originally introduced, they observe, nobody exercised themselves as to whether it had to derive from whale blubber or rather only to have the right structural properties. After all, why should anybody have bothered about this, given that hitherto the two options had been extensionally equivalent? So nothing in previous usage, the sensible tribe members argue, decides whether the new petrochemical product satisfies the requirements for “oil”. The only issue is whether the meaning of the term should now be made more precise, and if so in which direction.

Still, some of the tribe’s members are not satisfied. They think there is a crucial further issue that isn’t answered just by noting the new substance’s chemical and structural features—namely, do those features give rise to *oil* or not? So they deny that the labelling of the new stuff is just a matter for terminological decision. “How can it be arbitrary?” they ask. “The real issue is whether or not this new chemistry generates *oil*. That’s what the term ‘oil’ answers to, not just the underlying material features of the substance.”

I take it that these “oil dualists” would be misguided. The material features of oil don’t *generate* some further stuff present in genuine oil. There’s nothing but the material features. And if previous usage was unspecific about exactly which such features are definitive of oil, then there is nothing in reality to decide the question.

I say that we materialists should say just the same about consciousness. The concept of consciousness is introduced by example – “really all one can do it *point* to the phenomenon” – and this leaves it open exactly which material features are characteristic of consciousness. Given this, what could possibly determine whether consciousness in present in creatures that display some of the characteristic human features but not others? The situation seems quite analogous to the oil case. Many animals possess some but not all the features that are characteristic of consciousness in humans. This cannot but be indecisive as to whether they have the same kind of state as we point to in our own case – *that* kind of state. In my view, any remaining feeling that animal consciousness must be determinate can only stem from an implicit commitment to dualism, as with the oil dualists. The impression of determinacy is grounded in the intuition of distinctness.[[19]](#footnote-19)

If any readers are still feeling that consciousness must always be determinate (either the light is on or it is not) it is perhaps worth observing that some indeterminacy at least will be an inevitable consequence of *any* materialist analysis of consciousness, quite independently of the more specific arguments from the looseness of phenomenal concepts that I have offered. To see this, suppose that, despite the difficulties I have raised, we had somehow succeeded in putting our finger on some specific material property that determines consciousness. By the nature of the case, however, any such material property will still come in degrees—we can have more or less cortical resonance, or cognitive accessibility, or perceptual monitoring, or whatever. Yet it would be arbitrary to fix some precise point at which we have just enough for consciousness.[[20]](#footnote-20) Rather there will be borderline cases which are neither in nor out, as with any sorites-type phenomenon. So I say materialism itself pretty immediately implies that consciousness will sometimes be indeterminate. To the extent that we harbour an intuition that it can’t be, that can again only be because we haven’t yet fully embraced materialism and are still thinking of consciousness as something extra to the physical realm.[[21]](#footnote-21)

**8. Animal Welfare without Consciousness**

So far I have argued that consciousness is indeterminate. The aim of this paper, though, is to show that consciousness does not determine moral standing. It would be a mistake to move straight from the former claim to the latter. Even if consciousness is indeed indeterminate, it might yet be what determines moral standing. All that would follow is that moral standing is itself indeterminate.

This stance would make reasonable sense if the indeterminacy of consciousness were simply a matter of degree, arising where some sorites-like property enters its penumbra of vagueness. To illustrate, suppose once more, as at the end of the last section, that consciousness did depend on some identifiable material property. As I pointed out, since this feature would inevitably come in degrees, and any sharp cut-off point would be arbitrary, we would have borderline cases of consciousness. So, in between primitive animals that lack the requisite property, and more developed animals that clearly have it, there would be some borderline cases—jawless fish, perhaps—that are neither conscious nor not conscious. This wouldn’t force us to detach moral standing from consciousness, however. We could keep them tied together, and simply conclude that the moral status of jawless fish is itself indeterminate. On this view, while primitive animals like roundworms definitely lie outside the moral circle, and complex animals like cats definitely within, there would be no fact of the matter about the standing of jawless fish. We wouldn’t definitely be wrong to ignore their welfare, but we wouldn’t definitely be right either, precisely because of the indeterminacy of their consciousness.

I have argued, however, that the indeterminacy of consciousness is far more radical than sorites-style borderlines. And this blocks any prospect of keeping moral standing yoked to consciousness. The trouble is that moral standing is definite in many cases where consciousness is not.

After all, my analysis implies, not just that the consciousness is indeterminate for jawless fishes, but also that it is indeterminate for cats and crows and crabs and octopuses. With cats we face the “levels” question of whether they have *enough* of the features displayed by the paradigm conscious states of humans. And with crows, crabs and octopuses we face the further “realizer” question of whether they need, not just enough of those features, but also that they be realised by the sameevolved brain structures as in mammals.

So I say it is indeterminate whether cats and crows and crabs and octopuses are conscious or not. But I certainly don’t think there is no fact of the matter about their moral standing. These are complex creatures capable of much sophisticated behaviour. Of course they qualify as moral patients. Someone who appreciated their nature but felt that their moral standing was an open question would just be making a mistake. (I shall say more to back up these moral dicta in a moment.)

The consequence I draw is that moral standing must be grounded in something other than facts of consciousness. Moral standing is determinate in cases where consciousness is not. The features of non-human creatures that qualify them as moral patients must be more specific than whether they are conscious.

To say this is not to say that all questions of moral standing are fully determinate. I do not rule out some creatures falling in a borderline between those who definitely are moral patients and those who are not. My present point only requires that moral standing is *more* determinate than consciousness. Some creatures definitely have moral standing even though it is indeterminate whether they are conscious—cats and crows and crabs and octopuses, for instance. And this is enough to show that moral standing cannot be determined by consciousness.

**9. The Grounds of Moral Standing**

What might ground moral standing, if consciousness cannot? This is a difficult topic, and I will not attempt any definitive answer in this paper. Still, it will be worth making some general observations.

A first point to note it that we are concerned with a specific kind of moral fact. Moral standing involves things that are good or bad *for* some creature. It is facts of this kind that endow creatures with moral interests and thereby give them moral standing. It is because things can be good or bad *for them* that we ought to take their interests into account. (See Crisp 2021.)

We can call moral facts like these *welfare facts*. We can contrast them with *intrinsic* moral facts, ones that are good or bad independently of their being good or bad for any creature. It is perhaps contentious whether there are any intrinsic moral facts, but it is not implausible to hold, say, that a world with mountain ranges would be morally better than a homogenous desert, or a world with primitive life better than a dead world, even if there were no creatures for whom these things would be good.

So the question we need to address is: what makes something good or bad *for* some creature? (Equivalently, I shall take it, what endows a creature with *interests*, what is it for something to add or detract from a creature’s *welfare*?)

Now, one possible answer would be to say that any *living* organism has interests in the relevant sense. Evolved living creatures are imbued with teleology, in that their parts have functions that contribute to their survival and reproduction, and in this sense it will serve their interests that these functions be fulfilled.

However, I take it that this line of thought will lead to setting the bar too low. While some deep ecologists are prepared to argue in this way that all living beings have interests (Taylor 1986, Johnson 1993), I shall assume in what follows that we want a notion of welfare that excludes plants and bacteria from the category of moral patients.

But what principle then includes the cats, crows, crabs and octopuses, while excluding the plants and bacteria? At this point the attraction of appealing to consciousness becomes clear. Conscious subjects can have positive or negative experiences, experiences that are automatically good or bad *for them*. Plants and bacteria are not conscious and so are excluded from the moral circle.

However, as we have seen, the category of consciousness is insufficiently determinate to draw the lines that are needed. Still, perhaps it points us in roughly the right direction. Moral patients must be possessed of some kind of *mental* life. This is what makes things good or bad *for them*. They must have a point of view, a perspective from which they engage with the world (Godfrey-Smith 2020).

In one sense such a point of view will be possessed by any organism that distinguishes between itself and its environment and responds with activity appropriate to its circumstances. This minimal notion of a mental subject, however, is still too thin for our purposes. It is arguable that plants and even bacteria engage in “proto-cognition” when selecting responses to environmental conditions (Godfrey-Smith 2016). To exclude them from the moral circle we need a notion of mentality that requires more than merely distinguishing between self and environment.

One possible move at this point is to appeal to the notion of *desire*. This is the option favoured by two other authors who agree with me about the irrelevance of consciousness to moral standing. Peter Carruthers (2019) and Francois Kammerer (2022) have both recently argued, along lines similar to mine[[22]](#footnote-22), that our initial conceptual hold on consciousness is too weak to draw definite lines in the animal realm, and they too draw the conclusion that issues of animal welfare cannot depend on whether the animals are conscious. In response, both of them suggest that we should focus instead on whether animals possess desires, and that we should equate their welfare with whether or not their desires are satisfied.

Does not the idea of desire already presuppose consciousness? Not necessarily. We can think of desires functionally rather than phenomenologically, in terms of what they *do* rather than what they *feel* like. Thus we might equate them with any mental states which prompt behaviours apt to produce certain results, and perhaps which also reinforce those behaviours when those results are produced. This then opens the way to equating the welfare of animals with the satisfaction of their desires in this functional sense. Consciousness needn’t come into it. [[23]](#footnote-23)

**10. Desires and Animal Welfare**

I am sympathetic to the general idea of relating moral standing to the possession of desires. But I think the story needs to be made a bit more complicated.

For a start, there are familiar objections to a direct equation of welfare with desire satisfaction. First, is desire satisfaction always sufficient for positive benefit? If the satisfaction of a desire simply means the achievement of the desired result, why is it automatically a good thing, even pro tanto, that desires be satisfied? Suppose that I desire something this is of no worth at all, not even from my point of view—to count the blades to grass on my lawn, or to smoke a cigarette—and suppose it is further specified that I will in no sense enjoy these outcomes if they are achieved, nor feel bad if they are not. Would not the world be just as good a place if these results did not occur? That I desire the results, in the functional sense of being moved to pursue them, seems neither here nor there. If the results are worth nothing in themselves, why is it better for these desires to be satisfied than not?

This worry would seem to apply all the more in the animal world. While animals generally tend to pursue results that conduce to their flourishing, they are also often motivated to seek outcomes that are pointless from any perspective. Accidental learning contingencies can easily lead them to value worthless outcomes, and even innate urges can on occasion lead to behaviour whose results serve no good purpose.

And then there is a converse question about the necessity of desire for positive benefit. Plenty of circumstances would seem to contribute to animal welfare even when the animals concerned do not in any sense have desires directed specifically at them. For example, health, or companionship, or freedom to roam, are crucial to the welfare of many animals that lack the psychological sophistication to desire such abstract ends. While such animals will no doubt on occasion desire specific means to these ends, such as particular foodstuffs, or playmates, or paths, what matters to their welfare will be standardly be the abstract goods themselves, which would be equally well-served by alternative and undesired realizations of the abstract goods.

So I think it is a mistake to reduce animal welfare directly to desire satisfaction. We will do better to detail an “objective list” of outcomes that contribute to the welfare of animals, whether or not these are desired[[24]](#footnote-24). Such a list might include lack of pain, thirst, and hunger; health; procreation; companionship; space to roam; and so on—the relevant outcomes will no doubt be different for different creatures. In many cases it will be true that animals desire these outcomes. But this isn’t what constitutes the value of these outcomes for the animals. Desire can come apart from welfare, as we have seen—it is neither sufficient nor necessary for morally valuable results.

Having said this, I don’t want to hold that desires are irrelevant to moral standing. We still face the issue of what can make creatures sufficiently *minded* that things will be good or bad *for them*. And one reasonable answer would be that this depends on their being psychologically sophisticated enough to have desires, to care one way or another about things. I said that some things can be good for animals even though they do not desire them as such. But it remains plausible that these things are good for them only in virtue of their ability to desire other things. For example, health will arguably only matter to creatures that are motivated to pursue ends, for without even this minimal purposiveness why should it matter *to them* whether or not they are healthy?

So in the end I agree with Carruthers and Kammerer that moral standing hinges on the possession of desires, even if not quite in the way they have in mind. But this now brings us to another issue, one which might have been bothering readers for some time. What exactly counts as having desires? I said above that we might equate desires with “any mental states which prompt behaviours apt to produce certain results, and perhaps which also reinforce those behaviours when those results are produced”. That formulation already introduces a question. Is reinforcement learning needed for bona fide desires or not? And answering this only leads to further questions. If not, is any minimal set of desires enough for mindedness, or must they cover a non-trivial range of outcomes? If yes, is simple reinforcement learning enough, or must animals also be capable of acquiring new desires from “response-outcome” learning?

Clearly enough, merely requiring “desires” will not by itself decide all questions of moral standing. This requirement can itself be understood more or less strictly. However, I shall

leave this issue up in the air for now. As I said at the end of the last section, I am not committed to all questions of moral standing having fully determinate answers. I shall return in due course to the question of whether they can be rendered more determinate than they have been by the discussion so far.

**11. What Really Matters**

As we have just seen, turning away from consciousness does not automatically render all questions of moral standing determinate. But it does at least have the virtue of allowing us to focus on what is important. To my mind, this is a crucial benefit of detaching moral standing from consciousness. On the majority alternative view that moral standing rests on consciousness, the issue of consciousness comes first, and moral standing then follows. However, debates about consciousness often hinge on factors that are themselves of no moral significance. Those who take the majority line are thus in danger of concluding that moral standing is conditional on matters which are morally unimportant.[[25]](#footnote-25)

Let me first illustrate the point with a well-known science fiction example. In the classic Star Trek episode, “The Measure of a Man”, the central issue is the moral status of Commander Data, who behaves like a human, down to a fine level of functional detail, but is an artificial construction made of non-biological materials. Is Data conscious? His non-standard make-up creates room to doubt this (see in particular Block 2002). Even so, a natural reaction, encouraged by the Star Trek scriptwriters, is that Data’s welfare is no less important than that of a human being. The position we have reached supports this judgement. Once we put consciousness to one side, it is hard to see how Data’s non-standard material make-up could possibly be morally relevant. After all, it doesn’t prevent him from otherwise being organized exactly like a human being, and so bearing all the same relations to the good things in life, such as avoidance of hunger and pain, friendship, autonomy, prosperity and so on. Doubts about Data’s moral standing can thus only hinge on uncertainty as to whether the fine details of his physiology suffice for consciousness or not. Addressing the moral issues without bringing in consciousness consequently puts us in a position to see why these doubts are misplaced.

We can make the same point about non-mammalian animals. Corvids and octopuses, for example, have been shown to possess psychological powers that are comparable to those of cats and dogs. At the same time, their independent evolution means that these cognitive powers are realised in different kind of brains. As we saw earlier, this creates room to doubt whether they share the consciousness enjoyed by mammals. Still, if moral welfare depends on mental adequacy, not consciousness, then these doubts cease to be morally relevant. It is the cognitive sophistication of the non-mammals that determines whether things can go well or badly for them, not the details of their brain structures.

So I regard it as a virtue of the no-consciousness approach to welfare that it shows that differences in the *realization* of animal abilities are of no moral significance. By putting consciousness to one side, it confirms that it is the psychological powers of animals that matters, not their different kinds of brains.

**12. Putting Consciousness to One Side**

Some readers might be wondering whether I really am entitled to put consciousness to one side. Perhaps they might be prepared to allow, in connection with Commander Data, say, that psychological sophistication in itself is morally significant, independently of issues of consciousness. But surely it is of *additional* moral significance whether Commander Data, or indeed corvids and octopuses, have conscious experience. By assuming consciousness is morally irrelevant, am I not assuming what I need to prove?

Not at all. I take myself already to have provided sufficient reason to put consciousness to one side in discussing the moral standing of non-human creatures. If some readers still feel that I am begging the question, I say that this can only be because the remain attached to dualist thoughts. As I shall show in this section, they must be assuming that conscious phenomenology depends on some special non-physical stuff, and that any moral standing associated with consciousness hinges on whether creatures possess this phenomenology-delivering mind-stuff.

Of course, from a physicalist point of view, no such mind-stuff exists. And so it is a mistake to think that moral standing is affected by its presence or absence.

This is not to deny that consciousness matters to the well-being of humans. Of course it does. The states that are conscious in humans have a distinctive significance for their well-being. I can say this because the phenomenal concept of consciousness works well enough in picking out a definite set of human states, namely those states which unanimously display all of the suite of properties associated with consciousness (cognitive accessibility, unlimited associative conditioning, distinctive neural processes, and so on). And this unanimity allows us to by-pass the issue of exactly which physical aspects of consciousness mean that conscious states play a special role in human moral welfare (Is it the cognitive accessibility, or the associative conditioning, or the neural processes?)

It is only when we move away from humans to creatures who display some of these diagnostic properties but not others that things get complicated. The simple ostensive concept of phenomenal consciousness proves insufficiently focused to decide whether these creatures are conscious. And so we can no longer blithely assume that their well-being hinges on whether or they are “conscious”. Instead we have to consider which of the diagnostic properties that are yoked together in humans (cognitive accessibility, unlimited associative conditioning, distinctive neural processes, and so on) really count for moral standing, and how far the non-human creatures possess those.

In the case of non-human creatures, what else is there to go on? Which is why I reject the thought that the consciousness of non-human creatures has moral significance *as well* as their psychological organization and abilities. This reaction can only be driven by the idea of some phenomenological mind-stuff which is additional to physical make-up.

Let me use the whale oil analogy again. In the old days, all agreed that oil was really valuable, and didn’t stop to consider how much this was due to its pragmatic role and how much, say, to its reinforcing the spiritual affinity between whales and humans. But the advent of new oil brings these issues to the surface. Some say the new stuff is just as valuable as the old, given its pragmatic equivalence. Others feel that it is of significantly inferior worth, because of the way its divorces the tribe from its traditional environmental context. A substantial debate ensues about the relative importance of the pragmatic and expressive properties of oil. All sensible people can see that there is no further issue here, apart from whether the expressive association with whales is worth preserving. Except, that is, for the oil dualists—who allow that the pragmatic and expressive features might well matter to some extent, but urge that it is surely of additional significance whether the new stuff contains *oil*. After all, they insist, it was the oil itself that made the old product valuable, and so the value of the new stuff can’t be assessed without deciding whether it too contains oil.

The analogy should be clear. Just as the new oil shares functional properties with the old stuff, but not the traditional connection with whales, so do Commander Data and the relevant non-mammals share psychological sophistication with humans, but not their underlying make-up. I have argued that this psychological sophistication suffices for the moral standing of Data and the non-mammals, and that their basic make-up has no moral significance. Similarly, some member of the indigenous tribe might maintain that the practical features of the new oil make it just as valuable as the old, on the grounds, say, that the symbolic connection with whales is a trivial matter. If someone now objects to my view of Data and the non-mammals that, even so, their moral standing might be diminished by their lack of consciousness, then I say the objector must be thinking of consciousness as something extra to the material realm—otherwise how could it make an extra moral difference, over and above that of all the material features in play? The objector’s view would be just like thinking that the presence or absence of *oil* makes an extra difference to the value of the new stuff, over and above the value of its practical and symbolic features.

**13. Zombology**

A significant number of recent writers seek to elucidate the moral significance of consciousness by appealing to intuitions about “zombies” and “Vulcans”. These are supposed creatures who are materially just like us humans but who are consciously different. Zombies lack all consciousness. Vulcans are conscious but their mental states lack “valence”—they are consciously aware of events but experience no enjoyment or suffering. The consensus in this literature is that the zombies’ lack of consciousness would disqualify them from any moral standing. Many hold that the Vulcans’ lack of emotional reactions would disqualify them too, or at least render them of lower moral standing than humans. Some make the point that, even if zombies and perhaps Vulcans lack moral standing, in that nothing would be good *for them*, they could still add impersonal intrinsic value to the universe, akin to the intrinsic value of mountains. (See for example Kriegel 2019 and this volume, Siewert 2021, Chalmers 2022 chs 17-18, Smithies forthcoming, Shepherd 2024.)

On the face of things, this literature is in tension with my view that we should put consciousness to one side in considering the moral status of non-human creatures. The whole point of the zombological analysis is to show how certain non-human creatures would lack moral standing specifically because of their lack of (or limited) consciousness. After all, the zombies and Vulcans are specified to be exactly like humans in all material respects, and to differ from them only at the conscious level. If we are led to conclude that, because of this, they lack moral standing while we humans don’t, consciousness can scarcely be irrelevant to moral standing.

Still, from a materialist point of view there is something puzzling about this literature. Intuition is supposed to tell us that zombies and Vulcans would have a different moral standing from humans. But materialism about the mind implies that creatures who are materially just like humans must necessarily be like them in all respects, including mental respects. How then can they be morally different? Intuition must have gone wrong somewhere.

In response to this concern, some might reply (Uriah Kriegel has replied to me in correspondence), that zombies and Vulcans are still *conceptually* possible, on orthodox version of materialism, even if they are metaphysically impossible. This is an implication of the nature of phenomenal concepts. If the referents of these concepts are effectively fixed ostensively—"*that* kind of state”—then the concepts themselves will carry no conceptual implications about their referents. In particular, they will leave it conceptually open whether or not their conscious states are metaphysically dependent on material facts. So there will be no conceptual contradiction in imagining beings who are materially just like humans but lack consciousness or valenced versions thereof.

Given this, materialists need not necessarily object to the use of zombie or Vulcan thought experiments to explore the importance of consciousness for human well-being. Imagining creatures who differ from us humans only in conscious respects might well be a good way to bring out exactly *how* consciousness matters to human welfare. Moreover, such thought experiments might also offer a useful resource for investigating exactly *why* those aspects so matter. Is it because of the intrinsic worth of positively valenced conscious states themselves, or is it due to some more complex circumstance for which those states are a necessary condition?[[26]](#footnote-26)

We need to be careful, however, about what we take zombie or Vulcan thought experiments to show. They might help us to see how and why consciousness matters to moral standing in humans, but they certainly don’t show that that it matters *rather than* physical processes in the brain. To repeat, from a materialist point of view consciousness is nothing over and above those processes, and so the idea that it can matter when its material constitution doesn’t is metaphysically incoherent.

In particular, these thought experiments cannot show that the consciousness of creatures like Commander Data and non-mammalian animals matters to their moral standing *in addition* to the way that they are materially constituted. I grant that it is intuitively seductive to think like this. But that is just the intuition of distinctness once more. Intuition might tell us that we shouldn’t put consciousness to one side in addressing these difficult cases. But that is like intuition telling us that being *made from oil* adds value to pragmatic usefulness and affinity with whales. It only metaphysically coherent if you are a dualist.

Perhaps it will be worth making clear that I do not say that creatures like Data and mentally developed non-mammals are definitely not conscious. I am happy to grant that creatures who definitely aren’t conscious will lack moral standing. Rather my view is that the difficult non-human cases are indeterminate. Data and the complex non-mammals are not definitely conscious, but not definitely not conscious either. This is precisely why we have to put the issue of consciousness to one side in thinking about their moral standing. They have some of the material features that characterise consciousness in humans, but not others. We know that beings with all those features have moral standing, and those with none lack it. But with those who have only some we have to dig deeper. Rather than just asking whether or not they are conscious, we need to consider directly whether the material features they do have are sufficient to ground moral standing.

**14. AI Welfare**

Let me now turn to the moral standing of artificial computer-based creatures. This topic raises any number of issues, so I shall restrict myself to some relatively brief comments. My central observation will be that my analysis of animal welfare will apply all the more strongly to artificial creatures. Our intuitive concept of consciousness has no hope of drawing clear liens in the artificial realm, and so the moral standing of artificial creatures needs to be analysed without bringing in questions of consciousness.

In my view, this topic is likely to become urgent sooner rather than later. Recent advances involving deep neural nets, especially with large language models like GPT-4, suggest that human-like artificial intelligence is not now far off. Of course systems like GPT-4 are themselves of limited interest. They are mere next-word predictors with a superficial ability to continue conversations. Still, their undoubted mastery of human languages is likely soon to lead to systems whose linguistic facility is disciplined in ways that will allow them to match, and no doubt surpass, human intellectual attainment. (Chalmers forthcoming, Shea 2023.)

A computer system with human-like intelligence is one thing. A candidate for moral standing is another. So far we have been dealing with animals. Animals have bodies. They live, they procreate, they die. They select actions that are well suited to their goals given their circumstances. At first pass, none of this applies to existing computer systems. They have no bodies and pursue no goals.

How deep are these differences? Whatever the limitations of current artificial systems, there is no barrier to their being designed to pursue goals. Where the cognitive goals of animals have been bequeathed to them by phylogenetic and ontogenetic selection, artificial systems can be given goals by their designers, in the sense of being set up to choose outputs that, in the light of their information and experience, are likely to achieve specified results. Moreover, if necessary, artificial systems can further be designed to acquire derivative proximal goals, as in machine reinforcement learning, in cases where it is computationally non-optimal to aim outputs directly at ultimate goals. On these points at least the differences between animals and artificial systems would seem likely soon to disappear.

The question of bodies cuts deeper. Computer systems are not realized in perishable biological bodies, but in settings of switches in electronic circuits. Even if such circuits come to be housed in mobile androids, the resulting systems would still be unlike animals in many morally significant ways. They would not obviously be subject to death. They could be duplicated, reprogrammed or repurposed. They could be paused, or switched off, and later reactivated.

All these points make for potential moral differences between machines and animals. To take just one example, where would responsibility lie if one of a suite of computer clones violated some behavioural restriction? And which systems should be disciplined? A natural reaction would be that the programme type itself was at fault, rather than the token malefactor, in a way that wouldn’t apply to animal misbehaviour.

Some might feel that disanalogies like these will suffice to exclude artificial creatures from the moral realm. I am not so sure. In any case, this issue is likely soon to become the focus of public debate. I am thinking here of the rise of “social AI”. Commercial organizations like Replika, Xiaoice and Baidu are already marketing “virtual companions”, systems designed provide individual friendship and advice to humans. The earliest versions were effectively little more than chatbots, but newer versions now incorporate large language models, and we can soon expect such systems to become more and more humanlike. After all, they will be designed precisely to develop relationships that mimic human interactions.

There are already reports of humans treating artificial companions as genuine subjects.. This will no doubt increase along with the sophistication of these systems. We can expect many of the humans who engage with these systems to become concerned about their welfare, at least to the extent of wanting to please them, to gain their approval, not to upset them, and so on—that is, they will come to treat their companions as moral patients. Alongside such concerns we might also expect to see generalised worries about the welfare of social companions. Those who come to engage emotionally with their own companions might well become exercised about the extent to which other owners mistreat their own virtual friends.

At the same time, we can also expect these worries to be met with a great deal of scepticism. Naysayers will insist the artificial companions are nothing but machines, little more than stochastic parrots, no more conscious than a thermostat, and so on.

As has been observed before, two dangers threaten here (Shevlin 2020, Schwitzgebel 2023). On the one hand, the friends of virtual subjects might be too quick to attribute moral standing to what are in truth nothing but dumb machines. After all, humans will apply the “intentional stance” on the most minimal grounds, attributing beliefs and intentions to any entities whose behaviour seems at all responsive to circumstances. So we can anticipate that at least some concerns about the welfare of artificial systems will be misguided.

On the other hand, anti-machine convictions might lead to the neglect of real interests of genuine moral patients. Many people are likely to assume from the start that no mere assembly of electronic switches could possibly possess the attributes required for moral standing. Even the most flexible machines, they will insist, are at bottom just a fancy input-output system with no claims on our moral attention. This faction is likely to persist in dismissing the moral claims of artificial systems however sophisticated they become.

I take it that we should aim to resist both dangers. We do not want to mistake superficial appearances for genuine subjecthood, but at the same time we should not rule out on a priori grounds the possibility that an artificial system might have moral standing. From a materialist point of view, it is hard to see what could justify denying moral standing to any machine, however accomplished, apart from a conviction that biological systems alone are blessed with some sort of non-material spirit.

Still, how might we decide which moral systems have genuine moral interests? The aim would be to transcend both species of prejudice and reach some unbiased conclusion. But it is not at all obvious how to arrive there. The dispute is all too likely to collapse into unreasoned conflict with no hope of rational resolution.

I have only one recommendation. We should *not* address the issue by asking whether the machines are conscious. That is the natural first response to the problem, as manifested by a growing body of literature. But addressing the issue in this way will not help us find a solution, and indeed is only too likely to pour fuel on the fire of conflicting prejudices.

In truth, there will be no determinate facts about the consciousness of humanlike machines. They will share some of the features characteristic of human consciousness, but not others, and nothing in our loose concept of consciousness will determine which are decisive. And it will be misguided to hope that future advances in the “science of consciousness” will solve the problem. Science might come to tell us more and more about human and artificial cognition, but it will be impotent, for all the reasons I have rehearsed, to draw any clear line between conscious and non-conscious systems.

So we will have no real alternative except to focus directly on the cognitive workings of artificial systems. Perhaps this will suffice to make progress. We saw earlier how putting consciousness to one side can help bring out what really matters morally. Maybe the reactions of humans who are fully informed about the nature of intelligent machines will in time settle down to some consensus. We might expect continued interaction with machines to play a role in overcoming initial biases. Consensus will not be guaranteed, but the history of human attitudes to colonised people, slaves and farmed animals offers some reason to expect progress.

Of course, even if fully informed humans who engaged with intelligent machines reach consensus about their moral standing, this would not automatically settle the issue. Perhaps there is some higher standard beyond informed human reactions to which moral judgements must answer. I myself see no reason to posit such a standard, and indeed regard the idea as itself morally dangerous. But this is not the place to pursue these meta-ethical issues.

Let me return briefly to an issue left hanging at the end of section 10. I observed there that, even if we agree that the moral standing of animals hinges on their having desires, this leaves open the level of sophistication required. Desires come in degrees. Here again I can only offer the prospect that we might be able reach new agreements about moral standing as we come to understand more about the workings of non-mammalian and indeed mammalian animals. As before, I regard the issue of consciousness as a distraction. We need to attend directly to the cognitive workings of the candidates, and perhaps that itself will suffice to show us whether or not they are moral patients.

No doubt any such determination of moral standing will still leave us with borderline cases. There will be animals, and also machines, that lie between those that definitely have moral standing and those that do not. Still, this is only to be expected. At least the indeterminacy will less than that delivered by the loose concept of consciousness. Putting consciousness to one side, and attending directly to cognitive workings, will allow us to discern the boundaries of moral standing more clearly.

**15. Taking It Back**

Let me conclude with one final point. Despite everything I have argued, I do not necessarily want people to stop saying that non-human creatures have moral standing if and only if they are conscious. I would have no objection to their continuing to uphold this equation—as long as they did this reading from left to right, so to speak, rather than in the opposite direction. What I have in mind is that people might first decide, as a result of moral reflection about the relevant material facts, which creatures have moral standing, and then on this basis deem them to be conscious. That is, they would ground consciousness in moral standing, rather than vice versa. On this option, we shouldn’t think that crabs have moral standing, where fleas do not, say, because the former but not the latter are conscious. It’s rather that we should judge the crabs to be conscious because they have moral standing, and the fleas not to be conscious because they don’t.

I would expect that, to the extent ordinary people were persuaded by the arguments of this paper, this is the line they would adopt. They wouldn’t come to doubt that cats are conscious, because of the looseness of the concept of consciousness. Rather they’d continue to view their cats as having a conscious life, in line with taking them to be creatures who deserve moral consideration.

This concession might seem to be taking back the claims I have made about the indeterminacy of our concept of consciousness. (“There's the bit where you say it and the bit where you take it back” Austin 1962 2.) Hasn’t my main point been that there is no right answer to the question of whether cats are conscious? However, it would not be hard to make the concession consistent with the main substance of my argument. This could be done in one of two ways.

First, we could stick with the idea that the ostensive concept of phenomenal is radically indeterminate in application to non-human creatures. (Do they have states like *these*? But like these in which respects?) However, we could then add that we would have good pragmatic grounds for *refining* the concept to include any non-human creatures that we independently judge on the basis of their psychology to have moral standing. On this option, the concept of consciousness would initially have all the indeterminacy that I have argued for, but it would progressively be refined in content, as we reach conclusions on independent grounds about which creatures have moral standing.

Second, we could take the view that the concept always did contain a moral dimension. That is, we could hold that it always applied to any states which matter morally in the way that *these* do (pointing to our own conscious states). On this option, the concept of consciousness would be designed from the start to play a moral role. Applying it to non-human creatures would be a way of recognising their status as beings with moral interests.

On this second option, we would be denying the letter of my thesis that the concept of consciousness is radically indeterminate. But this option would still be in the spirit of the rest of this paper. Along with the first option, it will deny that ostension of our own conscious states suffices to draw any lines in the non-human realm. It is only when a moral element is added that the indeterminacy is reduced. So this second option, just like the first option, will still rule out any arguments of the form: this non-human creature can be shown to be conscious on scientific or metaphysical grounds, therefore it has moral standing. Rather it would require us to proceed in the opposite direction, from an independent determination of moral standing to a conclusion about consciousness.

The position we have now arrived at raises interesting further questions. For one, I have now in effect proposed that *consciousness* might be an essentially “thick” concept. For note that, while I am now suggesting that conscious facts can be grounded in moral facts, and moral facts can be grounded in material facts, I resist skipping the middle term and having conscious facts grounded directly in material ones. This raises questions about why moral judgement should play an ineliminable role in this kind of determination. (See Väyrynen 2021 especially section 2.)

For another, my final position has affinities with the Wittgensteinian view that the recognition of animals as minded creatures deserving of moral concern is a matter of affective human responses rather than scientific facts. Writers like Cora Diamond and Rai Gaita have urged that we cannot separate questions of animal psychology from our natural human inclinations to include certain animals in our moral circle (Diamond 1978, Gaita 1999). This raises questions, at both the ethical and meta-ethical level, about the warrant for so according a special status to human beings over other creatures.

But let me stop here. This paper has already addressed many questions. It is not a bad thing if it raises some more.[[27]](#footnote-27)

**References**

Baars, B. 1988 A Cognitive Theory of Consciousness Cambridge: Cambridge University Press

Bayne, T. and Shea, N. 2020 “Consciousness, Concepts and Natural Kinds” *Philosophical Topics* 48 65-83

# Birch, J. 2022 “The Search for Invertebrate Consciousness” *Nous* 56: 133-53

Birch, J. forthcoming *The Edge of Sentience*Oxford: Oxford University Press

Birch, J., Schnell, A. & Clayton, N. 2020 “Dimensions of Animal Consciousness” *Trends in Cognitive Sciences* 24: 789–801

Block, N., 1980 “Troubles With Functionalism” in Block, N. ed Readings in the Philosophy of Psychology Vol 1 Cambridge Mass: Harvard University Press.

Block, N. 1995 “On a Confusion about a Function of Consciousness” *Behavioral and Brain Sciences*  18: 227-47

Block, N. 2002 “The Harder Problem of Consciousness” *Journal of Philosophy* 91: 391-425

Brown, C. and Papineau, D. forthcoming “Illusionism and A Posteriori Physicalism; No Fact of the Matter”

# Butlin, P., Long, R. et al 2023 “Consciousness in Artificial Intelligence: Insights from the Science of Consciousness” <https://arxiv.org/abs/2308.08708>

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

Carruthers, P. and Gennaro, R. 2020 "Higher-Order Theories of Consciousness" in Zalta, E. and Nodelman, U. (eds) The Stanford Encyclopedia of Philosophy (Fall 2023 Edition)

Carter, O. et al "Conscious Machines: Defining Questions" *Science* 359: 400

Chalmers, D. 2003 **“**The Content and Epistemology of Phenomenal Belief**”** in Smith, Q. and Jokic, A. (eds) *Consciousness: New Philosophical Perspectives* Oxford University Press 220-72

Chalmers, D. 2022 *Reality+: Virtual Worlds and the Problems of Philosophy* New York: Norton

Chalmers, D. “Could a Large Language Model be Conscious?” unpublished

Crisp, R. 2021 "Well-Being" in Zalta, E. and Nodelman, U. (eds) The Stanford Encyclopedia of Philosophy (Winter 2021 Edition)

Dawkins, M. 2012 *Why Animals Matter: Animal Consciousness, Animal Welfare, and Human Well Being* Oxford: Oxford University Press

Dawkins, M. 2021 *The Science of Animal Welfare: Understanding What Animals Want* Oxford: Oxford University Press.

Dehaene, S., Lau, H. and Kouider, S. 2017 "What is Consciousness and Could Machines Have It?" *Science* 358: 486-492

Dennett, D. 1978 “Why You Can’t Make a Computer That Feels Pain” in his *Brainstorms* Cambridge, Mass: Bradford Books

Diamond, C. 1978 “Eating Meat and Eating People” *Philosophy* 53: 465-79

Feigl, H. 1958 “The ‘Mental’ and the ‘Physical’” in Feigl, H., Scriven, M. and Maxwell, G. (eds) *Minnesota Studies in the Philosophy of Science Volume II* Minneapolis: University of Minnesota Press

Gaita, R. 1999 *A Common Humanity: Thinking about Love & Truth & Justice* Melbourne: Text Publishing

Ginsburg, S. and Jablonka, E. 2019 *The Evolution of the Sensitive Soul: Learning and the Origins of Consciousness* Cambridge, Mass: MIT Press

Godfrey-Smith, P. 2016 “Mind, Matter, and Metaboilism” *Journal of Philosophy* 113; 481-506

Godfrey-Smith, P. 2020 *Metazoa* London: Collins

Godman, M., Mallozzi, A., & Papineau, D. 2020 “Essential Properties are Super-Explanatory: Taming Metaphysical Modality” *Journal of the American Philosophical Association* *6*: 316–34

Goodale, M. and Milner, A. 1992 “Separate Visual Pathways for Perception and Action” *Trends in Neurosciences* 15: 20-5

# Hall, G. 2023 “Is Consciousness Vague?” *Australasian Journal of Philosophy* 101:670-84

Johnson, L. 1993 A Morally Deep World: An Essay on Moral Significance and Environmental Ethics, Cambridge: Cambridge University Press

# Kammerer, F. 2019 “The illusion of conscious experience” *Synthese* 198:845-66

Kammerer, F. 2022 “Ethics without Sentience: Facing up to the Probable Insignificance of Phenomenal Consciousness” *Journal of Consciousness Studies* 29: 180-204

Kriegel, U. 2019 “The Value of Consciousness” *Analysis* 79: 503-520

Kriegel, U. “The Value of Consciousness to the One Who Has It” this volume

Lau, H. 2022 *In Consciousness We Trust: The Cognitive Neuroscience of Subjective Experience* Oxford: Oxford University Press.

Lee, A. 2023 “Degrees of Consciousness” *Nous* 57:553–575.

Lee, G. 2019 “Alien Subjectivity and the Importance of Consciousness” in Pautz, A. and Stoljar, D. (ed) *Blockheads! Essays on Ned Block’s Philosophy of Mind and Consciousness* Cambridge Mass: MIT Press

Levin, J. "Functionalism" in Zalta, E. and Nodelman, U. (eds) The Stanford Encyclopedia of Philosophy (Summer 2023 Edition)

Overgaard, M. and Fazekas, P. 2016 “Can No-Report Paradigms Extract True Correlates of Consciousness?” Trends in Cognitive Sciences 20: 241–242

Papineau, D. 1993 *Philosophical Naturalism* Oxford: Basil Blackwell

Papineau, D. 2001 “The Rise of Physicalism” in Gillett, C. and Loewer, B. (eds) *Physicalism and its Discontents* Cambridge: Cambridge University Press.

Papineau, D. 2002 Thinking About Consciousness Oxford: Oxford University Press

Papineau, D. 2003 “Could There Be a Science of Consciousness?” *Philosophical Issues* 13: 205-20

Papineau, D. 2007a “Phenomenal and Perceptual Concepts” in Alter, T. and Walter, S. (eds) *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism* Oxford University Press.

Papineau, D. 2007b “Kripke’s Proof is Ad Hominem not Two-Dimensional” *Philosophical Perspectives* 21: 475-94

Papineau, D. 2011 “What Exactly is the Explanatory Gap?” *Philosophia* 35: 5-19

Papineau, D. 2020 “The Problem of Consciousness” in Kriegel, U. (ed) *The Oxford Handbook of the Philosophy of Consciousness* Oxford: Oxford University Press 14-36

Papineau, D. 2022a “Swampman, Teleosemantics and Kind Essences” *Synthese* 200: 1-19

Papineau, D. 2022b “Review of Michael Tye’s *Vagueness and the Evolution of Consciousness: Through the Looking Glass*” *Notre Dame Philosophical Reviews* <https://ndpr.nd.edu/reviews/vagueness-and-the-evolution-of-consciousness-through-the-looking-glass/>

Schwitzgebel, E. 2023 "The Full Rights Dilemma for A.I. Systems of Debatable Personhood" *Robonomics* 4: 23

Shea, N. 2012 “Methodological Encounters with the Phenomenal Kind” *Philosophy and Phenomenological Research* 84: 307-44

Shea, N. 2023 “Moving Beyond Content-Specific Computation in Artificial Neural Networks” Mind & Language 38: 156-77

Shepherd, J. 2024 “Sentience, Vulcans, and Zombies: The Value of Phenomenal Consciousness” *AI and Society* 1-11

Shevlin, H. forthcoming “How Could We Know When a Robot was a Moral Patient?” Cambridge Quarterly of Healthcare Ethics

Siewert, C. 2021. “Consciousness: Value, Concern, Respect” *Oxford Studies in Philosophy of*

*Mind* 1: 3-40

Smart, J.J.C. 1959 “Sensations and Brain Processes” *Philosophical Review*, 68: 141–56

Smithies, D. forthcoming “Affective Consciousness and Moral Status” *Oxford Studies in Philosophy of Mind* 5

Taylor, P. 1986 Respect for Nature Princeton: Princeton University Press

Tononi, G. 2008 “Consciousness as Integrated Information: a Provisional Manifesto” Biological Bulletin 215: 216–42

Tsuchiya, N., Wilke, M., Frässle, S., and Lamme, V. 2015 “No-Report Paradigms: Extracting the True Neural Correlates of Consciousness Trends in Cognitive Science 19, 757–770

Väyrynen, P. "Thick Ethical Concepts" in Zalta, E. (ed) The Stanford Encyclopedia of Philosophy (Spring 2021 Edition)

1. Actually, quite a few say moral standing hinges on “sentience”, rather than “consciousness”, in order to make it clear that it does not demand *self*-consciousness or anything like that. However, I shall stick to “consciousness” in order to maintain contact with the main philosophical literature on the topic. [↑](#footnote-ref-1)
2. Modern defences of materialism start with Feigl 1958 and Smart 1959. I contribute to the debate in Papineau 2001, 2002 Chapter 1 and Appendix. [↑](#footnote-ref-2)
3. Barrs 1988 [↑](#footnote-ref-3)
4. Carruthers and Gennaro 2020 [↑](#footnote-ref-4)
5. Tononi 2008 [↑](#footnote-ref-5)
6. Lau 2022 [↑](#footnote-ref-6)
7. Ginsburg and Jablonka 2019 [↑](#footnote-ref-7)
8. Birch et al 2020 [↑](#footnote-ref-8)
9. Dehaene et al 2017, Butlin, Long et al 2023 [↑](#footnote-ref-9)
10. Carter et al 2018 [↑](#footnote-ref-10)
11. See for example Chalmers 2003, Papineau 2007 [↑](#footnote-ref-11)
12. Goodale and Milner 1992 [↑](#footnote-ref-12)
13. Dennett 1978 [↑](#footnote-ref-13)
14. Tsuchiya et al 2015 [↑](#footnote-ref-14)
15. Overgaard and Fazekas 2016 [↑](#footnote-ref-15)
16. Jonathan Birch (2022, forthcoming) combines Shea’s appeal to natural kinds with Ginsburg and Jablonka’s 2019 focus on unlimited associationist learning to develop a “theory-light” approach to animal consciousness. I take the issues I develop in this section to pose problems for this approach. [↑](#footnote-ref-16)
17. The non-mammals pose the familiar choice between “role” and “realizer” functionalism (Levin 2023). Does conscious *pain*, say, require only a state with a certain causal profile, or does it also depend on details of brain chemistry? The longevity of this debate testifies to its empirical intractability. [↑](#footnote-ref-17)
18. If certain mental states display a cluster of associative learning and stimulus recognition features across the whole animal kingdom, then won’t this itself constitute them as a natural kind, and won’t this kind have some unifying property, which might then ground consciousness, in line with Shea’s approach? However, this line of thought will lead us to the wrong kind of underlying explanatory property. The relevant cross-phylum kind will be a *functional* kind whose similarities are due to convergent evolution, and the unifying property for this kind will then be the common past selective pressures that have evolved similar structures despite the lack of a common neural basis. I take it that this kind of historical property is not a candidate for constituting consciousness. Consciousness is here-and-now. Whether or not a creature is conscious depends on its local properties, not on which long-past historical processes were responsible for them. (I do not have any general objection to mental kinds having historical essences. The point is peculiar to consciousness. Thus in Papineau 2022a I argue that *representation* is a functional kind and that representers must therefore have a specific kind of selectional history, and I use this point to argue that Swampman does not represent, just as orthodox teleosemantics has always maintained. However, I would not say that Swampman’s lack of selectional history deprives him of consciousness.) [↑](#footnote-ref-18)
19. How far does scientific research into the “neural correlates of consciousness” depend on an implicit assumption of some extra non-physical mind-stuff? Without this assumption, there seems nothing more for science to do than chart the causal roles of different kinds of brain processes, including their accessibility to central cognition. However, it would take us too far afield to pursue this issue here. [↑](#footnote-ref-19)
20. Andrew Lee (2023) defends the view that consciousness, like mass or size, is a quantity that comes in degrees. This seems reasonable enough, but does not alter the fact that it would be arbitrary to fix some definite point after which the consciousness quantity ceases to be zero. [↑](#footnote-ref-20)
21. I have urged that consciousness is indeterminate since my first writings about consciousness (see especially 1993, 2002, 2003). Peter Carruthers (2019) similarly argues that the concept draws no sharp boundaries outside humans. Michael Tye (2021) observes that orthodox physicalism must allow borderline cases of consciousness and responds by embracing a version of panpsychism. Papineau (2022b) argues that this does not help. Geoffrey Hall (2023) shows that the difficulty of imagining borderline cases of consciousness does not mean that there aren’t any. [↑](#footnote-ref-21)
22. Kammerer is in fact an illusionist, holding that our concept of consciousness a priori precludes any material referent (Kammerer 2019). But in his 2022 paper on animal ethics he assumes only that the concept fails to draw clear lines in the animal realm. For what it is worth, I regard the dispute between illusionism and materialism about consciousness as largely terminological (see Brown and Papineau forthcoming), but find it perverse, for pragmatic reasons, to end up saying that that consciousness does not exist. (That would be like the indigenous people in my oil fable concluding that “oil does not exist” on the uncompelling grounds that the concept of oil demands some non-physical realizer.) [↑](#footnote-ref-22)
23. Marion Dawkins has also argued that animal welfare is best decided by asking about the satisfaction of animal desires conceived non-phenomenologically (2012, 2021). Her rationale, however, is epistemological rather than ontological. She assumes that animal consciousness is determinate, but doubts that we have effective ways of finding out about it, and for that reason urges that we base decisions about animal welfare on other grounds. It should be noted that Dawkins also regards health as important for animal welfare, along with desire satisfaction. [↑](#footnote-ref-23)
24. See Parfit 1984 Appendix I. [↑](#footnote-ref-24)
25. François Kammerer (2022) emphasizes this point. I take Geoffrey Lee (2019) also to be sympathetic, given his deflationary view that consciousness is not a significant property. (Where I deny that the concept of consciousness refers determinately, Lee assumes that it does, but then concludes that the property picked out has no special significance. As I see it, our disagreement on the semantics does not stop us arriving at similar conclusions about the normative insignificance of consciousness.) [↑](#footnote-ref-25)
26. See in particular Uriah Kriegel’s contribution to this volume. It is worth noting that Kriegel fails to settle on a definite rationale for denying moral standing to zombies, and indeed entertains the possibility that this denial would be mistaken. [↑](#footnote-ref-26)
27. This paper started life as a talk at a conference on “ChatGPT and Other Creative Rivals” at the Institute pf Philosophy in London in 2023. Since then versions of it have been given at the University of Sheffield and Edinburgh, at the Digital Minds group in Oxford, and at Uriah Kriegel’s seminar series on the Value of Consciousness. I am grateful to all who made comments on those occasions. I would also like to thank Patrick Butlin and Brad Saad for written responses, and Jonathan Birch, Emma Borg, Alex Grzankowski, Anandi Hattiangadi, Geoff Lee, Nick Shea, Henry Shevlin and James Stazicker for helpful conversations. [↑](#footnote-ref-27)