The Philosophical Quarterly Vol. 0, No. 0 *2022*
ISSN 0031-8094 https://doi.org/10.1093/pq/pqab072

MATERIALISM AND THE MORAL STATUS OF ANIMALS

By Jonathan Birch

Consciousness has an important role in ethics: when a being consciously experiences the frustration or satisfaction of its interests, those interests deserve higher moral priority than those of a behaviourally similar but non-conscious being. I consider the relationship between this ethical role and an a posteriori (or 'type-B') materialist solution to the mind-body problem. It is hard to avoid the conclusion that, if type-B materialism is correct, then the reference of the concept Phenomenal Consciousness *is radically indeterminate between a neuronal-level property that is distinctive to mammals and a high-level functional property that is much more widely shared. This would leave many non-mammalian animals (such as birds, fish, insects and octopuses) with indeterminate moral status. There are ways to manage this radical moral indeterminacy, but all of these ways lead to profoundly troubling consequences.*

Keywords: consciousness, materialism, moral indeterminacy, ethics, animals.

Consciousness matters in ethics. To give one example: our moral obligations towards a patient in a chronic vegetative state plausibly differ from our obligations towards a patient who is minimally conscious (Kahane and Savulescu 2009). To give another example: if we can show that animals of a particular type (e.g. fish, crabs, insects, octopuses) have conscious experiences, this can form the basis of an argument that we have stronger moral obligations towards them than we might previously have supposed.

'Consciousness' is a notoriously ambiguous term, so what *type* of consciousness matters in case like these? It is not simply wakefulness, defined functionally in terms of sleep-wake cycles, because patients in a vegetative state have sleep-wake cycles, as do fish, crabs, insects and octopuses. There is no serious debate about this, and yet it does not settle the substantive ethical questions. There is more room for debate about so-called 'access consciousness' in Block's (1995) sense: the functional availability of information for reasoning and the rational control of action (including speech). In all of the above cases, there may be some availability of information for reasoning and rational control, or there may be none, depending on whether the being is capable of reasoning or rational action at all. But settling this issue would leave a big ethical question

unresolved. The question 'But do they *feel* any of this? Do they *experience* it?' would still arise.

I regard this as a question about *phenomenal consciousness* in Block's (1995) sense. A state is phenomenally conscious if and only if there is something it's like to be in that state. It is important to know, for unresponsive patients with brain injuries, whether there is something it's like to be them—whether they *experience* what is happening to them, whether they *feel* it. The same question matters in relation to non-human animals, particularly given our tendency to treat animals that are evolutionarily distant from us (such as crabs and lobsters) in ways that are ethically dubious if the animal does feel what is being done to it. In what follows, I will use the term 'conscious being' as a shorthand for a being capable of forming phenomenally conscious states.

Given the ethical importance of phenomenal consciousness, there are possible solutions to the mind-body problem that would create trouble for practical ethics. This is most obviously true of eliminative materialism about consciousness, also known (in recent literature) as strong illusionism.[1] This view holds that phenomenal consciousness does not exist. If strong illusionism is correct, and if phenomenal consciousness plays an important role in ethics, ethics may need significant revision. Kammerer (2020) has called this the 'normative challenge' for strong illusionism. But perhaps ethicists should not be unduly worried about this, since the challenge can be avoided by denying strong illusionism.

My focus here will be on a significantly more popular approach to the mind-body problem: a posteriori materialism, or 'type-B' materialism in Chalmers' (2010) terminology. The type-B materialist holds that, although there is an epistemic gap between physical and phenomenal truths that is exploited by anti-materialist arguments, this epistemic gap can be explained without positing any ontological gap between physical and phenomenal facts. In a particularly influential version of the view, the epistemic gap is to be explained by appeal to phenomenal concepts. These are concepts we use for thinking about our experiences, and they are posited to be radically different from the concepts we use to think about the physical world, with the result that there are few a priori connections (or none at all) between phenomenal concepts and non-phenomenal concepts (Carruthers 2000; Loar 1990; Papineau 2002). In short, we can be conceptual dualists without being ontological dualists. This has come to be known as the 'phenomenal concept strategy' (Balog 2012; Stoljar 2005). My aim here is not to motivate type-B materialism; it is enough to note that, at least in the eyes of its proponents, it represents the best way to reconcile the existence of consciousness with a naturalistic worldview on

[1] I am using the term 'strong illusionism' here in the sense of Chalmers (2018). Frankish (2016) uses the term in a slightly different sense.

which there is no credible entry point for any form of non-physical causation in the workings of the brain (see e.g. Papineau 2002: Appendix).

Since the type-B materialist does not deny the existence of phenomenal consciousness, it is not so obvious that that the view holds revisionary consequences for ethics. But I will argue that it does. I will argue that, if type-B materialism is true, then, given a plausible account of the relationship between phenomenal consciousness and moral status, facts of great ethical significance turn out to be indeterminate.

The next section discusses the ethical significance of phenomenal consciousness in greater depth, culminating in an attempt at a precise formulation of an attractive idea I call the 'higher priority principle'. The subsequent section argues that, if type-B materialism is true, then the reference of the concept Phenomenal Consciousness is indeterminate between a neuronal-level property that is distinctive to mammals and a high-level functional property that is much more widespread (here and throughout, I use capitalization to indicate that the concept Phenomenal Consciousness, and not the property of phenomenal consciousness, is at issue). These premises combine to yield the disturbing conclusion that, in many cases, it is indeterminate whether or not an animal's interests deserve higher priority than those of a behaviourally similar but non-conscious being. I then consider some ways to manage this radical moral indeterminacy—and find all the options troubling.

I. THE HIGHER PRIORITY PRINCIPLE

A being with *moral status* has at least some interests that matter morally *in their own right*, not just because they matter to some other being. I want to leave open the possibility that non-conscious beings can have moral status. Some maintain, for example, that plants, or foetuses not yet capable of forming conscious states, or unconscious patients who will never regain consciousness, have moral status. On some ethical theories, such as hedonic utilitarianism (Singer 1995) and Regan's (2004) animal rights theory, the capacity to have conscious experiences is a necessary condition for moral status. However, I want to avoid assuming any particular ethical theory.

To remain neutral on the question of whether a non-conscious being can have moral status, I want to focus on an idea I think can be a point of wide agreement. This is the idea that a conscious being's interests ought to factor into moral deliberation in a distinctive way, such that these interests receive *higher priority* than those of a behaviourally similar but non-conscious being, if we allow that the non-conscious have interests at all. When we find evidence that a patient thought to be permanently unconscious is in fact having conscious experiences at least some of the time, their interests (e.g. in adequate nutrition and hydration) appropriately receive higher moral priority than they otherwise

would. If forced to choose between the interests of a permanently unconscious patient and those of a conscious patient, we should prioritize the interests of the conscious. Likewise, when we find evidence that an animal thought to be wholly unconscious has conscious experiences, their interests (e.g. in humane treatment and good welfare) appropriately receive higher moral priority than they otherwise would. If forced to choose between the interests of a conscious animal or a non-conscious animal, we should prioritize the conscious.

The phrase 'higher priority' is intended to be compatible with the behaviourally similar non-conscious being having no morally significant interests at all—but also compatible with it having interests that deserve lower priority. It is also intended to be neutral between the interests of a conscious subject deserving *lexical priority* over the interests of the non-conscious (so that the interests of the non-conscious can never outweigh the interests of the conscious) and the interests of the former merely deserving *greater numerical weight* (a finite 'consciousness multiplier'), so that a weak interest of a single conscious subject might still be outweighed by the aggregated interests of many non-conscious entities.

To formulate the idea here as precisely as possible, we need to ask whether it is phenomenal consciousness *as such* or a *special type* of phenomenally conscious state that explains the moral claim to higher priority. Various authors (e.g. DeGrazia 1996; Korsgaard 2018; Shepherd 2018; Singer 2011) have proposed that it is not conscious experience as such but *valenced* conscious experience that matters. Valenced conscious experiences are experiences that feel bad or feel good. Negatively valenced experiences include pain, pleasure, distress, anxiety, boredom, tiredness, hunger and thirst. These experiences typically motivate actions to alleviate them. Positively valenced experiences include pleasure, joy, warmth, comfort, satiety and excitement. These experiences typically motivate actions to sustain them. Any capacity for valenced conscious experience will imply interests to escape, or to sustain, certain types of situation, and these interests matter morally.

One might wonder: Could valence alone be enough to explain the moral claim to higher priority, independently of phenomenal consciousness? This depends, first of all, on whether there can be valence without phenomenal consciousness. On one account of valence (that of Carruthers 2018), valence is the non-conceptual representation of value, and it is plausible that a state may represent value non-conceptually without being phenomenally conscious. Indeed, it seems necessary to posit non-conscious representations of value in order to explain the possibility of subliminal motivation (Pessiglione *et al.* 2007) and subliminal instrumental conditioning (Pessiglione *et al.* 2008, though cf. Skora *et al.* 2021). For example, Pessiglione *et al.* (2007) presented evidence that subjects find larger sums of money more strongly motivating than smaller sums, even if the amounts of money are presented subliminally. This suggests

(without conclusively showing) that the larger sums are assigned higher value than the smaller sums by unconscious motivational processes.

Granting that there can be valenced states that are not phenomenally conscious, it makes sense to ask whether a capacity to form valenced mental states might already suffice to explain the higher priority we initially thought to be dependent on consciousness. But I contend that a capacity for representing value would *not* suffice if the representations were always non-conscious. A capacity for representing value leaves open the question that lies at the heart of the claim to higher priority: 'But do they *feel* any of this? Do they *experience* it?' A reinforcement learning algorithm represents value, but it is usually taken to do so non-consciously, and in a way that intuitively confers no moral significance on its interests. Tomasik (2014) has argued that reinforcement learning algorithms do possess moral status (inspiring an organization called 'People for the Ethical Treatment of Reinforcement Learners'), but Tomasik rests his case on the idea that algorithms may have a capacity for phenomenal consciousness, not on the idea that non-conscious representation of value is already sufficient.

One might also wonder: Could phenomenal consciousness alone be enough to explain the moral claim to higher priority, independently of valence? At least in principle, there can be phenomenal consciousness without valence: experiences that feel neither bad nor good. It is not clear that humans can have such experiences: our overall conscious state arguably always contains an element of mood. But we can conceive easily enough of a being that has a subjective point of view on the world in which non-valenced states feature (it consciously experiences shapes, colours, sounds, odours, etc.) but in which everything is evaluatively neutral. Would such a being have the moral claim to greater weight associated with conscious experience?

Chalmers (quoted in Wiblin *et al.* 2019) has offered the example of a Vulcan. The original Vulcans in Star Trek are not wholly without valenced experiences, but we can conceive of a 'philosophical Vulcan' in which valenced experience is dialled down to nothing, while leaving conscious but valence-free perceptual experience, conscious thought, imagination, and episodic memory in place. Carruthers (2005) also considers such a being, which he names 'Phenumb'. Intuitively, a philosophical Vulcan has morally significant interests: it would be wrong to destroy such a being for no reason at all. Moreover, it seems intuitively wrong to give lower priority to its interests than to those of a human simply because of its dialled-down valence.

In opposition, Lee (2019a) offers the example of an animal that experiences a maximally simple non-valenced experience, such as an experience of slight brightness. The example is reminiscent of Ginsburg and Jablonka's (2007: 220) example of a being who experiences only 'white noise'—Ginsburg and Jablonka speculate that the first conscious experiences in the earliest nerve nets were something like this. Is the presence of conscious experiences of slight

brightness, or white noise, enough to justify giving higher moral priority to the animal's interests, relative to those of a behaviourally similar but white-noise-free animal? Plausibly, it is not.

Can we reconcile the conflicting intuitions elicited by these cases? What the philosophical Vulcan shows us, I suggest, is that morally significant interests can, in principle, be grounded independently of valence. An autonomous rational being capable of reflectively endorsing goals and projects has such interests, whether or not it has experiences of frustration, joy (and so on) associated with the success or failure of those projects. Note, however, that the Vulcan is still registering the promotion or frustration of its interests in experience. I propose that the step up in moral status associated with phenomenal consciousness is the change that comes when events that promote or thwart a being's interests are registered in experience. Valence has a special importance because it enables the conscious registration of one's interests being promoted or thwarted in beings who lack rational agency.

Rational beings who can reflect on their experiences, as we can, may endorse the pursuit of pleasure and other positive experiences (such as aesthetic experiences) for their own sake. For such beings, valenced experience acquires a *second* type of ethical significance. It is no longer just a currency in which interests register consciously; it is also *constitutive of* some of those interests. Consider, by way of analogy, the difference between someone who uses money as a currency and someone who comes to value money for its own sake. But my proposal is that even in beings with no capacity to reflect on their experiences, valenced experience matters by virtue of registering interests.

The overall picture, then, is one on which a capacity for phenomenal consciousness as such is a necessary condition for one's interests deserving higher moral priority. In addition, the ability to register the promotion or frustration of one's interests in experience, either in the form of valence or the rational endorsement of goals, must also be in place.

These ideas are brought together in the following principle:

> **Higher priority principle:** If a being has interests that register in conscious experience (e.g. by means of experiences with positive or negative valence), then these interests deserve higher moral priority than those of a behaviourally similar but non-conscious being.

To be clear, this is a proposal about the ethical significance of conscious experience, but it does not by itself offer a full *explanation* of that significance. One is admittedly still left wondering: why is it that *conscious* registration of one's interests being promoted or thwarted has a special significance that does not attach to non-conscious forms of registration? I leave this question open here. I also leave open the possibility that there is no deeper explanation—that the special significance of consciously experienced interests is ethical bedrock.

A consequence of the higher priority principle is that the question of which animals are capable of phenomenal consciousness has great ethical significance. This is because, for many animals, valence is in place: we already have good evidence of their ability to form representations of value and disvalue that guide flexible decision-making. The question that remains is whether the value and disvalue is consciously experienced.

To illustrate, consider Crook's (2021) recent study of responses to injury (injection of acetic acid) in Bock's pygmy octopus (*Octopus bocki*). Injured octopuses showed directed grooming at the site of the acetic acid injection that was abolished by a local anaesthetic, lidocaine. More than this, they came to prefer chambers in which they had been placed after receiving lidocaine, and disprefer chambers in which they had received an injury. This type of evidence is widely regarded in animal welfare science as evidence of pain, an exemplar of a valenced experience.

There is evidence of this general type (reviewed in Sneddon *et al.* 2014) for mammals, birds, reptiles, fish, cephalopod molluscs and decapod crustaceans. But this evidence invites the objection that to show valenced *experience*, it is not enough to show mere representation of value and disvalue—you also have to show that these representations are consciously experienced (Dawkins 2021; Paul *et al.* 2020). This task presents serious methodological challenges that are not the topic of this article, but we can hope that the science of consciousness will one day be able to overcome them, and that in the meantime it will produce relevant, albeit inconclusive evidence (for discussion of the methodological challenges, see Birch 2020). This hope, however, rests on there being a determinate fact of the matter to be found.

II. MILD AND RADICAL INDETERMINACY

Current scientific theories that seek to identify phenomenal consciousness with a cognitive/neurobiological natural kind tend to make use of concepts that are vague, in the sense of allowing for borderline cases.[2] To illustrate, suppose we think phenomenally conscious states are patterns of thalamocortical activity supported by pyramidal neurons in layer 5 of the neocortex (a hypothesis set out by Aru *et al.* 2019). Various neurobiological concepts are in play in this hypothesis: thalamus, neocortex, pyramidal neuron, layer 5. If we could see the evolution of these traits unfolding over time, we would expect to see borderline cases of all of these concepts. For example, when the laminated

[2] Tye (2021) and Schwitzgebel (2021) have made similar points. Tye goes on to argue that this vagueness has surprisingly radical metaphysical consequences—and even provides some support for a view with elements of panpsychism. If this is right, then I am wrong to label it 'mild'. But that is an issue for another occasion.

structure of the mammalian neocortex was in the process of evolving, it seems likely that there would have been borderline cases between laminated and non-laminated cortices. Moreover, although we should acknowledge that, in principle, current theories of consciousness could one day be superseded by theories that avoid vague concepts, it seems unlikely that the vagueness of current theories is a mere historical accident or a mark of the immaturity of consciousness science. It seems more likely to be a reflection of the gradual nature of evolution, in which new mechanisms and structures evolve through tiny incremental changes to their precursors without sharp transitions, making it entirely appropriate to use vague concepts to pick out those mechanisms and structures.

This vagueness is in tension with the intuition that phenomenal consciousness must always be determinately on or off, with no borderline cases, an intuition Anthony (2006) has called the 'intuition of sharpness' (see also Simon 2017). I take it that a materialist who wants to defend the correctness and completeness of an existing scientific theory of consciousness—or any successor theory that still employs vague concepts of gradually evolved cognitive/neurobiological kinds—should reject the intuition of sharpness and accept that borderline cases of conscious experience are possible (see also Godfrey-Smith 2020; Lee 2019b; Schwitzgebel 2021). To the extent that this is counterintuitive, it is just part of the price of the view: a sense in which it is counterintuitive, to be added to the counterintuitiveness (for some) of rejecting the possibility of zombies.

Several authors have noted that sorites-style vagueness, combined with a close connection between phenomenal consciousness and moral status, threatens to lead to borderline cases of moral status (Cutter 2017; Dunaway 2016; Godfrey-Smith 2020). However, there is no particular reason to expect sorites-style vagueness to lead to a widespread meltdown of our ethical deliberations regarding animals. First, there may not be any extant species occupying the borderline region for any of the relevant kinds (for example, all extant mammals determinately have a neocortex). Secondly, even if there are some extant borderline cases, we can reasonably hope that they will represent a very small fraction of species. This is implicit in the idea that neurobiological and cognitive kinds are natural kinds that 'carve nature at the joints', with most cases lying between the joints. If we found a putative cognitive/neurobiological kind such that a vast majority of animals were borderline cases with respect to that kind, that would give us reason to doubt that the putative kind did in fact carve nature at the joints—and a reason to look harder for more refined natural kinds. To be sure, sorites-style vagueness regarding moral status has interesting implications for meta-ethics, as Cutter (2017) has noted, since meta-ethical positions incompatible with vagueness about moral obligation will also be incompatible with plausible forms of materialism. Yet in so far as it need

not threaten our practical deliberations about what to do outside of a small number of cases, sorites-style vagueness is a mild form of indeterminacy.

Several prominent defenders of type-B materialism (Balog 2020; Carruthers 2019; Papineau 1993, 2002, 2003, 2020) have noted that it raises the spectre of a much more pervasive and troubling form of indeterminacy. The threat here is that the reference of the concept Phenomenal Consciousness is indeterminate between properties that, while coextensive in the paradigm case of a human who can report their experiences, differ radically in their extensions outside of this paradigm case.

Why would that be? For most concepts, the story of how the reference of the concept is fixed will normally give at least some role to *conceptions*: the stock of beliefs a subject associates with the concept. We need not be internalists about mental content to allow *some* role for conceptions in fixing reference. For example, my correct belief that platypuses are egg-laying mammals is likely to form part of the explanation for why my concept Platypus successfully refers to platypuses.

Yet for the type-B materialist, many of the conceptions we ordinarily associate with phenomenal consciousness are misconceptions. The type-B materialist parts ways here with the analytic functionalist (e.g. Lewis 1983), who takes us to conceive of conscious experiences as states that play a certain type of functional role. The type-B materialist agrees with the dualist that either there are no a priori links at all between phenomenal consciousness and any functional concept or else those links are too minimal to render zombies inconceivable. The concept of phenomenal consciousness is not the concept of a functional role or its realizer. Instead, we tend to think that phenomenal consciousness as having a list of special features the set it apart from functional properties: it is irreducible, intrinsic, qualitative, primitive, ineffable, physically inexplicable, unknowable from the outside, its essence is fully revealed to us first-personally, and so on. But the type-B materialist parts ways with the dualist by holding that many of these intuitive judgements are false. Rejecting these conceptions allows the rejection of dualism, but it means the type-B materialist must hold that Phenomenal Consciousness refers to a physical property *in spite of* many of the conceptions we associate with the concept, not because of them.[3]

I say 'many' conceptions to allow that some conceptions may survive the type-B materialist bonfire. In particular, a type-B materialist can endorse the conception that phenomenal consciousness is a property that the referents of our phenomenal concepts have in common, and the conception that phenomenal consciousness has effects in the physical world (without being definable in

[3] Alternatively, the type-B materialist may argue that no conceptions at all are associated with Phenomenal Consciousness, because it is a bare, indexical concept, its content being roughly 'this sort of thing' (see Carruthers 2019). This is still a view on which conceptions play no role in fixing reference.

terms of these effects). They can also endorse various contingent generalizations about how phenomenal consciousness typically relates to other human capacities, such as the conception that our perceptual and affective states are often conscious, that we often remember our conscious experiences, and that we can conjure new conscious experiences in imagination. The problem is that the remaining conceptions are thin, and seem unlikely to be enough by themselves to differentiate phenomenal consciousness from other mental properties.

When explaining the reference of a concept like ARTHRITIS, we can (as Burge 1979 argued) appeal to deference to experts to show how successful reference can be compatible with both thin conceptions and serious misconceptions. But it would be implausible to appeal to deference to experts to explain the reference of Phenomenal Consciousness, since it is generally supposed to be a concept that we can grasp intuitively by reflecting on our own conscious experiences. There are no textbook definitions for us to consult; the textbooks refer us back to our own experiences.

With deference off the table, there is not much the type-B materialist can say except that our applications of Phenomenal Consciousness successfully *track* a physical property, despite our substantive misconceptions about the nature of that property. They track the property in the sense that all and only the states to which we are disposed to apply the concept first-personally in fact possess that property. The materialist can then argue that successful tracking is enough for successful reference, even in the presence of substantial misconceptions. But now indeterminacy looms, because our applications of Phenomenal Consciousness are likely to successfully track more than one physical property.

I will focus here on a version of the problem from Papineau (2002), which I take to be the most troubling version. For Papineau, Phenomenal Consciousness successfully tracks properties at least two different levels of organization: a high-level functional property and its neuronal realizer (Papineau 2002: 214). Our first-person applications of Phenomenal Consciousness to our own states (as when I judge, for example, that I am consciously experiencing a perception of a blue sky, but am not consciously experiencing the digestion of my breakfast) will track a *high-level functional property*. The science and philosophy of consciousness gives us several important candidates for this functional property. In broad terms, it may be entry to a global workspace (Dehaene 2014; Dehaene and Changeux 2011) or something causally upstream of entry to a global workspace, such as entry to fragile short-term memory (Block 2007, 2011) or something causally downstream of entry to a global workspace, such as becoming the object of a higher-order thought (LeDoux 2019; Rosenthal 2005). Let us assume, perhaps optimistically, that the science of consciousness will eventually reach consensus about what this high-level functional property is, and let us call it property *F*.

Crucially, this high-level functional property, whatever it is, will be coextensive in humans with a particular *neuronal mechanism* (or set of mechanisms) that realizes it. For example, entry to the global workspace may be coextensive with activation of a neuronal mechanism involving pyramidal neurons in the prefrontal cortex and the 'global ignition' of many cortical regions (Dehaene 2014; Dehaene and Changeux 2011; Mashour *et al.* 2020). Again, let us assume optimistically that the science of consciousness will converge on a single such neuronal mechanism, and call it property *N*.

I want to assume as little as possible about the nature of *N*. In particular, I leave open the possibility that *N* will be shared by all mammals and will not be specific to primates or to humans. What seems very unlikely at this stage, however, is that *N* will be shared by a wide range of non-mammalian animals. This is because we have clear evidence that conscious experience is intimately related to mechanisms in the neocortex, a brain region that has evolved since the divergence of the mammals from other lineages. I will not review this evidence here (but see Dehaene 2014; Koch *et al.* 2016; Aru *et al.* 2020; Frith 2021). If *F* has evolved in other, non-mammalian lineages, then it must have a *non-cortical neuronal implementation* that differs substantially from its cortical neuronal implementation in mammals.

Papineau's point is this: there will be no way to resolve the question of whether Phenomenal Consciousness refers to *F* or to *N*. Moreover, this is not, he suggests, merely the result of epistemic limitations. This is a point of contrast with Block (2002), who assumes that there must be some fact of the matter about whether *F* or *N* is phenomenal consciousness, but argues that we cannot know this fact. Papineau contends that the reference of the concept Phenomenal Consciousness is indeterminate between *F* and *N*. We are disposed to apply the concept, in our own case, to states that instantiate both properties. There is nothing in the concept, or in its associated conceptions, or in our use of it, that could fix just one of these properties as the unique referent. They are equally eligible candidates for reference. And yet the distribution of *F* and *N* in the natural world may well be very different: *N* is likely to be specific to mammals for the reasons noted above, whereas *F* may turn out to be possessed by a very wide range of animals (birds, reptiles, fish, cephalopods, arthropods) which have evolved a different neuronal implementation of the same functional property. This point is highlighted in the context of global workspace theory (the source of one important candidate for *F*) by Dehaene, who writes 'I would not be surprised if we discovered that all mammals, and probably many species of birds and fish, show evidence of a convergent evolution to the same sort of conscious workspace' (2014: 246). *F* but not *N* may also be possessed by non-living entities, such as future AI systems and robots, as emphasized by Dehaene *et al.* (2017).[4]

[4] Papineau's argument has received surprisingly little discussion. Taylor (2013) and Balog (2020) are exceptions. Taylor rebuts a distinct argument from Papineau (2002), the so-called

Carruthers, another prominent type-B materialist, arrives at a similar conclusion via a different route.[5] Carruthers (2019: 155–60) draws an analogy with a person who sometimes remarks, of a neighbourhood, that it is 'that sort of neighbourhood'. Suppose there is more than one property that these judgements track, and that these properties happen to be coextensive in the part of the world where the person lives. Perhaps they track both low socioeconomic deprivation and low levels of gun ownership. Now we ask: which neighbourhoods would be 'that sort of neighbourhood' in a different country where these properties are no longer coextensive? To settle such a question in practice, we could present the person with those neighbourhoods (or pictures of them) and see what they judge. Whether or not we actually do this, the person will still have dispositions to judge counterfactual neighbourhoods as 'that sort' or 'not that sort', and these dispositions may be enough to triangulate a single property as the referent of 'that sort of neighbourhood'.

For Carruthers, the concept Phenomenal Consciousness works like the phrase 'that sort of neighbourhood'. We acquire the concept by picking out various particular experiences first-personally, and then forming a concept of 'that sort of thing'. Our applications of the concept track different properties that are coextensive in our own case. We then ask: To which non-human mental states does this concept apply? As with 'that sort of neighbourhood', the way to settle such a question would be to present that subject with the non-human mental states in question, first-personally, and see what they judge. We can't do that, but one might hope that—as in the neighbourhood case—we could use a subject's dispositions-to-judge regarding non-human mental states to triangulate a single physical property as the referent. But these counterfactuals, Carruthers argues, are non-evaluable: there is simply no fact of the matter about whether I would, or would not, first-personally judge a particular non-human mental state to be phenomenally conscious, given the chance.[6] Given this, Carruthers argues, we should accept that there is no fact of the matter about whether a non-human mental state is phenomenally conscious or not. We have run out of reference-fixing resources.

One possible way to resist the threat of radical indeterminacy is suggested by Shea (2012). Shea notes that there is one more reference-fixing resource available to the type-B materialist: the role played by Phenomenal Consciousness

'methodological meltdown' argument, but does not rebut the argument for referential indeterminacy between *F* and *N*.

[5] There are differences between Papineau and Carruthers that I lack the space to discuss here. For Carruthers, the main threat is not one of indeterminacy between *F* and *N*, but indeterminacy between functional properties specified at different grains of analysis. For example, the coarse-grained functional property of possessing a global workspace *of some kind* (F_1) is coextensive in humans with possessing a global workspace *with the specific cognitive architecture of the human global workspace* (F_2). Papineau (1993: 124) discussed a similar issue very briefly in earlier work. Shevlin (2021) discusses a related idea under the heading of 'the specificity problem'. See also footnote 9.

[6] Papineau (1993: 126, footnote 23) makes a similar point, briefly.

in inductive inferences.[7] Return here to 'that sort of neighbourhood'. If we find that the person uses this phrase in inductive inferences (e.g. 'it's that sort of neighbourhood, so litter on the street will soon be picked up'), we can ask what property explains the success of these inferences. This property, argues Shea, is a more eligible candidate for reference than a coextensive property that does less work, or no work at all, in explaining the inductive utility of the concept. For example, if low socioeconomic deprivation explains why 'that sort of neighbourhood' is inductively fruitful, but low gun ownership does not, low socioeconomic deprivation is a more eligible referent—it attracts the reference of the concept more strongly.

Can we use inductive considerations to discriminate between *F* and *N*? It is plausible that we sometimes use Phenomenal Consciousness in successful inductions. For example, there are inductive links between conscious experience and memory: if I am having a phenomenally conscious experience of perceiving a stimulus *S* now, I am likely to retain an episodic memory of perceiving *S* later; but if I perceive *S* unconsciously, it is very unlikely that I will form an episodic memory of perceiving *S*. There are also inductive links between conscious experience and imagination. If I have had phenomenally conscious experiences in a perceptual modality *M* (e.g. colour vision), I am likely to be able to imagine having experiences in *M*; whereas if I have had never such experiences, I will struggle to imagine what they would be like. Perhaps *F* will play a much greater role than *N*, or vice versa, in explaining why these inductions work.

Yet I am doubtful that the relation of realization allows enough space between *F* and *N* for one to be substantially more relevant than the other to the explanation of our inductive successes. If we can explain the inductive links between conscious experience, memory and imagination in our own case by appealing to *F* and its integration with the cognitive architecture of memory and imagination, then we can also explain them by appealing to *N* and its integration with the human neural implementation of memory and imagination. The explanation can proceed equally well at either level, cognitive or neural, and the two explanations will complement each other. Whatever successful induction we choose, there will be a cognitive-level explanation for its success that appeals to *F* and its connections to other cognitive properties, and a neural-level explanation that appeals to *N* and its connections to the human neural implementations of other cognitive properties.

To find successful inductions for which *F* and *N* differ in their explanatory relevance, we would have to admit (as relevant for reference-fixing purposes) successful inductions concerning systems without *N*, such as inductions about

[7] Shea (2012: 335): 'irrespective of whether we conceive of [phenomenal consciousness] as being the occupant of a functional role, our concept refers to whatever property underpins the successful inductions in which it is deployed'. Shea credits this idea to Millikan (2000).

how conscious experience relates to memory and imagination in insects or robots. But we cannot take it for granted in this context that any such inductions are actually successful—to regard them as successful would beg the question by assuming that PHENOMENAL CONSCIOUSNESS refers to *F* rather than to *N*.[8] In sum, appealing to inductive utility seems to help with 'that sort of neighbourhood', where the candidates for reference are distinct in a way that gives them very different types of explanatory significance, but it seems not to help with cases where the two properties vying for reference are related by realization.[9]

Given the above, it is plausible that, if type-B materialism is true, then the reference of PHENOMENAL CONSCIOUSNESS is indeterminate between *F* and *N*. It is worth considering briefly why the same problem does not resurface for different solutions to the mind-body problem. It does not straightforwardly arise for forms of dualism (interactionist or epiphenomenalist) or Russellian monism (including forms of panpsychism and panprotopsychism) because these views accept our intuitive conceptions about phenomenal consciousness, including conceptions that concern its non-functional, intrinsic, distinctively first-personal nature. These views posit a special type of property that answers to those conceptions. The hard questions for these views are why we should believe that such properties exist, how we can come to know about their distribution in non-human systems, and how to reconcile their existence with a scientific worldview—not whether (if they do exist) the concept PHENOMENAL CONSCIOUSNESS succeeds in picking them out. On the face of it, that particular question can be answered relatively straightforwardly by such views.

One qualification to this, however: Russellian monism notoriously faces the additional question of how the psychic (or protopsychic) properties of elementary particles *combine* to yield the 'macroconscious' states of whole animals (Brüntrup and Jaskolla 2017), and I do not rule out the possibility that the problem of indeterminate reference could recur for some approaches to this problem. For example, imagine a version of Russellian monism that borrowed from type-B materialism the idea that 'macroconsciousness' is the property our first-person applications of phenomenal concepts reliably track. Such a

[8] This is related to a point made by Michel (2019). Michel argues: to test the claim that pain is multiply realizable, we need to settle the question of whether it is present in any animals with different neural states that play the same functional role as pain. But to settle this question, we first need to know whether pain is multiply realizable. The threat here is one of epistemic circularity. But there is also a threat of semantic circularity for a semantic theory such as Shea's that ties reference to inductive success.

[9] Shea's response may be more useful, I suggest, for defusing Carruthers' concern about indeterminacy between F_1 and F_2 (see footnote 5). We might well find that one of these cognitive properties is more relevant than the other to the cognitive-level explanation of our inductive successes. There will plausibly be a cognitive property that includes just enough architectural specificity to explain what needs explaining (e.g. the links between consciousness, memory and imagination) but no more specificity than is necessary.

view would face the same problem of indeterminacy between *F* and *N*, but this time in relation to the concept of 'macroconsciousness'.

What about an analytic functionalist, such as Lewis (1983)? The analytic functionalist might deny that PHENOMENAL CONSCIOUSNESS refers to anything, favouring instead more specific concepts such as PAIN, for which the corresponding a priori functional role is more easily specified (they would then be a kind of strong illusionist). Alternatively, they might try to construct an a priori functional specification of PHENOMENAL CONSCIOUSNESS. Let us set aside for the moment the problem posed to this idea by the conceivability of zombies. If they were to take this route, they could further argue that PHENOMENAL CONSCIOUSNESS is ambiguous between two functionally defined concepts, one of which determinately refers to a role-property (what Lewis might have called the 'attribute of having phenomenal consciousness') and the other of which determinately refers to *N*, the neural state that realizes the phenomenal consciousness role in humans (cf. Lewis 1983: note 6). The corresponding problem for ethics would be one of choosing how to resolve this ambiguity in ethical contexts. That would be an interesting problem in its own right, but it is distinct from the problem that confronts the type-B materialist, and I will not discuss it further here.

III. MANAGING RADICAL MORAL INDETERMINACY

By combining the considerations from the last two sections, and assuming type-B materialism is true, we can run the following argument:

> *Premise 1:* If a being has interests that register in conscious experience (e.g. through experiences with positive or negative valence), then these interests deserve higher moral priority than those of a behaviourally similar but non-conscious being.
>
> *Premise 2:* For many non-mammalian animals (e.g. birds, fish, crabs, insects, octopuses), the question of whether they have interests that register in conscious experience hinges on whether their valenced states are phenomenally conscious.
>
> *Premise 3:* The reference of PHENOMENAL CONSCIOUSNESS is indeterminate between a neuronal realizer property (*N*) that is not shared by non-mammalian animals and a high-level functional property (*F*) that is shared by many non-mammalian animals.
>
> *Conclusion:* For many non-mammalian animals, it is indeterminate whether or not the animal's interests deserve higher moral priority than those of a behaviourally similar but non-conscious being.

I grant that one option for the type-B materialist is to reject the plausible link between phenomenal consciousness and ethics described in Premise 1. However, given the plausibility of this link, we should ask whether there are any other escape routes. Can we accept the conclusion, or would doing so

lead to a catastrophic meltdown of our ethical deliberations regarding the treatment of non-human animals?

It is helpful here to distinguish between objective obligations and what have been called 'subjective oughts' or 'decision oughts' (Williams 2017). If it is indeterminate whether or not an animal deserves moral priority over a non-conscious being, our objective moral obligations (e.g. when forced to choose between its interests and those of a non-conscious being) will be indeterminate in relation to that animal. But it seems we can still ask: in practical deliberation, ought I give its interests the special weight owed to the interests of a conscious being, or not? I must choose between these options; the structure of deliberation forces a choice upon me. So ought I treat the animal *as if* it deserved the moral status of a conscious being or *as if* it did not deserve that status? The ought in question is a decision ought.

If there is no fact of the matter in relation to decision oughts, then we do indeed face a meltdown of practical deliberation. But objective indeterminacy may not have to spill over into indeterminacy at the level of decision oughts. There are various principles we could endorse that would prevent indeterminacy from derailing practical deliberation, allowing us to move from objectively indeterminate moral status to determinate decision oughts. I will call these principles 'blocking principles'.

One possible blocking principle draws inspiration directly from Williams (2017). Williams proposes that 'a choice to X is decision permissible iff it is not determinately objectively impermissible to X' (2017: 670). If neither of two options is determinately objectively impermissible, then we may (subjectively, in our practical deliberations) treat both options as if they were permissible. In moving from indeterminacy to decision, we err on the side of permissiveness. A natural way to apply this idea to the present problem is the following:

> ***Blocking principle 1:*** If an animal is neither determinately conscious nor determinately non-conscious, then it is decision-permissible for an agent to treat it as if it were conscious and also decision-permissible to treat it as if it were non-conscious, as long as the agent's choices are diachronically consistent with the same agent's other choices.

The motivation for Blocking principle 1 is the same as the motivation for Williams' principle. Faced with indeterminate obligations, rational decision must avoid *neutral sanction*: sanction from the point of view of someone who takes no stand on indeterminate matters. Following Blocking principle 1 allows an agent to avoid neutral sanction.

Williams includes a diachronic consistency constraint: you ought to avoid not just neutral sanction for doing something objectively impermissible, but also neutral sanction for being objectively inconsistent. Accordingly, once you have made a judgement call about an indeterminate matter, you ought to decide consistently with that judgement call in the future, as long as your views about the other issues at stake do not change. So, if you initially choose

to associate *F* with phenomenal consciousness in practical contexts (e.g. when faced with the case of a bird), then all your subsequent decisions should be consistent with that (e.g. when faced with the case of a cognitively sophisticated invertebrate or robot).

Williams notes that such a principle is likely to generate 'queasiness' (2017: 671), since you may find yourself rationally compelled (by your initial arbitrary judgement call) to make long sequences of decisions that an alter-ego who made a different judgement call would find subjectively impermissible. In the present case, 'queasiness' seems too weak a word: the principle leads to profound unease. Blocking principle 1 might have been acceptable if the only indeterminacy we faced were the mild, sorites-style form, so that our initial arbitrary judgement call would only very occasionally constrain our future choices. But when the indeterminacy infects our dealings with a very wide range of non-mammalian beings (including birds, reptiles, fish, invertebrates and, potentially, future robots and AI), a huge amount seems to hang on the initial judgement call. That judgement call will constrain our future choices whenever there is conflict between the interests of a mammal and a non-mammalian candidate for consciousness.

The unease can be compounded by the following thought experiment: imagine an avian species one day evolves a human-level capacity for introspective and ethical thought. The avian creature constructs its own introspective concepts that are functionally analogous to our phenomenal concepts. Let us call these 'phenomenal* concepts'. Assuming type-B materialism, these concepts refer (indeterminately) to physical properties of its own brain. Suppose it constructs a concept of phenomenal consciousness* that refers to the sort of thing its phenomenal* concepts pick out. This concept comes to carry great ethical significance for the avian creature. Unfortunately, its reference is indeterminate between *F* and N^*, the neuronal realizer of *F* in the avian brain, for the same reasons our analogous concept is indeterminate between *F* and *N*. It endorses Blocking principle 1, permissibly chooses to treat all beings without N^* as if they were non-conscious, and regards mammals, including humans, as deserving no higher priority than behaviourally similar but determinately non-conscious beings.

Can this unease be avoided? A different approach begins with the way we would approach cases of *uncertain* moral status, assuming a sharp boundary between conscious and non-conscious life—and then aims to treat indeterminacy on the model of epistemic uncertainty (an approach of this general type, but in the case of sorites-style vagueness, is pursued by Dunaway 2016). When consciousness is uncertain but determinately present or absent, there is a strong case for applying a precautionary principle and erring on the side of treating the animal as if it were conscious in any case where we find widespread practices causing extreme negative valence (Birch 2017). The same general thought, carried over to the case of indeterminacy, leads to the suggestion

that, if it is indeterminate whether an animal is conscious or not, then it is decision-obligatory to treat it in all respects as if it were conscious in any context where it may be caused suffering.

But what happens when the interests of the determinately conscious conflict with the interests of indeterminate cases? Three main variants of 'treating indeterminate cases as if they were conscious' arise, representing different ways of handling such conflict. First:

> ***Blocking principle 2:*** If an animal is neither determinately conscious nor determinately non-conscious, then it is decision-obligatory to treat it in all respects as if it were conscious in any context where it may be caused suffering, drawing no distinction (*per se*) between determinate and indeterminate cases.

This version faces the criticism that it ignores the ethical relevance of indeterminacy. Faced with a choice between doing something determinately objectively obligatory and something indeterminately obligatory, it is plausible that our determinate moral obligations take priority (cf. Williams 2017: 655). For example, we plausibly have a moral obligation to intervene when a conscious being is tortured for no reason in front of us. Grant this, and suppose we are forced to choose between saving a determinately conscious being from torture and saving an indeterminately conscious being. On Blocking principle 2, we ought not take the determinacy into consideration, violating the principle that determinate obligations take priority.

The intuitive pull of granting some ethical relevance to determinacy can be captured by either of the following:

> ***Blocking principle 3:*** If an animal is neither determinately conscious nor determinately non-conscious, then it is decision-obligatory to treat it in all respects as if it were conscious in any context where it may be caused suffering, subject to the qualification that *lexical priority* should be given to the interests of determinately conscious animals.

> ***Blocking principle 4:*** If an animal is neither determinately conscious nor determinately non-conscious, then it is decision-obligatory to treat it as if it were conscious in any context where it may be caused suffering, while giving *greater weight* (a 'determinacy multiplier') but not lexical priority to the interests of determinately conscious animals.

Either principle might provide an adequate treatment of *mild* indeterminacy, where we face a small number of borderline cases representing transitional states between conscious and non-conscious animals. However, in the face of radical indeterminacy between *F* and *N*, both principles bring us back to the problem of genealogical unease raised by our example of the reflective avian creature. They involve giving either lexical priority or greater weight to beings that share our own neuronal mechanisms, simply because they happen to share our own neuronal mechanisms. The introspective avian would be entitled to do likewise, deprioritizing or giving reduced weight to the interests of mammals. That prospect should give us pause before devaluing non-mammals in our

own ethical deliberations. Of course, Blocking principle 4 faces an additional problem: that of finding a non-arbitrary determinacy multiplier.

These four blocking principles suggest that we are stuck between a whirlpool and a rock. The whirlpool is entirely denying the ethical significance of determinacy, and the rock is a form of taxonomic chauvinism that should leave us profoundly uneasy, since it would allow a moral agent with a different neuronal realization of *F* to reason its way to chauvinism about us. Type-B materialism appears to leave us with a choice between abandoning a strongly plausible link between phenomenal consciousness and moral status, succumbing to a meltdown of practical deliberation regarding animals, or endorsing a blocking principle with profoundly troubling consequences.

Could this be a reason to reject type-B materialism itself? The fact that a metaphysical position leads to a rather bleak predicament is not itself a reason to reject the view; that would be wishful thinking. What we can say is that it may be a reason to hope for the truth of some alternative view: a view on which consciousness slots into its ethical role more neatly. But whether any such view exists is a topic for further debate. Nothing I have said rules out the possibility that other views of the mind-body relationship are subject to even worse problems, or relatives of the same problem, and are consequently no more capable of vindicating the ethical significance of consciousness.[10]

FUNDING

This research is part of a project that has received funding from the European Research Council (ERC) under the European Union's (EU) Horizon 2020 research and innovation programme, Grant No. 851145.

REFERENCES

Antony, M. V. (2006) 'Vagueness and the Metaphysics of Consciousness', *Philosophical Studies*, 128: 515–38.

Aru, J. *et al.* (2019) 'Coupling the State and Contents of Consciousness', *Frontiers in Systems Neuroscience*, 13: 43.

—— (2020) 'Cellular Mechanisms of Conscious Processing' *Trends in Cognitive Sciences*, 24: 814–25.

Balog, K. (2012) 'In Defense of the Phenomenal Concept Strategy', *Philosophy and Phenomenological Research*, 84: 1–23.

—— (2020) 'Hard, Harder, Hardest', in A. Sullivan (ed.), *Sensations, Thoughts, and Language: Essays in Honor of Brian Loar*, 265–89. New York: Routledge.

Birch, J. (2017) 'Animal Sentience and the Precautionary Principle', *Animal Sentience*, 16. https://animalstudiesrepository.org/animsent/vol2/iss16/1/

[10] I thank Tim Bayne, Liam Kofi Bright, Campbell Brown, Susanne Burri, Giacomo Giannini, Anna Mahtani, Matthias Michel, David Papineau, Jonathan Parry, Lewis Ross, Nick Shea, Johanna Thoma, Robert Williams and two anonymous reviewers for their comments and advice.

—— (2020) 'The Search for Invertebrate Consciousness', *Noûs*, https://doi.org/10.1111/nous.12351
Block, N. (1995) 'On a Confusion about the Function of Consciousness', *Brain & Behavioral Sciences*, 18: 227–47.
—— (2002) 'The Harder Problem of Consciousness', *Journal of Philosophy*, 99: 391–425.
—— (2007) 'Consciousness, Accessibility, and the Mesh between Psychology and Neuroscience', *Brain & Behavioral Sciences*, 30: 481–99.
—— (2011) 'Perceptual Consciousness Overflows Cognitive Access', *Trends in Cognitive Sciences*, 12: 567–75.
Brüntrup, G. and Jaskolla, L. (2017) *Panpsychism: Contemporary Perspectives*. Oxford: OUP.
Burge, T. (1979) 'Individualism and the Mental', *Midwest Studies in Philosophy*, 4: 73–121.
Carruthers, P. (2000) *Phenomenal Consciousness: A Naturalistic Theory*. Cambridge: CUP.
—— (2005) *Consciousness: Essays from a Higher-Order Perspective*. Oxford: OUP.
—— (2018) 'Valence and Value', *Philosohoy & Phenomenological Research*, 97: 658–80.
—— (2019) *Human and Animal Minds: The Consciousness Questions Laid to Rest*. Oxford: OUP.
Chalmers, D. J. (2010) *The Character of Consciousness*. Oxford: OUP.
—— (2018) 'The Meta-Problem of Consciousness', *Journal of Consciousness Studies*, 25: 6–61.
Crook, R. J. (2021) Behavioral and Neurophysiological Evidence Suggests Affective Pain Experience in Octopus. *iScience*, 24: 102229.
Cutter, B. (2017) 'The Metaphysical Implications of the Moral Significance of Consciousness', *Philosophical Perspectives*, 31: 103–30.
Dawkins, M. S. (2021) *The Science of Animal Welfare: Understanding What Animals Want*. Oxford: OUP.
DeGrazia, D. (1996) *Taking Animals Seriously: Mental Life and Moral Status*. Cambridge: CUP.
Dehaene, S. (2014) *Consciousness and the Brain: Deciphering How the Brain Encodes Our Thoughts*. New York: Viking Press.
Dehaene, S. *et al.* (2017) 'What Is Consciousness, and Could Machines Have It?', *Science*, 358: 486–92.
Dehaene, S. and Changeux, J.-P. (2011) 'Experimental and Theoretical Approaches to Conscious Processing', *Neuron*, 70: 200–27.
Dunaway, B. (2017) 'Ethical Vagueness and Practical Reasoning', *Philosophical Quarterly*, 67: 38–60.
Frankish, K. (2016) 'Illusionism as a Theory of Consciousness', *Journal of Consciousness Studies*, 23: 11–39.
Frith, C. D. (2021), 'The Neural Basis of Consciousness', *Psychological Medicine*, 51: 550–62.
Ginsburg, S. and Jablonka, E. (2007) 'The Transition to Experiencing: I. Limited Learning and Limited Experiencing', *Biological Theory*, 2: 218–30.
Godfrey-Smith, P. (2020) *Metazoa: Animal Minds and the Birth of Consciousness*. New York: Farrar, Strauss and Giroux.
Kahane, G. and Savulescu, J. (2009) 'Brain Damage and the Moral Significance of Consciousness', *Journal of Medicine & Philosophy*, 34: 6–26.
Kammerer, F. (2020) 'The Normative Challenge for Illusionist Views of Consciousness', *Ergo*, 6: 891–924.
Koch, C. *et al.* (2016) 'Neural Correlates of Consciousness: Progress and Problems', *Nature Reviews Neuroscience*, 17: 307–21.
Korsgaard, C. M. (2018) *Fellow Creatures: Our Obligations to the Other Animals*. Oxford: OUP.
LeDoux, J. (2019) *The Deep History of Ourselves: The Four-Billion-Year Story of How We Got Conscious Brains*. New York: Viking Press.
Lee, A. Y. (2019a) 'Is Consciousness Intrinsically Valuable?' *Philosophical Studies*, 176: 655–71.
Lee, G. (2019b) 'Alien Subjectivity and the Importance of Consciousness', in A. Pautz and D. Stoljar (eds.), *Blockheads! Essays on Ned Block's Philosophy of Mind and Consciousness*. Cambridge, MA: MIT Press.
Lewis, D. K. (1983) 'An Argument for the Identity Theory', in *his Philosophical Papers, Volume 1*, 99–107. New York: OUP.
Loar, B. (1990) 'Phenomenal States', *Philosophical Perspectives*, 4: 81–108.
Mashour, G. A. *et al.* (2020) 'Conscious Processing and the Global Neuronal Workspace Hypothesis', *Neuron*, 105: 776–98.

Michel, M. (2019) 'Fish and Microchips: On Fish Pain and Multiple Realization', *Philosophical Studies*, 176: 2411–28.
Millikan, R. G. (2000) *On Clear and Confused Ideas*. Cambridge: CUP.
Papineau, D. (1993) *Philosophical Naturalism*. Oxford: Blackwell.
—— (2002) *Thinking about Consciousness*. Oxford: OUP.
—— (2003) Could There Be a Science of Consciousness? *Philosophical Issues*, 13: 205–20.
—— (2020) 'The Problem of Consciousness', in U. Kriegel (ed.), *The Oxford Handbook of the Philosophy of Consciousness*, 14–38. New York: OUP.
Paul, E. *et al.* (2020) 'Towards a Comparative Science of Emotion: Affect and Consciousness in Humans and Animals', *Neuroscience & Biobehavioral Reviews*, 18: 749–70.
Pessiglione, M. *et al.* (2007) 'How the Brain Translates Money into Force: A Neuroimaging Study of Subliminal Motivation', *Science*, 316: 904–6.
—— (2008) 'Subliminal Instrumental Conditioning Demonstrated in the Human Brain', *Neuron*, 59: 561–7.
Regan, T. (2004). *The Case for Animal Rights*, 2nd edn. Berkeley: University of California Press.
Rosenthal, D. M. (2005) *Consciousness and Mind*. Oxford: Clarendon Press.
Schwitzgebel, E. (2021) 'Borderline Consciousness, When It's Neither Determinately True nor Determinately False that Experience Is Present', <http://www.faculty.ucr.edu/~eschwitz/SchwitzPapers/BorderlineConsciousness-210817.pdf> accessed 13 December 2021.
Shea, N. (2012) 'Methodological Encounters with the Phenomenal Kind', *Philosophy and Phenomenological Research*, 84: 307–44.
Shepherd, J. (2018) *Consciousness and Moral Status*. London: Routledge.
Shevlin, H. (2021) 'Non-Human Consciousness and the Specificity Problem', *Mind & Language*, 36: 297–314.
Simon, J. A. (2017) 'Vagueness and Zombies: Why 'Phenomenally Conscious' Has No Borderline Cases', *Philosophical Studies*, 174: 2105–23.
Singer, P. (1995) *Animal Liberation*. 2nd revised edn. London: Pimlico.
—— (2011) *Practical Ethics*, 3rd edn. Cambridge: CUP.
Skora, L. I. *et al.* (2021) 'Evidence that Instrumental Conditioning Requires Conscious Awareness in Humans', *Cognition*, 208: 104546.
Sneddon, L. U. *et al.* (2014) 'Defining and Assessing Animal Pain', *Animal Behaviour*, 97: 201–12.
Stoljar, D. (2005) 'Physicalism and Phenomenal Concepts', *Mind and Language*, 20: 469–94.
Taylor, J. H. (2013) 'Is Consciousness Science Fundamentally Flawed?', *Journal of Consciousness Studies*, 20: 203–21.
Tomasik, B. (2014) 'Do Artificial Reinforcement-Learning Agents Matter Morally?', *arXiv*, <https://arxiv.org/abs/1410.8233> accessed 13 December 2021.
Tye, M. (2021) *Vagueness and the Evolution of Consciousness*. Oxford: OUP.
Wiblin, R., Koehler, A. and Harris, K. (2019) 'David Chalmers on the Nature and Ethics of Consciousness', The 80,000 Hours Podcast, 16 December 2019. <https://80000hours.org/podcast/episodes/david-chalmers-nature-ethics-consciousness/> accessed 13 December 2021.
Williams, J. R. G. (2017) 'Indeterminate Oughts', *Ethics*, 127: 645–73.

Centre for Philosophy of Natural and Social Science, The London School of Economics and Political Science, UK