Leonard Dung

Preserving the Normative Significance of Sentience

Abstract:
According to an orthodox view, the capacity for conscious experience (sentience) is relevant to the distribution of moral status and value. However, physicalism about consciousness might threaten the normative relevance of sentience. According to the indeterminacy argument, sentience is metaphysically indeterminate while indeterminacy of sentience is incompatible with its normative relevance. According to the introspective argument (by François Kammerer), the unreliability of our conscious introspection undercuts the justification for belief in the normative relevance of consciousness.
I defend the normative relevance of sentience against these objections. First, I demonstrate that physicalists only have to concede a limited amount of indeterminacy of sentience. This moderate indeterminacy is in harmony with the role of sentience in determining moral status. Second, I argue that physicalism gives us no reason to expect that introspection is unreliable with respect to the normative relevance of consciousness.

1. Introduction

According to the orthodox view, consciousness is normatively significant. In general terms, this means that what we ought to do depends on the distribution of consciousness in the world. The normative significance of consciousness encompasses specific mental representations, global states and creatures. That is, it seems that our obligations are sensitive to differences in whether specific mental episodes are consciously experienced or not, whether a human is in a globally conscious state rather than being, for instance, in a coma, and whether particular animal species have the capacity for conscious experience. To pick an especially salient example, we seem to have a pro tanto obligation to remove the pain and suffering of others. However, if a cognitive process is either individually not conscious or belongs to a being which is not capable of having conscious experiences (suppose it belongs to a non-sentient species or a human in a state like coma), then it is questionable whether it can constitute pain or suffering. If there is such a thing as non-conscious pain or suffering, then this unconscious state seems at least to be less normatively significant than its conscious counterpart.

However, in recent years, the orthodox view has attracted criticism. According to its opponents, consciousness is irrelevant to our ethical obligations (Carruthers 2020; Kammerer

Correspondence:
leonard.dung@fau.de
Centre for Philosophy and AI Research, University Erlangen-Nürnberg

2019; Kammerer 2022). Instead, we have to look to other kinds of properties to ascertain what we ought to do. These alternative properties may, for instance, be the presence of strong preferences or some form of psychological sophistication, as long as they are understood in a way which does not implicitly presuppose conscious experience. This paper aims to defend the orthodox view against the most important recent objections. I claim that – although properties other than conscious experience may be normatively relevant as well – consciousness ought to play an indispensable and central role in our ethical deliberation.

In section 2, I will explain the orthodox view further, distinguish distinct varieties of it and outline its appeal. In section 3, I will present considerations against this view. I will focus particularly on two objections which have been forcefully advocated by Kammerer (2022). According to Kammerer, illusionism or reductionism about consciousness create problems for the orthodox view. In particular, reductionism seems to entail that consciousness is ontologically indeterminate and both positions undercut introspective justifications for the orthodox view. Section 4 contains replies to both objections in turn. Section 5 concludes.

2. The normative relevance of sentience

2.1 Varieties of sentientism

This section will begin with laying some terminological groundwork. I define *sentience* as the capacity to have conscious experiences. The notion of consciousness at stake is 'phenomenal' consciousness. It concerns how mental states are subjectively experienced from a first-person perspective respectively "what it's like" (Nagel 1974) to be in a certain mental state. As a first approximation, *sentientism* is the view that consciousness – and derivatively sentience – is normatively relevant.[1] The literature contains different ways to spell out the normative relevance of consciousness.

For this reason, we can distinguish different versions of sentientism. First, authors differ in whether they treat sentience as the capacity for conscious experience in general (as I define it), or specifically for valenced conscious experience, i.e., experiences that feel good or bad, like pain, fear, joy or relief (Birch et al. 2021). Thus, we can introduce *narrow sentientism* as the view that only valenced conscious experiences are normatively relevant and *wide sentientism* as the claim that some non-valenced conscious experiences are normatively relevant as well. I remain neutral between those two views.[2]

[1] Note that sentientism does not imply that every conscious state is normatively relevant, but only the weaker claim that some conscious states are normatively relevant in virtue of their particular phenomenal character, i.e., how they feel. For a criticism of the stronger view, see Lee (2018).

[2] For a discussion of narrow and wide sentientism (as defined here), see Chalmers (2022).

Furthermore, we need to make the notion of *normative relevance* more precise. In the literature, two conceptions of what consciousness is relevant for are most prominent: moral status (Jaworska and Tannenbaum 2021; Shevlin 2020; Singer 2011) and intrinsic value (Kammerer 2019; Kriegel 2019). A creature possesses moral status iff it matters morally for its own sake. A creature with moral status can be wronged, i.e., actions can be wrong because they violate *its* interests as opposed to someone else's. Similarly, a situation contains intrinsic value iff it is good or bad for its own sake, i.e., independently of its (especially causal) relations to other situations. Hence, we can stipulate *status sentientism* to be the thesis that consciousness intrinsically contributes to moral status and *value sentientism* to be the view that consciousness intrinsically contributes to intrinsic value. Again, I do not choose sides between these positions. In what follows, I will frame objections to sentientism either as objections to value or to status sentientism, depending on which framing is more typical in the literature. However, all objections to be mentioned apply to both versions of sentientism.

The notion of *intrinsic contribution* captures that consciousness is relevant to moral status or intrinsic value in a non-instrumental way, i.e., not because it is related to some other third property. Different theorists can flesh this notion out by positing their favorite relations of ontological dependency, for example grounding or constitution, as explanations of the intrinsic contribution of consciousness to moral status or value.

As a final piece of terminological clarification, I add the distinction between *strong sentientism* and *weak sentientism*. Strong sentientism is the view that *only* consciousness is normatively relevant such that other properties do not possess normative relevance whereas weak sentientism does not commit to this further claim. As you may have come to expect by now, I do not make a claim regarding the truth of strong sentientism. The task of motivating weak sentientism is rather modest. All I need to do is to support the contention that there are some cases where consciousness is normatively relevant.

2.2 The appeal of sentientism

Why should one believe in the normative significance of consciousness? First of all, the view is widely shared by experts and the broader public alike. With few but important exceptions, almost all experts believe in sentientism. Indeed, most think that consciousness is very significant for ethics. Notably, the objections to sentientism we will mention later stem mostly from philosophers of mind. Ethicists almost unanimously accept weak sentientism (e.g., Jaworska and Tannenbaum 2021; Kagan 2019; Nussbaum 2007; Singer 2011). Laws regarding the treatment of patients with post-comatose disorders and of animals also reflect the assumed

relation between moral rights and sentience (Browning and Birch 2022; Browning and Veit 2022). Nevertheless, even if sentientism is widely accepted, one should ask whether there actually are reasons that warrant this (near) consensus.

There are three main avenues to justify sentientism: via introspection, via general theoretical considerations and via the method of cases. First, when we are in pain, this experience seems to immediately strike us as bad (Kammerer 2022; Muehlhauser 2017). Some think that one directly introspects the pain's badness. This is underscored by the observation that it is hard to conceive of the pain not as something bad.[3] As will become crucial later, Kammerer (2022) especially emphasizes this point:

> There appears to be no need for careful reflection on elaborated moral laws, ethical principles, or contractual agreements, to know that *this* (focusing on our current pain) is bad because of the way it *feels*, and hence that, when someone else feels the same thing, it is also bad.

So, arguably sentientism is an obvious datum of introspection.

Second, there are some general theoretical considerations which favor sentientism. Valenced conscious experience seems conceptually entangled with values and moral status. Violating someone's interests without a good reason seems to wrong that being. For having one's interests frustrated is bad for one. If this is true, having interests is sufficient for possessing a moral status. In addition, it seems plausible that some valenced conscious experiences suffice for the possession of interests. A being which consciously feels pain seems to have a vested interest in the disappearance of that pain experience. If this is true, then certain conscious experiences are sufficient for the possession of moral status.

In addition, the truth of many plausible ethical beliefs can be *explained* by sentientism. There are (at least) pro tanto obligations to refrain from hurting someone, to anaesthetize patients subjected to (otherwise painful) surgery and to treat people with respect. If sentientism is true, then this arguably explains why these obligations hold: Hurting humans, consciously experienced surgery or treating other humans without respect arguably tends to increase long-run suffering and thus – if value sentientism is true – makes the world worse.

Third, the method of cases is the typical way to determine whether particular states are intrinsically valuable. Using this method, we have to construct various fictional scenarios to consult our intuitions about them. More precisely, we should imaginatively contrast two situations which are almost identical, only that one contains conscious experiences or conscious

[3] I thank an anonymous Reviewer for pointing out that some forms of pain may not be introspected as something bad, e.g., pain for masochists or pain asymbolics (Grahek 2001). This observation may be a challenge for the view that *all* conscious experiences are normatively relevant. However, it does not undermine the view that *some* conscious experiences are normatively relevant which is all sentientism (as defined here) entails.

beings while the states and beings in the contrasting situation are fully unconscious (Kriegel 2019; Levy 2014). If these two situations differ in perceived intrinsic value, then the value difference has to obtain in virtue of the consciousness difference. In this case, we can conclude that consciousness intrinsically contributes to intrinsic value.

Employing this method clearly suggests that consciousness is normatively relevant. To mention some examples: While it is very bad when someone experiences excruciating conscious pain, unconscious pain processing causing reflexive pain behavior is not nearly as bad.[4] Torturing animals is at least partially bad because animals feel (conscious) pain in response. If torture would not cause pain experience, it would not nearly be as bad. Being subject to an invasive operation is, abstracting from its potentially damaging instrumental effects, less bad when one is anaesthetized such as not to feel pain. This is, after all, a paramount reason why we use anesthesia. Finally, it is intrinsically bad when locked-in patients feel unhappy, while no such concerns apply in the case of comatose patients.

To summarize, introspection, theoretical considerations and particular cases converge in their support for sentientism. For this reason, (at least) weak sentientism seems to be on pretty solid grounds. In the next section, we will encounter two important arguments against sentientism. Subsequently, we will evaluate whether they are strong enough to warrant a rejection of sentientism.

3. Objections to sentientism

3.1 The indeterminacy objection

In this section, I will introduce the two objections to sentientism which will occupy us for the remainder of the paper.[5] The *indeterminacy objection* (Kammerer 2022) to sentientism is based on the claim that, for many beings and mental states, there is no fact of the matter whether they are conscious. Focusing on the consciousness of beings, rather than states, and moral status, rather than value, it can be characterized schematically as follows:

P1: For a wide range of beings, it is metaphysically indeterminate whether they are sentient.

[4] Both kinds of pain may be similarly bad instrumentally, for instance if they cause a trauma, but in terms of their contribution to intrinsic value, there is a sharp contrast.

[5] Other objections to sentientism are based on the illusionist theory of consciousness (Kammerer 2019) or the epistemic inaccessibility of sentience (Danaher 2020; Gunkel 2019; Shevlin 2020). For a response to the former argument, see Dung (2022a), and to the latter, see Dung (2022b).

P2: If sentientism is true and – for a wide range of beings – it is metaphysically indeterminate whether they are sentient, then – for a wide range of beings – it is metaphysically indeterminate whether they have moral status.
P3: It is not the case that – for a wide range of beings – it is metaphysically indeterminate whether they have moral status.

C: Sentientism is not true.

Let's discuss the three premises in order. Arguments to the effect that sentience may be metaphysically indeterminate in some sense have been presented by a wide range of authors (Birch 2022; Carruthers 2020; Cutter 2017; Hall 2022; Papineau 2002; Tye 2021). That being said, these arguments are based on metaphysical and semantic presuppositions which differ from and sometimes conflict with each other. In addition, they are presented with varied aims in mind and reach different kinds of conclusions regarding the extent to which consciousness is indeterminate.[6] For this reason, I will first stick to Kammerer's presentation of the argument for P1 since he uses it in the context that interests us: a general attack on sentientism.

Kammerer starts with assuming that reductive materialism implies that the conception associated with our concept of consciousness is, to some extent, illusory.[7] For – to the extent that this conception is anchored in our introspection of conscious states – it plausibly depicts consciousness as intrinsic, immediately known and irreducible to physical or functional properties. All these putative properties of consciousness are denied by reductive physicalists. Given this, the conception that introspection attaches to our concept of consciousness – which characterizes consciousness as essentially non-physical – cannot pick out a (unique) referent. To fix the reference of our consciousness concept, we must instead posit that the concept refers to the physical-functional property (e.g., broadcasting in a global workspace, having certain kinds of higher-order representations etc.) which actually covaries most systematically with the introspective application of our concept. This physical property is what constitutes consciousness in humans.

[6] Importantly, many of the authors mentioned don't ultimately hold that consciousness is indeterminate. In particular, some use the argument that indeterminacy of consciousness follows from certain assumptions as *reductio* against one of these assumptions.
[7] More precisely, Kammerer starts out by talking about the illusory nature of our (presumably sub-personal) introspective representations of conscious states, not our concepts. However, since his argument ultimately relies on the truth-conditions of consciousness ascriptions and thus on features of our (personal-level) concept of consciousness, I prefer this way of setting up the argument.

However, following Carruthers (2020), Kammerer emphasizes that the physical property constituting consciousness in humans will never be *exactly* shared by animals and machines. Suppose consciousness is, in humans, constituted by global broadcasting of information (Baars 1988; Carruthers 2020; Mashour et al. 2020).[8] Various animals have cognitive mechanisms which resemble human global broadcasting to various degrees and in multi-facetted, cross-cutting ways. Our introspective representations cannot decide whether the broadcasting mechanisms in, say, rabbits are similar enough to human broadcasting to count as conscious. Since rabbits plausibly do not have introspective representations of their broadcast contents, it is not clear which fact could determine which cognitive processes in rabbits are sufficiently similar to human broadcasting to constitute consciousness and which are not. Crucially, this indeterminacy does not indicate limitations of our knowledge of animal consciousness. Instead, there is no fact of the matter, since nothing fixes whether the human concept of consciousness refers to rabbit broadcasting processes or not.

Similar to Birch (2020), Kammerer considers that global broadcasting might be a natural kind, i.e., a division in nature which is independent from our practices of conceptual classification. Such kinds are posited to be *referentially magnetic*, such that our consciousness concept might determinately refer to whatever property is a member of the general kind global broadcasting.[9] However, Kammerer retorts that there would be too many natural kinds which exert this referential magnetism on our consciousness concept to make its reference determinate. In particular, adapting an argument by Birch (2022) and Papineau (2002), he claims that the reference would be indeterminate between the functional kind global broadcasting and the specific neural process which implements – and is always co-instantiated with – this functional kind in humans. This would mean, for instance, that it is indeterminate whether artificial systems which instantiate global broadcasting but do not have a brain are conscious. The same might apply to bird brains which are functionally complex but do not contain a cortex (Birch 2022). Hence, there would be many types of animals and machines for which it would be metaphysically indeterminate whether they are conscious. The analysis of the remaining premises of the indeterminacy argument will be briefer.

P2 is true if 'sentientism' is understood as 'strong sentientism'. If sentience is the only property that is relevant for whether beings have moral status and the sentience of an animal is

[8] Since Carruthers (2020) endorses the global workspace theory of consciousness, this is the only example he focuses on.

[9] Crucially, this kind is assumed to have coarse-grained identity conditions, such that, e.g., rabbits and humans could instantiate the same kind, despite manifold differences in their detailed psychological and neural makeup.

indeterminate, then its moral status is indeterminate as well. Weak sentientists can allow that further properties contribute to the determination of moral status. They might say that – in beings whose sentience is indeterminate – other properties gain primacy in fixing moral status such that moral status comes out as determinate, nonetheless. However, if the metaphysical indeterminacy of sentience is as widespread as the arguments in favor of P1 attempt to show, namely if on most controversial discussions regarding sentience in animals or AI there is no fact of the matter, and if sentience is irrelevant to the moral status of these beings, then the normative contribution of sentience would be very limited. The resulting very weak form of sentientism is something that proponents of sentientism might want to avoid.

Finally, P3 states that moral status is determinate or that cases of indeterminacy are at least rare. The main argument for this claim rests on the link between moral status and ethical deliberation. We care about the distribution of moral status because, in many situations, our ethical obligations depend on which of the beings affected by the decision possess moral status. If moral status is indeterminate, then ethical facts turn out to be indeterminate as well. The orthodox response to this conundrum is to make a terminological *stipulation* which settles per convention when we treat beings as sentient and therefore as possessors of moral status. Yet, since there is no independent fact of the matter, this terminological decision is essentially arbitrary such that our ethical obligations become arbitrary as well. At least if one shares the inclinations of an ethical realist, this consequence is unacceptable.[10]

Instead of conferring determinate moral status per convention, Birch (2022) examines which subjective norms for decision-making might obtain in the explicit condition of indeterminacy of moral status. While I cannot recapitulate Birch's reasoning here, he discovers that all available options seem to be unsatisfactory. They all downplay the ethical significance of indeterminacy: Either the difference between having a determinate moral status and having an indeterminate moral status or the difference between determinately not having a moral status and having an indeterminate moral status is levelled in a problematic manner.

In conclusion, there seem to be strong arguments for all three premises of the indeterminacy objection. Even if P2 is not strictly speaking correct, since one may hold on to a very restricted version of sentientism, the indeterminacy objection nevertheless implies that the role of sentience in fixing moral status – and the normative significance of sentience more generally – is small. Next, we will encounter the second objection to sentientism.

[10] To be clear, as it is frequently defined, ethical realism may be logically consistent with the metaphysical indeterminacy of many, although not of all, ethical facts. That being said, the intuitions animating realist views are in tension with the admission that many ethical decisions can be settled purely conventionally.

3.2 The introspective objection

The introspective objection, developed by Kammerer (2022), is as follows:

P1: Reductive physicalism and illusionism both (individually) entail that beliefs about consciousness which are claimed to be justified by introspection are actually unjustified.

P2: Our belief that phenomenal consciousness grounds moral standing is claimed to be justified by introspection.

C: Our belief that phenomenal consciousness grounds intrinsic value is unjustified.

As can be seen from P1, the introspective argument shares the starting point of the indeterminacy objection. Reductive physicalism and illusionism both imply that our introspective grasp of conscious experience – which portrays it as fundamentally non-physical and irreducible – heavily misleads us. If our introspection of consciousness deceives us in this case, we arguably cannot trust it in further cases. I will not challenge this premise here.[11]

The reasoning in favor of P2 has already been presented last section. According to Kammerer, the plausibility of sentientism is based first and foremost on our immediate introspection of the goodness or badness of our conscious experiences. Kammerer grants that there may be non-introspective justifications of sentientism. However, even in this case, the overall justification of sentientism would be weakened extremely once introspective rationales are undermined. He claims (Kammerer 2022):

> Our impression that phenomenal consciousness has some significant intrinsic value at best becomes on a par with other impressions we might have about other valuable things, which means that we should treat it with great caution and consider it highly defeasible (think about Greeks and slavery, Kant and masturbation, etc.)

That is, given that we cannot rely on an introspective justification for sentientism, we should treat it as on a par with other questionable ethical views. There may be some highly defeasible reasons in favor of sentientism and it may ultimately turn out to be correct, but its epistemic status is not much different from culturally influenced prejudices.

This relates to a difference between the indeterminacy and the introspective objection: While the indeterminacy objection concludes that sentientism is false, the introspective

[11] However, the premise does rely on some controversial assumptions. In particular, certain solutions to the meta-problem of consciousness (Chalmers 2018), i.e., the problem of explaining why it seems like there is a special and unique problem of explaining consciousness in physical terms, might rule out this objection. For instance, if our intuitions that consciousness is irreducible are mostly caused by contingent cultural and socio-historical factors (Rosenthal 2019), then one may argue that there are no analogous undermining factors present in respect to our intuition that consciousness grounds intrinsic value. Thus, one may say that we can trust this latter intuition more.

objection merely undercuts the justification for sentientism. It concludes that sentientism is not (sufficiently) justified, not that it is false. Nevertheless, if Kammerer is correct and sentientism's epistemic standing is comparable to the justification beliefs like the permissibility of slavery enjoyed, sentientism would be in deep trouble.

In this section, we have encountered two objections to sentientism. The first is based on the putative indeterminacy of consciousness, the second on the illusory nature of our introspection of it. Prima facie, the indeterminacy and the introspective objection both have some appeal. That being said, in the next section, I aim to refute both objections.

4. Rebutting the objections

4.1 Varieties of physicalism and moderate indeterminacy

In this sub-section, I present my counterargument to the indeterminacy objection. My reply is based on two theses: First, the arguments supplied above demonstrate only a modest, not a widespread indeterminacy of sentience. Cases of indeterminate sentience are conceivable, but – if some are actually realized – they are few and far between. Second, even opponents of sentientism should be prepared to admit a modest amount of indeterminacy regarding moral status and value.

The indeterminacy objection is based on the fact that three claims are incompatible: sentience is metaphysically indeterminate, moral status is determinate and sentience determines moral status. I will discuss the first two claims to show that the extent to which sentience is indeterminate fits well with the extent of moral indeterminacy we should expect independently. Let's start with the indeterminacy of sentience. Kammerer conjoins arguments by Birch and Carruthers to make the case that the indeterminacy of sentience concerns a wide range of animals and machines. As seen last section, Kammerer's argument depends on the claim that a unique reference of the term consciousness cannot be established by treating consciousness, viz. the physical process underlying it, as a natural kind. Given the argument he ascribes to Birch, the reference of our consciousness concept would nevertheless be indeterminate between a broad functional kind – like global broadcasting – and the concrete neuroscientific kind which implements it. For consciousness correlates equally well with both putative natural kinds in humans. Since there are many animal species and possible future AI systems which instantiate the functional kind but not the specific neuroscientific kind, there would be widespread indeterminacy of sentience.

However, Kammerer's argument neglects that Birch's (2022) argument explicitly rests on a specific version of physicalism, namely "type-B" physicalism (Chalmers 2003).[12] Type-B physicalism is characterized by Chalmers as a view which acknowledges that there is an "epistemic gap" between consciousness and the physical world such that, e.g., zombies are conceivable, but denies an ontological gap. Thus, it is a moderate position which concedes that phenomenal consciousness cannot be *understood* in physical terms in the way ordinary properties can, while its ultimate nature is physical, nonetheless.

This is curious since the views of consciousness Kammerer is attracted to seem not to be moderate at all, but radically physicalist: Either they reject the existence of phenomenal consciousness outright (and thereby repudiate any epistemic gap in understanding the sort of consciousness we do have) or take a reductive stance.[13] That being said, Kammerer's (2022) view is compatible with type-B physicalism, since he understands 'reduction' in a metaphysical, not in an epistemic sense.[14] Yet, this does not change that one can disarm the indeterminacy objection by opting for a different kind of physicalism.

Let's look at why specifically type-B physicalism is necessary for Birch's argument. The reason is that Birch's argument relies on the type-B physicalist claim that either (i) the conception associated with our consciousness concept is largely false or (ii) there is no conception attached to our consciousness concept (Birch 2022; Papineau 2002; Tye 2008). Type-B physicalists endorse this claim because they posit an epistemic gap between our grasp of consciousness and our understanding of the physical world. Thus, they need to hold that we don't conceive of consciousness in physical terms. Yet, they are physicalists which commits them to the view that consciousness is physical. Thus, on type-B physicalism, our conception of consciousness is either largely false, uninformative or there is no descriptive conception attached to our consciousness concept.

With ordinary non-phenomenal concepts, the referent of a concept may partially be determined as being the kind of entity which satisfies the conception associated with the concept. But since type-B physicalists hold that either our consciousness concept does not come with a conception of consciousness or the part of the conception which is true is not very

[12] In Chalmers' seminal paper, this view is called 'type-B materialism'.

[13] Type-A physicalism is not inconsistent with Kammerer's (2022) view that "phenomenal consciousness (introspectively) seems to have properties that it does not have in reality". While type-A physicalists hold that our concept of consciousness is functional, they can nevertheless allow that introspection deceives us about the nature of consciousness (making it seem non-functional). This is because they hold that our concept of consciousness is formed independently of introspection.

[14] Metaphysically, property F reduces to the more fundamental property G iff there obtains a sufficiently close relation of ontological dependency between F and G, like identity or constitution. Epistemically, F reduces to G iff complete knowledge of all G-facts is (for an ideal reasoner) sufficient to come to know all F-facts.

informative, this conception cannot contribute substantially to fixing the reference of our consciousness concept. Then, Birch's argument succeeds. Given that the occurrence of this concept covaries with at least one neuroscientific and one functional kind[15] and introspection and deference to experts (Birch 2022) cannot determine the reference either, there is no fact which can establish a determinate reference for our consciousness concept.

The alternative to type-B physicalism – type-A physicalism – rejects an epistemic gap between consciousness and the physical world. According to type-A physicalists, we conceive of consciousness as a certain kind of physical or functional property (e.g., Dennett 1991; Lewis 1966). According to this view, there is no principled difference in the epistemic relation between the physical world and consciousness on the one hand and between the physical world and other high-level entities like organisms, thoughts or economies. For example, functional truths like the following may compose our conception of consciousness:

a) When a being reports the presence of an object, it is usually conscious of the object.

b) When a being's cognitive architecture is organized as a mere look-up table, it is not conscious (Negro 2020).

c) Having a brain as complex and interconnected as the human brain is normally sufficient for consciousness.

Many subtle ideas about the connection between consciousness and behavioral properties, cognitive architecture and neural mechanism may be part of our conception of consciousness. This conception can then help narrow down the candidate referents of our consciousness concept. For instance, if our conception of consciousness includes many claims consistent with functionalism, then a broad functional mechanism will turn out to be the referent of our concept. If, on the other hand, the conception links consciousness to specific neural properties, then a neural process may win out as the referent.

Crucially, even if a statement is part of our conception of consciousness, it does not need to be part of the proper concept. For instance, even if b (from above) belongs to our conception of consciousness, it is not necessarily the case that cognitive systems implementing a look-up table cannot be conscious. This is the case because the conception can serve as a mere reference-fixer which is contingently associated with the concept.

To see this, compare how the reference of natural kind terms like water is determined (Putnam 1975). A conception – something akin to "a transparent, odorless, liquid which freezes

[15] This way of putting the argument assumes that functional kinds exist (*pace* Buckner 2014). However, arguably, there are also multiple non-functional kinds which correlate with our concept of consciousness.

at 0 degree Celsius" – serves to pick out a class of potential referents but the concept water then refers to whatever shares the underlying chemical structure – H20 – of most objects in this class. Ultimately, the concept of water refers to some objects which do not satisfy the conception, as long as they belong to the natural kind which is formed by the chemical structure H2O. Similarly, the concept of consciousness may refer to a natural kind whose members do not all fully satisfy the associated conception.

While conceptions may play a role in fixing the reference of natural kind concepts, it is even more plausible that they are vital for determining the reference of other classes of concepts. Arguably, often concepts just refer to whatever satisfies the associated conception (Lycan 2019). Therefore, if our concept of consciousness does not differ much from more mundane concepts, conceptions contribute to reference determination. If conceptions contribute to reference determination, then it does not follow from the argument above that the reference of our consciousness concept is indeterminate. Hence, if sentientists are not type-B physicalists, they don't have to fear the indeterminacy objection.[16]

I have shown that type-A physicalism does not entail widespread indeterminacy of consciousness. Nevertheless, both views are compatible. Is there any reason to expect that, in practice, the view that our consciousness concept is associated with a reference-fixing conception will not be associated with widespread indeterminacy? Yes, on at least two views of how the reference of our consciousness concept is fixed. First, we said that the functional conception could serve as a contingent reference-fixer, as with natural kind terms (e.g. water). Since natural kinds "cut nature at the joints", their instances are mainly determinate (Birch 2022). Thus, reference will be determinate. Second, the correct conceptual analysis of our consciousness concept may determine its reference. This analysis will be based on how we actually use the concept *consciousness*. Since both the folk and the philosophers' conception seem to suggest that consciousness is not widely indeterminate, the default expectation is that consciousness will come out as determinate in this conceptual analysis.

Let us discuss to possible replies by proponents of the indeterminacy objection. First, they might insist that type-B physicalism is a more plausible view than type-A physicalism. If so, my reply to the indeterminacy objection fails. I cannot negotiate this complex metaphysical debate here, but let's view things from a detached perspective: The indeterminacy objection

[16] The easiest way to see that type-A physicalist views of consciousness (at least) do not entail indeterminacy of consciousness is that – according to type-A physicalism – our concept of consciousness may work the same way as other concepts of high-level properties, mental or otherwise. If one does not believe that life, photosynthesis, memory and income are radically metaphysically indeterminate, then one is not committed to radical metaphysical indeterminacy of sentience either.

shows that type-B physicalism, the normative relevance of sentience and the (predominant) determinacy of moral status are mutually inconsistent. Something has to give. As the objection presupposes, the view that moral status is mostly determinate is indispensable, unless we accept serious ethical problems. As explained earlier, the normative relevance of consciousness is normally even seen as self-evident, except by researchers engaging with the objections discussed in this paper, and a central ingredient in the accounts of moral status of virtually all ethicists. Thus, there are strong reasons to retain commitment to sentientism, if possible. By contrast, type-B physicalism is – despite its virtues – a very contentious view. Given that there are alternative physicalist views in good standing which avoid widespread indeterminacy of moral status and the rejection of sentientism, it seems currently advisable to reject type-B physicalism.

Second, Kammerer (2022) also alludes to a different line of support for premise P1 of the indeterminacy objection. Namely, if physicalism is true, it seems that consciousness is vague, i.e., admits of borderline cases.[17] In respect to every broadly physical or functional, high-level property it seems clear that there are borderline cases of the instantiation of this property (Cutter 2017; Tye 2021). For instance, the functional and biological features which characterize global broadcasting, e.g. information integration, number and complexity of consumer systems or number and length of long-range pyramidal neurons, all admit of variation on a continuous scale. Thus, some area on this scale covers borderline cases where a being is neither determinately conscious nor determinately unconscious.

However, there is no reason to think that these borderline cases of consciousness are extremely widespread. The relevant physical property might be delineated quite, although not perfectly, sharply. Furthermore, as Birch (2022) notes, the claim that few or no actual species are borderline cases of instantiating the relevant properties is "implicit in the idea that neurobiological and cognitive kinds are natural kinds that 'carve nature at the joints' with most cases between the joints". If a theory of consciousness would identify consciousness with a property which admits of exceedingly many borderline cases, we would need to refine the theory such as to refer to a kind which better captures the actual divisions within cognition.

In virtue of the familiar reasoning of the indeterminacy argument, limited indeterminacy of sentience plausibly translates into limited indeterminacy of moral status. Is limited indeterminacy of moral status a problem for sentientists? No, sentientists have two good replies available. First, in ethics, the existence of moral vagueness is widely (Constantinescu 2014;

[17] According to Antony (2006), every view according to which the physical correlate of consciousness does not belong to fundamental physics – including dualist theories – entails that consciousness is vague.

Dougherty 2017; Schoenfield 2016), although not universally (Dworkin 2011), presupposed for independent reasons. In general, there is no obvious compelling reason why an ethical realist should reject indeterminacy of ethical statements, if he accepts that there is empirical vagueness. At the same time, Sorites-like considerations – known from descriptive contexts – suggest the presence of borderline cases of the applicability of fundamental ethical predicates. For instance, suppose I ought to donate $10 to charity every month. Also, I ought to give $10.01, $10.02 etc. It is not the case that I ought to give a million dollar to charity every month. However, there is a nearly continuous transition from the one supposed obligation to the other. Plausibly, a difference in donation of $0,01 cannot change a determinate moral obligation to donate to the determinate absence of a moral obligation to donate. Thus, there will be an amount of dollars for which it is indeterminate whether one ought to donate that much (Constantinescu 2014).

Second, and more clearly, any competing account of moral status implies the presence of borderline cases. For there is no plausible account of moral status according to which moral status is entirely directly grounded in properties of fundamental physics.[18] Yet, given physicalism, any imaginable ground of moral status which is not a fundamental physical property admits of borderline cases. No matter whether one sees moral status as grounded in the possession of preferences (Shevlin 2020), rationality and the capacity for reflection or something else: Given physicalism, this ground is ultimately a broadly physical property (e.g., a neuroscientific kind), which is indeterminate in the same way the possible candidate processes underlying sentience are indeterminate. In conclusion, borderline cases of sentience do not threaten sentientism because (i) there are independent reasons to accept borderline cases of moral status and (ii) any putative ground of moral status consistent with physicalism admits of borderline cases of moral status.

Let's recap why the indeterminacy objection fails. The original objection is not sound because it presupposes type-B physicalism. This allows sentientists to adopt type-A physicalism and to subsequently claim that the conception associated with our concept of consciousness helps to make the latter's reference determinate. Thus, sentientists can hold that there is no widespread indeterminacy of sentience. While one should grant the existence of borderline cases of sentience, borderline cases of moral status are consistent with sentientism. Now that the indeterminacy objection has been defused, we will tackle the introspective objection.

[18] A sentientist view according to which consciousness is identical to a fundamental physical property might be the exception.

4.2 Reliable and unreliable introspection

According to the introspective objection, sentientism is not justified because belief in sentientism is mainly supported by introspection while illusionism and reductive physicalism demonstrate that we cannot trust our introspection of conscious states. Non-introspective sources of justification, it is claimed, do not suffice to grant sentientism a privileged epistemic status relative to other once prevalent prejudices, like the permissibility of slavery.

My counterargument targets P1, i.e., the premise that illusionism and reductive physicalism undermine the evidential value of introspection when it comes to beliefs about consciousness. The introspective objection is based on the assumption that introspection of conscious states is unreliable. However, things are not that clear-cut. According to the physicalist views at issue, introspection portrays consciousness as intrinsic, immediately known and irreducible to physical or functional properties: all properties which consciousness is denied having. However, physicalists do grant that humans know things like when they are in pain, what they are consciously seeing etc. If people have this knowledge, they must have it via introspection. So there seem to be "good" and "bad" types of cases: Types of cases where introspection systematically deceives us and types where it is very reliable. To find out whether introspection of value belongs to the good or the bad type, we have to ascertain what members of the bad type have in common.

It seems like all members of the bad type encapsulate a common core of dualist intuitions: in showing consciousness as intrinsic, immediately knowable and irreducible, introspection depicts consciousness as something which is not a typical functional or physical property. By contrast, introspectively acquired beliefs like "I am in pain" and "I am currently seeing a red tomato" are metaphysically neutral. They are not in tension with the belief that consciousness is a functional or physical property.

What about the introspectively acquired belief that consciousness is normatively relevant? Prima facie, this belief is compatible with holding that consciousness is reducible to physical properties. For there is no contradiction involved in thinking that the same property is purely physical and intrinsically valuable. However, even if the claim that consciousness is intrinsically valuable is compatible with different accounts of its metaphysical nature, a proponent of the introspective objection may hold that the normative relevance of consciousness is not justified without anti-physicalism. She may elaborate as follows: We only have the intuition that conscious pain is bad because our introspection conceives of it as a non-functional process. If we would regard consciousness as a purely functional process, e.g. a

process of making contents in the brain globally available (Dehaene 2014; Mashour et al. 2020), why would it be valuable? There is nothing about consciousness, conceived in third-person terms, which seems normatively relevant. Thus, introspection of consciousness as normatively relevant, and its justificatory role, does stem from introspecting it as irreducible.

I disagree. I submit that the justificatory role introspection has for the normative relevance of consciousness does not depend on the problematic kind of introspection which portrays consciousness as non-physical. Introspection of consciousness as valuable does not depend on the introspection of consciousness as non-physical. As Kammerer (2022) explains, when introspecting conscious pain, for instance, we seem to recognize that *this* (kind of state) is bad because it *feels* bad. Similarly, Muehlhauser (2017) argues that our intuitions about the value of experience stem from demonstrative judgements. According to him, when my foot hurts (say), my intuitive judgement about the pain is roughly of the form "Whatever *this* is, *this* is really bad" (Dung 2022a). If our introspection picks out the value of consciousness demonstratively, then it is neutral regarding its metaphysical status. In particular, it is neutral between whether consciousness is a physical property or not. Thus, introspection of consciousness as normatively relevant does not depend on anti-physicalist introspection. For this reason, physicalist views do not undermine the justification of the view that consciousness is normatively relevant.

5. Conclusion

There appear to be multiple strong reasons to think that sentience is normatively relevant, namely general theoretical considerations, introspection and the consideration of particular relevant cases. Neither arguments based on the putative metaphysical indeterminacy of sentience nor doubts about the reliability of our introspective access to consciousness severely undermine sentientism. The first argument presupposes type-B physicalism. If one opts for type-A physicalism instead, then one has to accommodate only a modest indeterminacy of consciousness, which does not threaten sentientism. The second argument exaggerates the degree to which introspection of consciousness is untrustworthy: No version of physicalism undermines the reliability of introspective judgements about the normative relevance of consciousness. In conclusion, sentientism should remain the default view on the grounds of value and moral status.

Acknowledgements

I would like to thank two anonymous reviewers as well as François Kammerer for helpful feedback and fruitful discussion on an earlier version of this manuscript.

Funding

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - project number GRK-2185/2 (DFG Research Training Group Situated Cognition).

References

Antony, M. V. (2006). Vagueness and the Metaphysics of Consciousness. *Philosophical Studies*, *128*(3), 515–538.

Baars, B. J. (1988). *A Cognitive Theory of Consciousness*. Cambridge University Press.

Birch, J. (2020). Global Workspace Theory and Animal Consciousness. *Philosophical Topics*, *48*(1), 21–38.

Birch, J. (2022). Materialism and the Moral Status of Animals. *The Philosophical Quarterly*. https://doi.org/10.1093/pq/pqab072

Birch, J., Burn, C., Schnell, A. K., Browning, H., & Crump, A. (2021). Review of the Evidence of Sentience in Cephalopod Molluscs and Decapod Crustaceans. *London School of Economics and Political Science*. https://www.lse.ac.uk/business/consulting/reports/review-of-the-evidence-of-sentiences-in-cephalopod-molluscs-and-decapod-crustaceans.aspx. Accessed 30 November 2021

Browning, H., & Birch, J. (2022). Animal Sentience. *Philosophy Compass*, *n/a*(n/a), e12822. https://doi.org/10.1111/phc3.12822

Browning, H., & Veit, W. (2022). The Sentience Shift in Animal Research. *The New Bioethics*, *0*(0), 1–16. https://doi.org/10.1080/20502877.2022.2077681

Buckner, C. (2014). Functional Kinds: a Skeptical Look. *Synthese*. https://doi.org/10.1007/s11229-014-0606-z

Carruthers, P. (2020). *Human and Animal Minds: The Consciousness Questions Laid to Rest*. Oxford, New York: Oxford University Press.

Chalmers, D. (2003). Consciousness and its Place in Nature. In S. P. Stich & T. A. Warfield (Eds.), *Blackwell Guide to the Philosophy of Mind* (pp. 102--142). Blackwell.

Chalmers, D. (2018). The Meta-Problem of Consciousness. *Journal of Consciousness Studies*, *25*(9–10), 6–61.

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

Constantinescu, C. (2014). Moral Vagueness: A Dilemma for Non-Naturalism. In *Oxford Studies in Metaethics* (Vol. 9). Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198709299.003.0007

Cutter, B. (2017). The Metaphysical Implications of the Moral Significance of Consciousness. *Philosophical Perspectives*, *31*(1), 103–130. https://doi.org/10.1111/phpe.12092

Danaher, J. (2020). Welcoming Robots into the Moral Circle: A Defence of Ethical Behaviourism. *Science and Engineering Ethics*, *26*(4), 2023–2049. https://doi.org/10.1007/s11948-019-00119-x

Dehaene, S. (2014). *Consciousness and the brain: Deciphering how the brain codes our thoughts* (p. 336). New York, NY, US: Viking.

Dennett, D. C. (1991). *Consciousness Explained*. Penguin Books.

Dougherty, T. (2017). *Vagueness and Indeterminacy in Ethics*. Routledge Handbooks Online. https://doi.org/10.4324/9781315213217.ch11

Dung, L. (2022a). Does Illusionism Imply Skepticism of Animal Consciousness? *Synthese*, *200*(3), 238. https://doi.org/10.1007/s11229-022-03710-1

Dung, L. (2022b). Why the Epistemic Objection Against Using Sentience as Criterion of Moral Status is Flawed. *Science and Engineering Ethics*, *28*(6), 51. https://doi.org/10.1007/s11948-022-00408-y

Dworkin, R. (2011). *Justice for Hedgehogs*. Cambridge, MA, USA: Harvard University Press.

Grahek, N. (2001). *Feeling Pain and Being in Pain* (2nd ed.). MIT Press.

Gunkel, D. J. (2019). No Brainer: Why Consciousness is Neither a Necessary nor Sufficient Condition for AI Ethics. In *AAAI Spring Symposium: Towards Conscious AI Systems*.

Hall, G. (2022). Is Consciousness Vague? *Australasian Journal of Philosophy*, *0*(0), 1–15. https://doi.org/10.1080/00048402.2022.2036207

Jaworska, A., & Tannenbaum, J. (2021). The Grounds of Moral Status. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2021.). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/spr2021/entries/grounds-moral-status/. Accessed 14 November 2021

Kagan, S. (2019). *How to Count Animals, more or less* (1st ed.). Oxford University Press. https://doi.org/10.1093/oso/9780198829676.001.0001

Kammerer, Francois. (2019). The Normative Challenge for Illusionist Views of Consciousness. *Ergo, an Open Access Journal of Philosophy*, *6*. https://doi.org/10.3998/ergo.12405314.0006.032

Kammerer, François. (2022). Ethics Without Sentience. Facing Up to the Probable Insignificance of Phenomenal Consciousness. *Journal of Consciousness Studies*. https://philarchive.org/rec/KAMEWS. Accessed 7 March 2022

Kriegel, U. (2019). The Value of Consciousness. *Analysis*, *79*(3), 503–520. https://doi.org/10.1093/analys/anz045

Lee, A. Y. (2018). Is Consciousness Intrinsically Valuable? *Philosophical Studies*, *175*(1), 1–17. https://doi.org/10.1007/s11098-018-1032-8

Levy, N. (2014). The Value of Consciousness. *Journal of Consciousness Studies*, *21*(1–2), 127–138.

Lewis, D. K. (1966). An Argument for the Identity Theory. *The Journal of Philosophy*, *63*(1), 17–25. https://doi.org/10.2307/2024524

Lycan, W. G. (2019). *Philosophy of Language: A Contemporary Introduction* (3rd ed.). New York: Routledge. https://www.routledge.com/Philosophy-of-Language-A-Contemporary-Introduction/Lycan/p/book/9781138504585. Accessed 14 March 2022

Mashour, G. A., Roelfsema, P., Changeux, J.-P., & Dehaene, S. (2020). Conscious Processing and the Global Neuronal Workspace Hypothesis. *Neuron*, *105*(5), 776–798. https://doi.org/10.1016/j.neuron.2020.01.026

Muehlhauser, L. (2017). Report on Consciousness and Moral Patienthood. *Open Philanthropy*. https://www.openphilanthropy.org/2017-report-consciousness-and-moral-patienthood. Accessed 7 March 2022

Nagel, T. (1974). What is It Like to Be a Bat? *Philosophical Review*, *83*(4), 435–50. https://doi.org/10.2307/2183914

Negro, N. (2020). Phenomenology-first versus Third-person Approaches in the Science of Consciousness: the Case of the Integrated Information Theory and the Unfolding Argument. *Phenomenology and the Cognitive Sciences*, *19*(5), 979–996. https://doi.org/10.1007/s11097-020-09681-3

Nussbaum, M., C. (2007). *Frontiers of Justice: Disability, Nationality, Species Membership*. Harvard University Press.

Papineau, D. (2002). *Thinking about Consciousness*. Oxford: Oxford University Press. https://doi.org/10.1093/0199243824.001.0001

Putnam, H. (1975). The Meaning of "Meaning." *Minnesota Studies in the Philosophy of Science*, *7*, 131–193.

Rosenthal, D. (2019). Chalmers' Meta-Problem. *Journal of Consciousness Studies*, *26*(9–10), 194–204.

Schoenfield, M. (2016). Moral Vagueness Is Ontic Vagueness. *Ethics*, *126*(2), 257–282. https://doi.org/10.1086/683541

Shevlin, H. (2020). Which Animals Matter?: Comparing Approaches to Psychological Moral Status in Nonhuman Systems. *Philosophical Topics*, *48*(1), 177–200. https://doi.org/10.5840/philtopics20204819

Singer, P. (2011). *Practical Ethics:* (3rd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511975950

Tye, M. (2008). *Consciousness Revisited: Materialism without Phenomenal Concepts*. Cambridge, MA, USA: MIT Press.

Tye, M. (2021). *Vagueness and the Evolution of Consciousness: Through the Looking Glass*. Oxford, New York: Oxford University Press.