Animal suffering and felt unpleasantness: a sensory-motivational account

Introduction

Does it hurt lobsters when they are boiled alive? Do laboratory rats experience discomfort if they are kept severely underweight for experimental reasons? Do farm animals who are kept in confined conditions thereby experience emotional trauma? Questions such as these are central to our ethical treatment of animals, yet there is little philosophical or scientific consensus about how to answer them. Different jurisdictions differ wildly on the forms of protection they offer to various animal species, and there are no widely agreed-upon criteria for determining if a creature is undergoing pain or other forms of suffering.

In moving towards a more systematic approach to animal welfare, one key philosophical requirement is that we provide a theory of *felt unpleasantness*. For while states such pain, hunger, and negative emotions differ in their typical causes and effects, they have an important common aspect, namely that they can, when consciously experienced, feel unpleasant. This is part of what makes these states of special ethical significance, since experiences of felt unpleasantness arguably have a constitutive negative connection to well-being: creatures are worse off for having them. To be sure, felt unpleasantness plays a critical role in learning and behavior. Pain can help us avoid further damaging an injured limb, and anxiety can prompt us to take action to address potential problems in our lives. However, considered apart from their broader regulatory functions, unpleasant states are a source of immediate disvalue to us: to be in pain is worse, *ceteris paribus*, than not being in pain, and experiencing anxiety is worse than being in a state of contentment.

In dealing with adult humans, we have comparatively little difficulty in determining whether someone is in a state of felt unpleasantness: we can simply ask them. However, no such resource may be available when dealing with animals, neonates, or (perhaps one day) artificial systems. The question of how we can more objectively determine when a state of felt unpleasantness is occurring, then, is of no small practical and ethical significance.

In this paper, I aim to take some initial steps towards an account of felt unpleasantness with practical applicability. I begin in Section 1 by explicating the notion of felt unpleasantness and arguing for it ethical significance. In Section 2, I argue that the goal of giving a theory of felt unpleasantness is well-founded insofar as we have some initial reason to regard experiences of felt unpleasantness as constituting a distinctive *psychological kind*, marked by its intrinsic negative motivational force and commensurability among its

instances. In Section 3, I outline and defend a new account of felt unpleasantness that I term the *Sensory-Motivational* account. In Section 4, I explore how this account could be applied to the practical assessment of animal welfare. In Section 5, I consider some objections to the Sensory-Motivational account concerning its claimed connection between felt unpleasantness and motivation. Finally, in Section 6, I consider whether philosophical worries about animal consciousness create an insurmountable stumbling block to the practical assessment of felt unpleasantness.

1. What is felt unpleasantness and why does it matter?

Assessing the well-being of both humans and animals is a fraught and normatively complex matter. Though some bold attempts have been made to naturalize and measure concepts like human thriving (Kringelbach and Berridge 2017), considerable disagreement remains even among psychologists as to which measures of happiness best track well-being (Greene, Morrison, and Seligman 2016)**.** Moreover, many philosophers have suggested that well-being may not be a purely psychological matter at all, but instead depend on non-experiential factors. Thus we might speak of someone as being in some sense harmed, for example, if their autonomy is compromised, even by events outside their ken (Nozick 1974). Likewise, it might be claimed that our well-being can be influenced by whether our hearts' desires are truly satisfied, independent of whether this resulted in experiences perceived to be unpleasant (Heathwood 2006).

Setting aside such complex questions, however, we can nonetheless recognize that a central and important form of disvalue comes in the form of experiences that simply *feel unpleasant*, such as pains, bodily discomforts, and negative emotions. While, as noted above, these states often play an important role in our behavioral economy, they often serve to inflict tremendous suffering without contributing much of value: cluster headaches, chemotherapy-induced nausea, and panic attacks are all extremely detrimental to their sufferers' well-being, even considered apart from the broader negative influence they exert in people's ability to fulfill their goals and lead rewarding lives.

This capacity for states of felt unpleasantness to cause suffering marks it out as an important target for philosophical and scientific understanding. However, it is has still further importance in light of the fact that for non-human animals, felt unpleasantness is likely to be the primary if not the sole source of harm they undergo. While humans may care about things like autonomy or the completion of long-standing personal projects, the main harms

experienced by a dog or a fish are likely to arise from states such as hunger, pain, and stress. Thus insofar as we are concerned with minimizing animal suffering, we would benefit tremendously from having a clearer grasp of the phenomenon of felt unpleasantness, and in particular, from developing behavioral or neuroscientific measures that might allow us to identify and quantify its occurrence.

The project of understanding and assessing felt unpleasantness in objective terms is of course highly challenging, and requires many distinct contributions from philosophers and scientists. Nonetheless, in what follows, I will attempt to provide some early steps towards such a framework. Despite the weak epistemic status of these efforts, however, I would suggest they may have considerable practical value. As matters stand, our assessment of felt unpleasantness, especially in the case of animals, is guided by a mix of intuition, legal precedent, and political compromise. Different jurisdictions afford varying protections to different species and indeed phyla of animals. Thus while the British Animals (Scientific Procedures) Act of 1986 extends protections to all vertebrates as well as octopuses, the corresponding American legislation (7 U.S.C. § 2131-2156) makes provision only for warm-blooded animals. While these regulations are certainly guided by ethical and scientific considerations, they are not made on the basis of any kind of settled framework. In light of this, I would suggest that even faltering progress towards a more systematic approach may be of considerable benefit.

Note that I will not attempt to offer any more precise characterization of felt unpleasantness other than to say that, in my intended senses of the terms, it should be eminently familiar to anyone who has undergone pains and bodily discomforts or who has endured mental states such as dread, despair, or grief. However, two brief points of clarification are in order.

First, I recognize that some will be disinclined to count states that feel only *mildly* unpleasant as having any ethical status. Certainly, it would seem melodramatic to refer to the mild feeling of pain in a stubbed toe, or one's hunger for lunch, or mild anxiety that one might miss the bus as instances of *suffering* (Dennett 1996: 166). I am inclined nonetheless to see such cases as admittedly borderline instances of a broader mental phenomenon that is of intrinsic ethical significance, and includes, in its extremes, desperate pain, wracking nausea, and profound anxiety.

In part, I am motivated in this contention by the simple observation that unpleasant sensations such as headaches can progress continuously from being barely noticeable to wholly unbearable, all the while being broadly disagreeable but without a subjectively

apparent moment in which they begin to harm us. As such, it seems reasonable to me to regard them as more or less intense instances of a single phenomenon involving intrinsic detriment to well-being. To further make the case that even mild cases of discomfort possess some ethical significance, consider Heathwood's (2007) example of Tina the Torturer, who subjects her victims to wearing itchy sweaters. While she might not count as a moral monster, she is surely doing something harmful. I am thus inclined to regard felt unpleasantness *tout court* as tracking a morally significant mental state.

A second point key point to note is that I take it that experiences of felt unpleasantness are bad for their subjects at least partially in virtue of being conscious. This may seem a trivial point, insofar as it is hard to make sense of the idea that someone could be in a state of felt unpleasantness without there being *something it was like* for them to undergo such a state. However, it might be claimed that what makes experiences of felt unpleasantness bad is their content, or their contribution to the organism's function, such that the same state would be just as bad for the creature if it occurred unconsciously (Carruthers 2005; Pereplyotchik 2017).

I certainly do not deny that pains, nausea, and so on can be bad for us in a host of different ways; by interfering with our ability to achieve our goals, for example, or preventing us from partaking in pleasures. However, it also strikes me that a fundamental way such states are detrimental to our well-being is by inducing in us a kind of ghastly *feeling* that does not admit of or require further justification. While it is hard to give an argument to this effect that does not beg theoretical questions of my opponents, I would note that, quite independent of any downstream effects on behavior, most of us would prefer that a painful operation be conducted under anesthetic, if only for the reason that we be spared unpleasant experience.[1]

With these considerations in mind, then, I would suggest that there is considerable ethical value in attempting to offer a scientifically applicable account of felt unpleasantness, where this is understood as a species of conscious experience that is intrinsically bad for the subject who undergoes it.[2] As I will now describe, there is some reason to be optimistic about the prospects of such a project.

---

[1] Note that I will return to some of these questions in Section 6, below.

[2] For the restricted purposes of this paper, I leave open the possibility that there are other forms of experience, besides felt unpleasantness, that are immediately bad for the subject who undergoes them. However, I see no clear-cut examples of such states. Thus while certain beliefs (such as the conviction that one is worthless, or that one's hopes have come to naught) might be associated with negative outcomes and unpleasant experience, I am unconvinced that simply having these beliefs is *intrinsically* bad for a subject.

2. Can we give a scientific theory of felt unpleasantness?

I hope that the foregoing discussion has motivated the idea that a providing scientifically-applicable account of felt unpleasantness is a worthy endeavor. However, one might question whether felt unpleasantness is properly considered a single cohesive phenomenon in its own right. My approach in this paper assumes, perhaps somewhat ambitiously, that felt unpleasantness can properly understood as a psychological kind. Broadly speaking, I take this to amount to the claim that felt unpleasantness has a distinctive psychological role: experiences of felt unpleasantness contribute to mental function in a way that marks them as a unified class of phenomena and distinguishes them from other mental states. This in turn makes them a proper object of study for cognitive psychology, and in turn makes reasonable the aspiration that we might be able to reliably identify their occurrence using objective measures from behavior and neuroscience.

However, I recognize this claim might be contested; perhaps, for example, felt unpleasantness is best considered a generic con-attitude that we sometimes adopt towards a certain class of experiences. If that is right, then the attempt to give a rigorous psychologically-grounded account of the phenomenon – as opposed to one embedded primarily in social and linguistic practices – might flounder from its outset, in much the same way as an attempt to give a properly psychological account of eloquence or charm might be ill-advised. The concern of the present section will be to respond to this objection.

First, note that it is undeniable that the notion of felt unpleasantness plays an important and distinctive role in folk moral psychology. We frequently appeal to felt unpleasantness in explanations of our own actions and the actions of others and to ground claims of deserved sympathy. Thus someone might claim, with immediately intelligibility, that she has not been herself because she has been suffering from chronic pains; likewise, we might cite the apparent distress of a child as a reason to intervene to help it.

Nonetheless, it does not follow from the fact that we employ a term in everyday psychological explanation that it is a properly homogeneous kind that can be incorporated into scientific psychology. The Ancient Greeks used the term *thumos*, for example, to refer to the set of emotions and attitudes associated with an individual's pride and standing in a community, but no directly comparable notion is used by psychologists today. Indeed, contemporary philosophers have challenged whether even central folk psychological notions like "emotion" properly constitute a single kind (Griffiths 1997).

Why might one doubt whether felt unpleasantness is a properly homogenous kind?

One simple reason concerns their disparate phenomenology. While we may speak of states like pains, nausea, and anxiety as feeling unpleasant, these states differ dramatically in how they feel. Of course, we should expect as much, insofar as they possess different somatosensory and affective components: nausea is felt in one's innards, and pain in the affected body part. Once we have accounted for such differences, however, it is far from clear that there is any further phenomenal property that they have in common.[3] It is with this worry in mind that Korsgaard asks "What do nausea, migraine, menstrual cramps, pinpricks, and pinches have in common that makes us call them all pains?" (Korsgaard 1996: 104). Likewise, Prinz claims, "[t]here is no feeling of unpleasantness; there are just unpleasant feelings" (Prinz 2006: 178).

While, for my part, I am minded to think that there is some phenomenal core common to experiences of felt unpleasantness, the disparate intuitions on this question make clear that phenomenology alone cannot ground the claim that experiences of felt unpleasantness share any kind of unified psychological basis. And while there is some neurobiological evidence supporting a common role for certain brain areas in feelings of unpleasantness, notably the anterior cingulate cortex (Foltz and White 1962; Aydede 2000), this evidence is hardly dispositive given the wide variety of functions performed by the anterior cingulate cortex, and the still unsettled nature of the science.

Instead, I would suggest that the best evidence for the claim that felt unpleasantness constitutes a proper psychological kind comes from two related observations. The first is that experiences of felt unpleasantness are *subjectively commensurable*. We are typically able (to some extent) to quantify and compare unpleasant states in respect of how bad they feel (Schroeder 2004: 87) For example, if we are given the choice between drinking a foul-smelling medicine and undergoing a moderately painful injection, we can reflect on which would be the more unpleasant. In doing so, we might imagine what the two experiences would be like and gauge which would feel worse. We can do this even though the two states feel very different and occur in different sensory modalities.

Such comparison is not always easy, of course; someone contemplating a knee replacement operation might struggle to decide whether the prospect of reduced chronic pain warrants the brief but more intense pain and discomfort involved in surgery, for example. But when we are forced to choose between two unpleasant experiences, we normally take

[3] This is sometimes referred to as the heterogeneity problem, and applies to pleasure as much as pain (if not more so). See Heathwood (2007) for discussion.

ourselves to have a *reason* for the choice we make, where this reason makes reference to how we expect the experiences to feel.

There is a contrast to be drawn here between this kind of commensurability and the methods we use to make broader decisions about our well-being. If considering, for example, whether to take on a work assignment that would remove me from my family for six months, I will weigh up numerous considerations, including many that are not straightforwardly experiential (for example, the loss that would be involved in missing seeing my children grow up for a period). Conversely, in the case of weighing up the relative unpleasantness of two pains, or a pain and some other unpleasant state, we can arguably rely solely on imaginatively projecting ourselves into the relevant situations and gauging which would feel worse.

The direct in-principle commensurability of states of felt unpleasantness provides some initial reason, I would suggest, for thinking that they may constitute a psychological kind. A second reason for thinking that felt unpleasantness is a reasonably unified phenomenon concerns its connection to motivation. While experiences of felt unpleasantness differ in many of their typical causes and effects, there is at least one effect they all arguably have in common: their nature is such that they serve to motivate us to put an end to the relevant states, or, as I will put it hereafter, they are intrinsically *negatively motivating*. Thus after we discover that a food makes us feel nauseous, or that a given medication induces vertigo, we are typically inclined to avoid it in future just in virtue of that feeling. Similarly, an experience's being unpleasant seems to provide a reason *in its own right* for not wishing to undergo it: if someone is asked to ride a rollercoaster and declines on the grounds that it makes them feel unpleasantly nauseous or dizzy, we can immediately make sense of their preference just on these very grounds.

In the same vein, it is hard to make sense of the idea that an experience could be truly unpleasant without making its subject, other things being equal, at least somewhat disinclined to undergo it. Thus imagine that we found someone self-administering electric shocks, who claimed to be doing so for no other reason than that it felt unpleasant. In such a case, we would doubtless say that the individual concerned was either lying or was subject to some further pathological delusion or desire.

Of course, that is not to deny that we are regularly motivated to undergo unpleasant states: we eat foul-tasting food so as not to offend a host, swallow bitter medicines to cure a headache, and engage in physical exercises that are sometimes painful in the name of health. In such cases, however, there is some further *overriding* motivation that makes the relevant

unpleasant state worthwhile. Even in more difficult cases in which such subjects voluntarily undergo unpleasant states, such as eating disorders or self-harm, many psychological explanations of the relevant pathological behavior will invoke some further goal or some (perhaps compensatory) positive feeling accompanying any felt unpleasantness, such a desire for self-punishment or a feeling of control.[4]

Tying these two considerations together, then, I would suggest that experiences of felt unpleasantness are marked out by their subjective commensurability *inter se* and their intrinsically *negatively motivating* character. If we grant that these features are individually necessary for a state to be one of felt unpleasantness, we might further ask whether they are also jointly sufficient. While I do not take myself to be in the business of giving a conceptual analysis of felt unpleasantness, I can think of no mental states that obviously fit these criteria that are not instances of felt unpleasantness. Someone might have an unconsciously conditioned aversion to spiders, for example, that causes them to leave the room whenever there is a suggestion that a spider is present, even without having any strong feelings. However, insofar as this case *ex hypothesi* relies on a conditioned behavioral response unmediated by feelings of fear or disgust, there would be no possibility of subjectively comparing the intensity of the relevant feelings to other negative states.

In any case, if I am correct in claiming that instances of felt unpleasantness can reliably be identified by these features of intrinsic negative motivational force and commensurability, then I would suggest we have promising initial grounds for treating felt unpleasantness as a candidate psychological kind.

3. A sensory-motivational account of felt unpleasantness

Thus far, I have made two main claims. The first is that felt unpleasantness is an important ethical concept with particular relevance to our treatment of animals. The second is that there are some initial promising reasons to think felt unpleasantness may be understood as a psychological kind, specifically one marked out by its qualities of commensurability and intrinsic negative motivational power. I now turn to the task that will occupy me for the remainder of the paper, namely offering a more developed account of felt unpleasantness that can be applied to the practical assessment of animal welfare.

[4] I recognise these cases are psychologically complex and ethically fraught, and detailed analysis of them is beyond the scope of this paper. The remarks provided are intended just to show that they do not offer uncontroversial counterexamples to the claim that felt unpleasantness is negatively motivating.

3.1 – What kinds of state can be felt as unpleasant?

The first task for such a theory is that we establish the *kinds* of states that can give rise to experiences of felt unpleasantness. While this may seem an esoteric philosophical question, we can put it more plainly if we simply ask: what kinds of states can feel unpleasant?

I suggest that the best answer to this question is *sensory* states and *affective* states. When I feel a pain in my foot, my experience is not simply one of generic unpleasantness, but rather of a particular *kind* of unpleasantness - such as an aching or burning sensation – located in a particular part of my body. Note that the character of these states is not exhausted purely by their (characteristic) felt unpleasantness: the sensory element in such states conveys certain information to me about where the unpleasant sensation arises, the kind of motor action liable to make it worse, its intensity relative to other aches or pains, and so on. Likewise, if I feel a sudden rush of anxiety, my experience presents itself not just as a state of general emotional unpleasantness, but as a more specific kind of psychological episode involving physiological arousal and, perhaps, a sense of some imminent threat or the need for urgent action.

In asserting that only sensory and affective can feel unpleasant, I do not deny that there may be psychological states that are broadly speaking *detrimental to our well-being* that lack a sensory-affective basis: one might perhaps feel generically out of sorts or in a bad mood, for example, with negative consequences for one's well-being. However, it would somewhat misleading to suggest that prolonged periods of low mood involving a persistent accompanying feeling of unpleasantness in the same way as pain or nausea. To the extent that they indeed involve episodes of felt unpleasantness, there will be corresponding sensory or affective specific states – such as more troublesome pains, or frequent gripes of despair – in which the felt unpleasantness would inhere.

I would suggest, then, that experiences of felt unpleasantness always involve a sensory or affective state.[5] That such a component is required is further suggested by the fact

[5] A central question in the philosophy of emotion concerns whether affective phenomenology can be understood in terms of somatosensory phenomenology (James 1884). If so, then one might claim that felt unpleasantness is strictly a sensory phenomenon. I am happy to remain neutral on this issue, however, insisting only that felt unpleasantness must involve some sensory *or* affective component, where the latter can be spelled out in various different ways.

that there are, to my knowledge, no clear cases in which an individual undergoes felt unpleasantness without *any* accompanying sensory-affective state.[6]

If this claim is correct, then we have taken our first step towards a scientifically applicable account of felt unpleasantness. Specifically, in assessing whether a given state feels unpleasant to a given subject, we can confine ourselves to sensory and affective states. This is still, to be sure, casting a very broad net: at any moment, most creatures will be undergoing numerous sensory and perhaps some affective states that might qualify. However, it marks a first step in the right direction.

3.2 – What makes a given state feel unpleasant?

The next task before us is to ask what it is that *makes* a sensory or affective state feel unpleasant in a given case. What is it about *this pain* or *this* feeling of anxiety that makes it feel unpleasant? It may not be immediately obvious why such an account is required. Is it not simply obvious that certain states feel bad and others do not? To attempt to understand felt unpleasantness simply by formulating a list of the various states that typically feel unpleasant would be insufficient for present purposes, however. For one, it would be highly uninformative. As noted earlier, pains, nausea, anxiety, and the like are quite different experiences. While I argued that all states of felt unpleasantness have features in common – namely their commensurability and their intrinsic negative motivational force – nothing I have said thus far illuminates *why* these features should be present in states of felt unpleasantness.

Additionally, we should acknowledge that states like pain, nausea, and even anxiety do not always feel bad. The clearest cases of this come from the medical literature, and specifically the phenomenon of pain asymbolia (see Grahek 2011 for a review). Patients with this condition have normal sensitivity towards pain and spontaneously report feelings of pain, but seem indifferent towards them. This, I would suggest, is naturally interpreted as their pains simply not feeling unpleasant to them.[7]

Even in daily life, moreover, it seems that states that are typically felt as unpleasant can vary considerably in their negative character. An experience commonly reported by those

[6] Note that Grahek (2011) discusses a case where a patient reports an unpleasant feeling somewhere in his arm, but was unable to specify whether it was a pain or some other sensation. However, this does not seem to be a case of truly 'free-floating' unpleasantness, but rather one in which the unpleasantness simply has a relatively less determinate sensory character.

[7] However, see Klein (2015) for an alternative interpretation of pain asymbolia.

on analgesic drugs such as morphine is that their pains persist but cease to be particularly troublesome. Likewise, a given pain can bother us more or less on an individual occasion: a headache can be at one moment a mild nuisance, and at another a deeply troubling intrusion. Similarly, many of us will have experienced cases in which the degree of unpleasantness associated with a given emotion fluctuates over time. Someone tasked with giving an intimidating public lecture, for example, might go from a pleasant feeling of anticipation to an unpleasant feeling of anxiety even while the somatosensory aspects of their affective state – their physiological arousal and alertness, say, and their sense of an impending challenge – remain fairly constant.[8]

In light of this, then, an essential part of a theory of felt unpleasantness will be an account of how unpleasantness can arise in varying degrees in connection with different sensory and affective states at different moments. Many different answers to this question have been given by philosophers. Korsgaard (1996: 147), for example, takes the position that the painfulness of pain (in present terms, its unpleasantness) is constituted "the fact that these are sensations that we are inclined to fight." More specifically, she suggests that painfulness involves "the *perception* that we have a reason to change our condition." What is presumably lacking, then, in cases where pains are not felt as unpleasant is precisely this perception. Parfit offers a somewhat similar view, claiming that "the badness of a pain consists in its being disliked, and… it is not disliked because it is bad" (Parfit 1984: 501).

It should be noted that both Korsgaard and Parfit are concerned specifically with pain, rather than felt unpleasantness, and this only in the context of much more ambitious theoretical endeavors. Other theorists address the question of what makes states unpleasant more directly. Heathwood (2007), for example, spells out in considerable detail a view according to which a state's quality of felt unpleasantness consists in its being associated with a specific sort of desire that the state not be occurring, where this desire has that very state as its content (in Heathwood's preferred terms, a *de re* desire). Klein (2015b) considers two slightly different approaches to the question. The first of these, drawing from Kant, and with clear parallels to Korsgaard's account, holds that unpleasantness consist consists in a higher-order recognition of the negative motivational force of some first-order state. As Klein puts it, what makes states feel unpleasant is that they are accompanied by "judgments that those

[8] One might insist that it is part of the *definition* of characteristically unpleasantness emotions such as sadness and anxiety that they feel bad. While I would contend that this claim is somewhat at odds with our folk psychological understanding of varied subjective character of such states, I am happy to grant that, as a definitional matter, we individuate certain emotional states both by their characteristic psychological role and physiological effects *and* their hedonic tone.

states bear a certain motivational relationship to us." The second view Klein considers (and seems to prefer) differs from the first in respect of the type of content of the relevant higher-order state. Rather than being merely an appreciation or recognition of the motivational force of the first-order state, it is itself a kind of imperative or command with motivational force: "Don't have that sensation!"

I have considerable sympathy for some of these approaches. In particular, it seems clear to me (in light of the discussion in the previous section) that felt unpleasantness has constitutive connections to motivation. However, I differ from the approaches just mentioned insofar as they analyze felt unpleasantness in terms of some higher-order attitude such as disliking, desiring-that-not, or being inclined to fight, adopted towards a given sensory state.

This seems to be a fraught path to take, for two main reasons. The first is that it seems to me to risk *overintellectualising* felt unpleasantness. In feeling a sudden electric shock or wave of nausea as unpleasant, I do not need to recognize the source of my discomfort for it to be a ghastly experience. Likewise, it is not uncommon to undergo unpleasant experiences while in highly confused or disoriented states, as is sometimes the case when a person has a high fever or undergoes traumatic sleep events like night terrors. It seems to me somewhat implausible to insist that in all these cases we are able to so quickly adopt a metacognitive attitude towards our experience.

Of course, defenders of the theories like those mentioned above can respond that the relevant higher-order attitudes underpinning felt unpleasantness can be activated reflexively and extremely rapidly. While I do not discount this response, it strikes me as worryingly *ad hoc*, and I take it as advantage of my own account, to be discussed below, that it does not need to appeal to any further occurrent mental state in order to explain felt unpleasantness.

A further risk associated with overintellectualising felt unpleasantness is that it may dramatically constrain the range of animals that can undergo unpleasant states. This is not simply the worry, discussed below, that we may lack scientific grounds for assigning consciousness to animals. Rather, the worry is that, even if we judge that an animal is a promising consciousness candidate (Birch 2017; see discussion in 6, below), if it lacks a capacity for higher-order mental states (as it is commonly assumed most animals do) then its pains will not hurt nor will its anxiety trouble it.

Of course, perhaps this view is quite correct; we certainly cannot dismiss it lightly. But note that, at the very least, it might make felt unpleasantness seem peculiarly epiphenomenal, and thereby sever the initially promising connection between felt unpleasantness and a straightforward account of how they motivate us. To illustrate, let us

suppose that pain and nausea exist in animals and humans alike, perhaps even consciously, but only in the latter case are they accompanied by felt unpleasantness. In that case, it seems that felt unpleasantness *per se* would play little significant role in motivating the behavior we normally associate with pain and nausea; after all, an electric shock will motivate a dog or a rat as surely as it would a human. Felt unpleasantness, then, would be left as a peculiarly human phenomenon, perhaps connected to capacities such as reflection, but severed from the kinds of aversive responses we typically associate with states that feel bad.

Note that this objection applies primarily to views (such as Korsgaard's, and the first position considered by Klein) that take unpleasantness to consist in a higher-order appreciation or representation of some first order state. For higher-order views that take unpleasantness to consist in a higher-order *motivational* state (such as Heathwood, or Klein's second view), it would be the case that animals that lack appropriate higher-order states do indeed differ from humans in their broader motivational responses to pains, nausea, and so on. What this difference would be however, is far from clear, given that most animals exhibit very similar responses to characteristically unpleasant stimuli as human beings. Even Klein admits this view has a whiff of mystery, wondering why we should have "a state that drives you to get rid of pain… After all, you might think, the pain motivates you to get rid of pain in the best possible way, namely, by protecting yourself" (Klein 2015b).

3.3 – The Sensory-Motivational account of felt unpleasantness

I conclude, then, that we should prefer an account of felt unpleasantness, if one is available, that does not spell out felt unpleasantness in higher-order terms. I would suggest that such an account is available if we focus simply on the *motivational role* of states like pain, nausea, and anxiety. Specifically, I would suggest that what makes a sensory or affective state feel unpleasant is that it serve to motivate us in a specific way, namely to put an end to the state in question. Crucially, however, this does not require that a creature form any kind of higher-order mental state to the effect that it should put an end to the pain, nausea, and so on, but merely that it adopt behaviors that will have that effect. The felt unpleasantness will thus not consist in the occurrence of pain or nausea in isolation, nor on the occurrence of such a state accompanied by some further higher order state, but rather, the *effect* that that first-order state has on a creature's behavioral dispositions.

Let us call this theory *Sensory Motivation* theory, or SM for short. It can be formulated as follows.

(SM) Sensory-Motivation Theory: a subject is in a state of felt unpleasantness iff she undergoes a state with sensory or affective phenomenology and is behaviorally motivated to end or diminish that state.

If something like this view is correct, experiences of felt unpleasantness consist in having some kind of conscious sensory or affective state that negatively motivates a subject. To give a crude example, if I have a conscious sensory state such as a headache, and a consequence of having that headache is that my behavior is altered in such a way that I am disposed to put an end to it, then it would follow that that state was felt as unpleasant. The relevant motivation need not, of course, trump all other standing desires and motivations: I may have good reason for enduring pain in certain instances. What matters is that the state negatively affect my disposition to remain in that state.

Someone might wonder how a pain or other typically state could affect a creature's behavioral dispositions without that state being appreciated as bad or a reason for action. However, I would note that many motivational processes, including those underpinning instrumental learning, are accomplished by *subpersonal* mechanisms. Such mechanisms are extremely widespread in nature, and very simple creatures, including seaslugs and even nematode worms, display learned aversion to dangerous stimuli (Zhang, Lu, and Bargmann 2005) presumably without any metacognitive capacities (however, see Shea 2012). This is not to suggest that such creatures undergo felt unpleasantness, of course; there are doubtless important differences in the relevant mechanisms of learned aversion across different animal phyla, to be spelled out by comparative neurobiology (see also Section 4, below). However, I take this point to demonstrate that we can understand the notion of a state's having negative motivational force without reference to further personal-level psychological states.

I would suggest, then, that SM theory very simply captures the intrinsic negative motivational force I earlier argued was a basic characteristic of states of felt unpleasantness: what it is for a state to feel unpleasant, on this view, is precisely that it play such a motivational role. Additionally, the view can explain the typical commensurability of states of felt unpleasantness: such states, by their nature, negatively motivate us to varying degrees, and when we reflect on which of two experiences would feel worse, we are thereby

anticipating which feeling we would be more strongly disposed to end or diminish.[9]

However, the theory as stated is too simple, and vulnerable to various counterexamples. In particular, it seems that many conscious sensory or affective states alter our dispositions such that we act to put an end to them, without involving felt unpleasantness. Thus imagine that I am watching a televised documentary about the negative health effects of television and thereby acquire a motivation to turn off the set. In this case, we know that the visual and auditory sensations involved in watching the program do not constitute a case of felt unpleasantness. What seems to be missing in this case is instead the relevant sort of causal connection between the sensation and the motivation: rather than being motivated just by the character of the sensations themselves, I am motivated by the information conveyed in the broadcast. We might thus attempt to amend the theory as follows.

> **(SM-2) Sensory-Motivation Theory: a** subject is in a state of felt unpleasantness iff she undergoes a state with sensory or affective phenomenology and is thereby motivated to end or diminish that state in virtue of how it feels to her.

The 'in virtue of' clause here indicates simply that the motivation must be suitably connected to the relevant sensory state. Note that it does not require the recognition on the behalf of the subject that the way the state feels is a *reason* to put an end to it. Instead, a promising way to spell this out would be in terms of a relationship of reliable covariation between the state and its motivating role. For example, as the subjective intensity of the relevant unpleasant sensation increases, so too should the creature's motivation.

However, this version of the theory still requires a further amendment to deal with counterexamples in which a subject's other beliefs and desires might lead us to attribute felt unpleasantness inappropriately. For example, imagine that I taste a delicious piece of chocolate cake, and realize it tastes so delicious that I should stop eating it now, before I am overcome with greed and devour the whole thing. Although not a case of felt unpleasantness, this would fit the account given above: one would expect a reliable covariation between the subjective deliciousness of the taste of the cake and my motivation (the more delicious it is, the more worried I will be about eating it). Thus one final change is required.

---

[9] Note, of course, that most non-human animals likely lack the introspective capacities to engage in this kind of mental comparison of different states. Nonetheless, it seems fair to say that such states would still be *in principle* commensurable insofar as their varying motivational force would be available for introspection for any creature that possessed the requisite capabilities.

(SM-3) Sensory-Motivation Theory: a subject is in a state of felt unpleasantness iff she undergoes a state with sensory or affective phenomenology and is behaviorally motivated to end or diminish that state just in virtue of how it feels to her.

By requiring that the subject is or would be motivated *just* in virtue of the character of the state in question, and not depend on any other beliefs or desire she possesses, we can avoid the counterexample above; there, my motivation not to eat the cake depends on my desire not to be gluttonous, and is thus ruled out.

4. Towards scientific tractability

In this form given above, Sensory- Motivation Theory seems to offer a promising preliminary framework for understanding felt unpleasantness. I will now consider how it might be made applicable to the assessment of animal welfare, before considering some objections to the view. Note that I will set aside questions about consciousness for the time being, returning to them in Section 6, below.

An initial point to note in attempting to apply the theory to the assessment of animal welfare is that, without further elaboration, it risks ascribing felt unpleasantness extremely widely. I agree that the notion of negative motivation as stated is simply too broad. For example, forms of negative reinforcement can be found in suborganismal systems such as isolated sections of rats' spinal cords (Grau 2014). Other evidence shows that fish initiate escape behaviors even after having all frontal areas of their brain removed (Overmier and Gross 1974). In both of these cases, it seems unlikely (though not impossible) that any felt unpleasantness is occurring.

However, I would suggest that by the lights of the account given thus far we have *principled* grounds for wanting to constrain the relevant form of negative motivation to exclude such cases. In particular, note my earlier claim that it is essential to states of felt unpleasantness that they be (in principle) commensurable. In the human case, our ability to compare different states in respect of their degree of felt unpleasantness surely relies on quite sophisticated cognitive capacities such as introspection and imaginative projection that are not available to most animals. However, this does not mean that the relative felt intensities of different unpleasant states play no role in non-human animals.

Instead, in at least some cases, states like pain, nausea, and anxiety exhibit varying

degrees of influence on the selection of behavior on the part of the creature, leading to a phenomenon known as motivational tradeoff. This is the tendency of some organisms to willingly undergo one negative state in order to avoid another even more negative one, or come to highly value a previously neutral stimulus if it will relieve a negative stimulus.

This provides us with a way of assessing of gauging which sensations and emotions exert the strongest motivational effect on animal's behavior, namely by measuring which they most strenuously act to avoid. Note that such motivational tradeoff behavior in various forms has been demonstrated in a wide range of animals (see Dawkins 2012: 150-75, for a review) as follows.

- Rats, in a reversal of their normal preferences, will prefer a light chamber to a dark chamber in order to avoid unpleasant mechanical stimulation of an injured paw (LaBuda and Fuchs 2000).
- Broiler chickens, who are normally strongly disinclined to jump over high barriers unless food-deprived, will do so quite spontaneously without any food deprivation in order to get away from highly-crowded enclosures (Buijs, Keeling, and Tuyttens 2011).
- Zebrafish normally prefer to swim to an environmentally enriched chamber rather than a barren, brightly lit one, but their preferences are reversed if the fish are injected with an irritating acid and the bright chamber is filled with an analgesic (L. Sneddon 2013).
- Trout, highly social animals, will endure normally disliked electric shocks in order to avoid being isolated from other fish (Braithwaite 2010: 104-5).
- Experiments along these lines can also discern quite subtle preferences. One set of experiments (Rushen and Congdon 1986) for example, suggested that sheep prefer being manually restrained while being sheared compared to being electrically immobilized.

In all of these cases, we see evidence of animals engaging in behavior they would normally be disinclined to perform in order to avoid some seemingly unpleasant stimulus (or to obtain relief from an existing negative state). This suggests that they are undergoing or have undergone states that do not merely negatively reinforce single behaviors, but give rise to a set of weighted preferences that can flexibly influence longer term decision-making at the

whole organism level. These states extend well beyond pain, including other forms of apparently unpleasant state, such as the distress involved in being in crowded enclosures or being kept away from conspecifics. Additionally, because motivational tradeoff behavior requires that a creature's behavior be sensitive to multiple competing interests and involve the adjustment of typical behavior in accordance with circumstances, it is hard to see how it could simply be a matter of hardwired reflex action, but instead seems to reflect the presence of a broader psychological economy sensitive to positive and negative states of varying intensity. It thus provides a principled way of discounting highly inflexible forms of aversion learning like those mentioned earlier.

I conclude, then, that, motivational tradeoff seems to offer the basis for a promising and scientifically applicable notion of motivation to fill out the theory given above. Specifically, those sensory and affective states that can serve a negative motivational role within a broader set of motivational tradeoff behavior seem plausible candidates for states of felt unpleasantness. We could thus incorporate motivational tradeoff behavior into the framework given earlier with the following condition.

> **Evidence for felt unpleasantness:** we should attribute felt unpleasantness to a creature if we have good reason to think that it is undergoing a conscious sensory or affective state and is behaviorally motivated to end or diminish that state just in virtue of how it feels, where this motivation exhibits flexibility in accordance with the creature's other goals.

While this condition does not spell out what is involved in a state's being conscious to begin with, it could, when supplemented with an account of consciousness, provide us with a way of assessing felt unpleasantness in animals.

Note that I do not suggest that any and all negative motivational tradeoff behavior automatically involves suffering. Animals may have motivations (perhaps including some instinctive aversions or appetitive compulsions) that are reflected in their behavioral preferences but not appropriately sensitive to any kind of sensory or emotional state at all, thus falling short of the conditions for experiential suffering laid out above.

More work on the mechanisms of motivational tradeoff is also required, and it is possible that this apparently sophisticated behavior is underpinned by different mechanisms in different creatures; comparable behavior may be present in surprisingly simple systems where we are inclined to rule out felt unpleasantness (see Castro-González, Malfaz, and

Salichs 2013 for robots that can engage in superficially similar behavior). It is also possible that this account underestimates the reach of felt unpleasantness; perhaps it could occur in simpler creatures that are incapable of the relatively sophisticated behavior seen in motivational tradeoff paradigms. As matters stand, however, motivational tradeoff may serve as fairly compelling evidence for the presence of suffering.[10]

5. Problems for the Sensory-Motivational account

The account of felt unpleasantness I have argued for thus far has many virtues. It captures what I took to be the central traits of felt unpleasantness as a psychological kind, namely its motivational force and its commensurability, and does so without the need for appeal to higher-order states. It accurately circumscribes the instances of felt unpleasantness in human experience with which we are familiar, while also lending itself naturally to the assessment of suffering in non-human animals.

Nonetheless, it faces some serious challenges. In the present section, I will consider two objections concerning the connection between felt unpleasantness and motivation, before going on in Section 6 to consider some worries relating to consciousness.

The first objection under present consideration comes from experiments on 'liking' and 'wanting' behavior. The work of Kent Berridge has demonstrated a divergence between both the behavioral and neural measures of positive experience, 'liking', and motivation, 'wanting' (see Berridge, Robinson, and Aldridge 2009 for a review). One important such experiment is that of Peciña et al. (2003). In this paradigm, ordinary mice and hyperdopaminergic mice were both given rewards for learning to perform certain tasks. The mice's motivation (their "wanting" behavior) was measured using several different metrics, such as speed and the number of pauses in performing the task. The amount of pleasure they took in the reward (their "liking" behavior) was measured by characteristically happy behavior such as licking of the mouth and paws. Results showed that the hyperdopaminergic mice exhibited significantly more "wanting" behavior and were more motivated to complete the relevant tasks, but did not display a similar increase in "liking" behavior. Combined with neuroscientific evidence, this supports the idea that dopamine circuits underwriting motivation can be disassociated from those involved in subjective pleasure.

This creates the following worry for the sensory-motivational view of felt

[10] Note that several other authors have suggested motivational tradeoff as a signature of pain (L. U. Sneddon et al. 2014), consciousness (Godfrey-Smith 2017), and sentience (Birch 2017).

unpleasantness.[11] Given that motivation and positive experience dissociate for pleasure, is it not plausible that they may also come apart in the case of felt unpleasantness? More specifically, the sensory-motivational account relies on the idea that we can assess the occurrence of felt unpleasantness by extrapolating from the extent that organism is negatively motivated by a given state within the broader context of a motivational tradeoff paradigm. However, if the motivational role of negative reinforcement can be severed from felt unpleasantness, then this idea looks to be in jeopardy.

There are a few moves available to the sensory-motivational theorist at this point. First, it is not obvious that the dissociation between "liking" and "wanting" also applies to cases of felt unpleasantness" and "not wanting": there are many important asymmetries between positively- and negatively-valenced experiences.[12] Still, we can seemingly imagine a case in which two animals might exhibit different responses to an aversive stimulus while experiencing the same degree of felt unpleasantness. Certainly, we think of such a phenomenon as occurring in the human case: two pains felt by different people might result in one screaming and yelling, and the other merely biting their lip, even while being felt as equivalently unpleasant.

The best response on behalf of the sensory-motivation theorist to cases such as these, I would suggest, is to grant that the relationship between degree of felt unpleasantness and aversive behavior is not an absolute one: different creatures may have the same experience of felt unpleasantness while behaving in quite different ways.

This might seem at odds with the claim that felt unpleasantness consists in the negatively motivating role played by a given sensory or affective state. However, this depends on how we construe the negative motivational force of a given stimulus. Rather than linking it absolutely to aversive behaviors such as speed of attempted escape, we might instead *index* it to the broader range of behavioral responses available to a given animal. Somewhat roughly, the suggestion is that for any two behaviors, B_1 and B_2, B_1 could be understood as involving a greater degree of felt unpleasantness than B_2 if it resulted in a greater aversive response (for example, leading the creature to prefer B_2 to B_1 in motivational tradeoff frameworks).

Admittedly, this would limit the degree to which we could make interpersonal or inter-animal assessments of relative degrees of felt unpleasantness: simply by observing that

[11] For a more developed version of this objection, directed specifically against the idea that unpleasantness is a function of aversiveness in pain, see Corns (2014).
[12] See Shriver (2014) for a more developed argument to this effect.

two creatures exhibited different aversive behavioural responses to a given stimulus, it would not follow that they underwent different degrees of felt unpleasantness. However, it would still allow us to determine, first, whether felt unpleasantness is occurring in a given creature, and second, which of two situations that a creature might be exposed to resulted in a greater degree of felt unpleasantness. It also preserves what I take to be the core insight of Sensory-Motivation theory, namely that felt unpleasantness consists in a kind of direct alteration of a creature's behavioral dispositions brought about by a sensory or affective state.

A second important objection relating to felt unpleasantness and motivation comes from the worked on 'learned helplessness' in animals. This phenomenon was famously shown by Martin Seligman and colleagues, who administered electric shocks to different groups of dogs (Seligman 1972). One group of dogs was able to terminate the electric shock by pressing a lever, while the other had no control over the duration of the shocks. In a subsequent learning task, the two groups of dogs were placed in a situation in which they received electric shocks, but could avoid them by leaping over jumping out of their enclosure. The group of dogs that had learned to terminate the electric shocks in the earlier task readily learned this new escape behavior, but the dogs that were formerly 'helpless' did not, instead simply lying down and whimpering while the shocks were administered. This result has been replicated with other animals, and has also been applied to understanding various forms of 'helpless' behavior in human beings, notably in cases of depression (Abramson, Seligman, and Teasdale 1978).

The basic phenomenon of learned helplessness may seem to pose a threat to the sensory-motivational view of experiential suffering since it seems likely that the helpless dogs in Seligman's original experiment were undergoing unpleasant experiences, yet seemingly failed to take advantage of the opportunity to put an end to their shocks. Given this, one might infer that they were not *motivated* to put an end to the shocks that were being administered. If correct, this case would be a clear counterexample to my theory.

However, it is a matter of contention whether the helpless dogs in Seligman's experiment were not motivated to escape. In particular, they *did* commence escape behaviors if they were shown how to perform them – specifically, by having their legs moved by an experimenter in such a way as to remove them from the enclosure. In other words, once it was demonstrated to the dogs that there was indeed a viable approach to avoiding the shocks, they took advantage of it. This suggests that learned helplessness does not involve a deficit in motivation per se, but rather, a pessimism by a creature about its ability to escape a given

negative stimulus. After all, someone can be highly motivated to do something, but if they do not believe that there is any viable way for them to achieve it, this will not be evident in their behavior (at least up until the point where they are convinced otherwise).

6. Suffering and consciousness

A final important set of objections for Sensory-Motivation Theory concern consciousness. I have claimed that only conscious states can feel unpleasant, while also suggesting that we might be able to make reasonable empirical assessments of whether a given animal is in a state of felt unpleasantness. This in turn assumes that we could, at least in principle, determine whether a given non-human animal is in a conscious state. Putting it mildly, this is quite a demand.

In defense of this proposal, I would note that, quite independent of questions about consciousness, we require a theory of felt unpleasantness if we are to improve our treatment of non-human animals: even if we could establish that a given animal was conscious, it would not answer the further question of whether it could suffer.

However, a vital worry for the theory would arise if it was held that there were no good grounds, even in principle, for attributing consciousness to animals. In light of this, it is worth considering important lines of criticism levied by Dawkins (2008, 2015) and by Carruthers (Carruthers 2018a, 2018b) against the idea that consciousness should play a central account in a theory of animal welfare.[13]

In essence, Dawkins' argument is that we have so little grasp on the nature of consciousness or the states or behaviors with which it is associated that it cannot ground a theory of animal welfare. As she puts it, "[t]he problem is that we know so little about human consciousness… that we do not know what publicly observable events to look for in ourselves, let alone other species, to ascertain whether they are subjectively experiencing anything like our suffering" (2008). Instead, Dawkins suggests, we should base our assessment of animal welfare on more scientifically tractable considerations, specifically the health of animals and the satisfaction of their strong preferences.

While Carruthers is sympathetic to Dawkins' suggestions for how we should approach questions of animal welfare (Carruthers 2018a), his reasons for setting aside

[13] Both Dawkins and Carruthers raise their criticisms within complex and sophisticated frameworks for thinking about animal cognition, hence the following discussion is somewhat cursory. Nonetheless, I hope it will suffice to at least illustrate the main points of divergence between the approach defended here and their perspectives.

questions of animal consciousness are somewhat different. His position derives from a commitment to the Global Workspace Theory of consciousness (Dehaene and Naccache 2001) which he takes to have considerable scientific and philosophical support. He then presents a dilemma relating to animal consciousness for those who share some sympathy for Global Workspace Theory concerning two positions he terms "qualia realism" and "qualia irrealism". Somewhat crudely, he takes qualia realism to be the view that phenomenal consciousness "attaches to" globally broadcast states without being strictly reducible to it, instead involving some further (perhaps neurobiological) property. By contrast, qualia irrealism is the view that "phenomenally conscious experience just *is* globally broadcast nonconceptual content."

Carruthers suggests that, for the qualia realist, "the question of animal consciousness becomes intractable", on the grounds that without recourse to report, we cannot know if a given behavioral response is associated with consciousness. Nor will observed similarities between animal behavior and behavior that is typically conscious in humans suffice to answer such questions (cf. Tye 2017), since, Carruthers claims, we will be unable to determine in any given case whether it is qualia per se that underpin the relevant behavior, or instead unconscious states with similar functional roles (2018b: 197).

For qualia *irrealists* who endorse Global Workspace Theory, Carruthers takes the view that questions about animal consciousness simply do not arise and do not matter. The differences in cognitive architecture between humans and animals are such, he suggests, that no non-human animal will straightforwardly instantiate the kind of cognitive properties entailed by global broadcast in humans such as metacognitive access and reportability of one's first-order states. That need not stop us drawing comparisons between the cognitive architectures of humans and those of animals, of course, but there will be no principled basis, Carruthers claims, for determining what degree of similarity will be sufficient for consciousness. As he puts it, "as we transition from species whose cognitive architecture is quite unlike that of human global broadcasting through species whose networks are increasingly similar to our own, nothing lights up, and nothing magical appears… [t]here is just nonconceptual content that is available to a greater range of systems."

Dawkins and Carruthers have well developed arguments in support of their views, and detailed engagement with them is beyond the scope of this paper. However, it will be helpful to draw attention to two reasons why I do not regard their arguments as fatal for the present endeavor.

The first point I would stress, in response to both Dawkins and Carruthers, is that states of consciousness, and specifically experiences of felt unpleasantness, play a central role in our folk psychological understanding of suffering and moral value. A creature that is not conscious or sentient has no intuitive command on our moral sentiments. Granted, we might value it for reasons of biodiversity, or natural beauty, or scientific utility; but if there is nothing it is like to *be* that creature, then any attitude of sympathy we adopt towards it will be misplaced. Singer puts the point well.

> The capacity for suffering and enjoying things is a prerequisite for having interests at all, a condition that must be satisfied before we can speak of interests in any meaningful way. It would be nonsense to say that it was not in the interests of a stone to be kicked along the road by a schoolboy. A stone does not have interests because it cannot suffer. (Singer 1989)

In light of this, the idea that we should develop an ethics of animal welfare that does not make reference to consciousness at all can be interpreted in two ways. Either it amounts to the claim that we radically revise our intuitive ethical practices and dispense with notions like animal suffering all together in favor of concepts like health and preference satisfaction, or else the claim we should adopt a purely anthropocentric notion of animal welfare, in which the moral status of animals consists solely in their contribution to human interests.

It should scarcely need to be said that either option would entail radical revisions to our moral attitudes and practices. We do not, as matters stand, regard animal health and the satisfaction of their preferences as ends in themselves, independent of their connection to suffering or human interests; it is hard to see why the physical health of a brain-dead animal should be of any ethical significance, after all. Likewise, nothing in Dawkins' framework explains why – or whether – we should value the health and preferences of intelligent animals such as whales or chimpanzees above those of insects or shellfish, even though we certainly do regard the former as more important targets of moral concerns.

I should stress that I regard it as an open possibility that Carruthers and Dawkins are correct in claiming that a science of animal welfare grounded in animal consciousness will prove to be impossible or void of philosophical interest. However, as I will argue below, this claim is far from obvious in light of the current unsettled status of consciousness science. Given the drastic revisions to our moral attitudes and practices that would be required by their proposals, then, I would suggest that Dawkins' and Carruthers' worries would require

far stronger motivation before they deter us from the hope of giving a principled account of felt unpleasantness in non-human animals.

The second reason I think Dawkins' and Carruthers' arguments can be set aside for present purposes is that, in short, their views of the state of consciousness science are respectively overly pessimistic and overly optimistic. While Dawkins sees little prospect of a scientifically applicable theory of consciousness, Carruthers takes Global Workspace Theory to be a conceptually adequate account of consciousness, with only fine details to be resolved.

Neither position, so it seems to me, quite captures the current state of scientific work on consciousness. As Carruthers acknowledges, current opinion among philosophers and scientists is divided among a range of different approaches to consciousness, including higher-order thought theories (Rosenthal 2005), integrated information theory (Massimini and Tononi 2018), several varieties of global workspace theory (Dehaene and Naccache 2001; Kouider et al. 2010), and views that link consciousness to some form of fragile short-term memory (Block 2007; Shevlin 2017). Matters are not hopeless, however, and considerable progress and development has been made across *all* these research programs, and our understanding of the neural and psychological basis of different forms of conscious experience is steadily improving. Additionally, it should be noted that much scientific work on consciousness has been fairly narrowly focused on humans, and in many areas of the field, it is only relatively recently that serious efforts have been made towards integration with biological perspectives on the evolution of consciousness and its function (Feinberg and Mallatt 2016; Godfrey-Smith 2017)

Of course, the lack of current consensus does not mean that philosophers and scientists should refrain from giving arguments and presenting evidence in support of their preferred approaches. However, in developing broad scientifically-informed approaches to animal welfare, I think it premature to draw firm conclusions about the neural and psychological basis of consciousness.

As a final point, note that one might reasonably worry, in light of the current unsettled nature of the science of consciousness, that the current enterprise is similarly premature. How can one hope to give a consciousness-based account of felt unpleasantness if there is so little consensus on the neural and cognitive mechanisms underlying consciousness to begin with? Here, I would suggest that the notion of a *sentience candidate* introduced by Birch (2017) may prove to be useful. Rather than tie ourselves rigidly to a particular account of consciousness, we can instead evaluate how different animals fare according to different measures of consciousness. We can make then make assessments of the probability of

consciousness across different animal species. These, to be sure, are likely to be highly preliminary and inaccurate to begin with, but can be updated in light of progress across different branches of the science of consciousness. These initial assessments of the likelihood of consciousness in different animals can then be combined with an account of felt unpleasantness like Sensory-Motivation theory to enable us to make defeasible but informed judgments about the occurrence of unpleasant experience in a given case.

Conclusion

The primary goal of this paper has been to defend a novel account of felt unpleasantness, Sensory-Motivation theory, and to suggest how it could be applied to the practical assessment of animal welfare, namely via motivational tradeoff paradigms. As noted earlier, I acknowledge that this theory is a merely preliminary attempt to grapple with the daunting problem of how we might develop an objective framework for assessing the occurrence of felt unpleasantness. It is my hope that the theory can be developed and refined in accordance with philosophical responses and new empirical evidence.

The paper has had several other goals along the way, however. I have argued that felt unpleasantness is a concept with considerable ethical import, and that we might fruitfully consider it a putative psychological kind. Likewise, I have argued that philosophical worries concerning the challenges involved in assessing animal consciousness should not deter us from attempting to set the assessment of animal welfare on a firmer theoretical footing. Hence even if the reader remains unconvinced of the merits of Sensory-Motivation theory in particular, it my hope that these others claims may ring true, and contribute to the ongoing interdisciplinary project of understanding and assessing felt unpleasantness in non-human animals.

References

Abramson, L. Y., M. E. Seligman, and J. D. Teasdale. 1978. “Learned Helplessness in Humans: Critique and Reformulation.” *Journal of Abnormal Psychology* 87 (1): 49–74.

Aydede, Murat. 2000. “An Analysis of Pleasure Vis-a-Vis Pain.” *Philosophy and Phenomenological Research* 61 (3): 537–70.

Berridge, Kent C, Terry E Robinson, and J Wayne Aldridge. 2009. “Dissecting Components of Reward: ‘Liking’, ‘Wanting’, and Learning.” *Current Opinion in Pharmacology* 9 (1): 65–73. https://doi.org/10.1016/j.coph.2008.12.014.

Birch, Jonathan. 2017. “Animal Sentience and the Precautionary Principle.” *Animal Sentience* 2: 16(1).

Block, Ned. 2007. “Consciousness, Accessibility, and the Mesh between Psychology and Neuroscience.” *Behavioral and Brain Sciences* 30 (5): 481--548.

Braithwaite, Victoria. 2010. *Do Fish Feel Pain?* OUP Oxford.

Buijs, Stephanie, Linda J. Keeling, and Frank A. M. Tuyttens. 2011. “Using Motivation to Feed as a Way to Assess the Importance of Space for Broiler Chickens.” *Animal Behaviour* 81 (1): 145–51. https://doi.org/10.1016/j.anbehav.2010.09.027.

Carruthers, Peter. 2005. *Consciousness: Essays From a Higher-Order Perspective*. Oxford University Press UK.

———. 2018a. “Comparative Psychology without Consciousness.” *Consciousness and Cognition* 63: 47–60.

———. 2018b. “The Problem of Animal Consciousness.” *Proceedings and Addresses of the American Philosophical Association* 92.

Castro-González, Álvaro, María Malfaz, and Miguel Angel Salichs. 2013. “An Autonomous Social Robot in Fear.” *IEEE Transactions on Autonomous Mental Development* 5 (2): 135–151.

Corns, Jennifer. 2014. “Unpleasantness, Motivational Oomph, and Painfulness.” *Mind and Language* 29 (2): 238–54.

Dawkins, Marian. 2015. “Chapter Two - Animal Welfare and the Paradox of Animal Consciousness.” In *Advances in the Study of Behavior*, edited by Marc Naguib, H. Jane Brockmann, John C. Mitani, Leigh W. Simmons, Louise Barrett, Sue Healy, and Peter J. B. Slater, 47:5–38. Academic Press. https://doi.org/10.1016/bs.asb.2014.11.001.

Dawkins, Marian Stamp. 2008. “The Science of Animal Suffering.” *Ethology* 114 (10): 937–45. https://doi.org/10.1111/j.1439-0310.2008.01557.x.

———. 2012. *Why Animals Matter: Animal Consciousness, Animal Welfare, and Human Well-Being*. OUP Oxford.

Dehaene, S., and L. Naccache. 2001. “Towards a Cognitive Neuroscience of Consciousness: Basic Evidence and a Workspace Framework.” *Cognition* 79 (1–2): 1–37.

Dennett, Daniel C. 1996. *Kinds of Minds*. Basic Books.

Griffiths, Paul, E. 1997. *What Emotions Really Are: The Problem of Psychological Categories*. University of Chicago Press.

Feinberg, Todd E., and Jon M. Mallatt. 2016. *The Ancient Origins of Consciousness: How the Brain Created Experience*. MIT Press.

Foltz, E. L., and L. E. White. 1962. “Pain ‘Relief’ by Frontal Cingulumotomy.” *Journal of Neurosurgery* 19 (February): 89–100. https://doi.org/10.3171/jns.1962.19.2.0089.

Godfrey-Smith, Peter. 2017. *Other Minds: The Octopus and the Evolution of Intelligent Life*. HarperCollins UK.

Grahek, Nikola. 2011. *Feeling Pain and Being in Pain*. MIT Press.

Grau, James W. 2014. “Learning from the Spinal Cord: How the Study of Spinal Cord Plasticity Informs Our View of Learning.” *Neurobiology of Learning and Memory* 0 (February): 155–71. https://doi.org/10.1016/j.nlm.2013.08.003.

Greene, Joshua D., India Morrison, and Martin E. P. Seligman. 2016. *Positive Neuroscience*. Oxford University Press USA.

Heathwood, Chris. 2006. “Desire Satisfactionism and Hedonism.” *Philosophical Studies* 128 (3): 539–63.

———. 2007. “The Reduction of Sensory Pleasure to Desire.” *Philosophical Studies* 133 (1): 23–44.

James, William. 1884. “What Is an Emotion?” *Mind* 9 (34): 188–205.

Klein, Colin. 2015a. “What Pain Asymbolia Really Shows.” *Mind* 124 (494): 493–516.
———. 2015b. “What the Body Commands : The Imperative Theory of Pain.”
Korsgaard, Christine M. 1996. *The Sources of Normativity*. Vol. 110. Cambridge University Press.
Kouider, Sid, Vincent de Gardelle, Jérôme Sackur, and Emmanuel Dupoux. 2010. “How Rich Is Consciousness? The Partial Awareness Hypothesis.” *Trends in Cognitive Sciences* 14 (7): 301–7.
Kringelbach, Morten L., and Kent C. Berridge. 2017. “The Affective Core of Emotion: Linking Pleasure, Subjective Well-Being, and Optimal Metastability in the Brain.” *Emotion Review : Journal of the International Society for Research on Emotion* 9 (3): 191–99. https://doi.org/10.1177/1754073916684558.
LaBuda, C. J., and P. N. Fuchs. 2000. “A Behavioral Test Paradigm to Measure the Aversive Quality of Inflammatory and Neuropathic Pain in Rats.” *Experimental Neurology* 163 (2): 490–94. https://doi.org/10.1006/exnr.2000.7395.
Massimini, Marcello, and Giulio Tononi. 2018. *Sizing Up Consciousness: Towards an Objective Measure of the Capacity for Experience*. Oxford University Press.
Nozick, Robert. 1974. *Anarchy, State, and Utopia*. Vol. 7. Basic Books.
Overmier, J. B., and D. Gross. 1974. “Effects of Telencephalic Ablation upon Nest-Building and Avoidance Behaviors in East African Mouthbreeding Fish, Tilapia Mossambica.” *Behavioral Biology* 12 (2): 211–22.
Parfit, Derek. 1984. *Reasons and Persons*. Oxford University Press.
Peciña, Susana, Barbara Cagniard, Kent C. Berridge, J. Wayne Aldridge, and Xiaoxi Zhuang. 2003. “Hyperdopaminergic Mutant Mice Have Higher ‘Wanting’ but Not ‘Liking’ for Sweet Rewards.” *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 23 (28): 9395–9402.
Pereplyotchik, David. 2017. “Pain and Consciousness.” In *The Routledge Handbook of Philosophy of Pain*. Routledge, London.
Prinz, Jesse J. 2006. *Gut Reactions: A Perceptual Theory of Emotion*. Oup Usa.
Rosenthal, David. 2005. *Consciousness and Mind*. Oxford University Press UK.
Rushen, J., and P. Congdon. 1986. “Relative Aversion of Sheep to Simulated Shearing with and without Electro-Immobilisation.” *Australian Journal of Experimental Agriculture* 26 (5): 535–37. https://doi.org/10.1071/ea9860535.
Schroeder, Timothy. 2004. *Three Faces of Desire*. Oxford University Press.
Seligman, M E P. 1972. “Learned Helplessness.” *Annual Review of Medicine* 23 (1): 407–12. https://doi.org/10.1146/annurev.me.23.020172.002203.
Shea, Nicholas. 2012. “Reward Prediction Error Signals Are Meta-Representational.” *Nous (Detroit, Mich.)* 48 (2): 314–41. https://doi.org/10.1111/j.1468-0068.2012.00863.x.
Shevlin, H. 2017. “Conceptual Short-Term Memory: A Missing Part of the Mind?” Text. 2017. https://www.ingentaconnect.com/contentone/imp/jcs/2017/00000024/f0020007/art00009;jsessionid=an9n312plnkkr.x-ic-live-01.
Shriver, Adam. 2014. “The Asymmetrical Contributions of Pleasure and Pain to Animal Welfare.” *Cambridge Quarterly of Healthcare Ethics* 23 (2): 152–62.
Singer, Peter. 1989. “All Animals Are Equal.” In *Animal Rights and Human Obligations*, edited by Tom Regan and Peter Singer, 215--226. Oxford University Press.
Sneddon, Lynne. 2013. “Do Painful Sensations and Fear Exist in Fish?” In *Animal Suffering: From Science to Law, International Symposium*, edited by T.A. van der Kemp and M. Lachance, 93-112. Toronto: Carswell.
Sneddon, Lynne U., Robert W. Elwood, Shelley A. Adamo, and Matthew C. Leach. 2014. “Defining and Assessing Animal Pain.” *Animal Behaviour* 97 (November): 201–12. https://doi.org/10.1016/j.anbehav.2014.09.007.
Tye, Michael. 2017. *Tense Bees and Shell-Shocked Crabs: Are Animals Conscious?* Oxford University Press.
Zhang, Yun, Hang Lu, and Cornelia I. Bargmann. 2005. “Pathogenic Bacteria Induce Aversive Olfactory Learning in Caenorhabditis Elegans.” *Nature* 438 (7065): 179–84. https://doi.org/10.1038/nature04216.