

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository:<https://orca.cardiff.ac.uk/id/eprint/135985/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Irvine, Elizabeth 2020. Developing valid behavioural indicators of animal pain. *Philosophical Topics* 48 (1), pp. 129-154.

Publishers page: <https://www.jstor.org/stable/48628589>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See <http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.



Developing Valid Behavioural Indicators of Animal Pain

Abstract:

Identifying which non-human animal species are capable of feeling pain is important both for understanding pain mechanisms more generally, and for informing animal welfare regulations, particularly in genera that are not yet widely protected. A common way to try to provide evidence of pain experiences is through behavioural indicators. In this paper I use a very simple interventionist approach to experimentation, and the contrast case provided by *C. elegans*, to argue that behavioural indicators commonly used for identifying pain in non-human animals are much less robust than typically presented. Indeed, I argue that many behavioural indicators of pain are invalid as they are currently described. More positively, this analysis makes it possible to identify what valid criteria might look like, and where relevant, identify existing evidence related to them. Based on this I propose that the best way to make progress on questions around animal pain is to clearly ally them with questions about animal consciousness more generally, and to productively use conceptual and empirical work in both areas to develop more theoretically defensible behavioural indicators.

Acknowledgments: TBA

1. Introduction

Like many philosophers, I am open to the possibility of consciousness (including pain experiences) being widespread across the animal kingdom¹. Much of the evidence levered in favour of consciousness in the more 'controversial' genera across fish, crustaceans and insects, outlining their behavioural and neurophysiological complexity, seems strongly suggestive at the least (e.g. Barron & Klein, 2016; Elwood, 2019; Sneddon, 2013; Sneddon et al., 2014). However, in this paper I use a simple interventionist approach to experimentation, and the contrast case provided by *C. elegans*, to argue that behavioural criteria for identifying non-human animal pain are much less robust than typically presented. Indeed, I argue that many behavioural indicators of pain are invalid as they are currently described. This is not intended to undermine existing claims about pain experiences in these genera, but is part of an attempt to evaluate what valid criteria might look like, and where relevant identify existing evidence related to them. Overall, I propose that the best way to make progress on questions around animal pain is to clearly ally them with questions about animal consciousness more generally, and to productively use conceptual and empirical work in both areas to develop more theoretically defensible behavioural indicators. Getting this right is important both for understanding pain mechanisms more generally, and in informing animal welfare regulations, particularly in genera that are not yet widely protected.

¹ Though see e.g. Carruthers, this volume, on whether this claim is meaningful.

There has so far not been much systematic philosophical engagement with behavioural criteria used to identify non-human animal pain experiences, particularly as applied to the more controversial genera. One possible reason for this is that challenges to claims about animal pain are usually associated with the no-cortex-no-consciousness challenge, made very prominently by Key (2016; see also e.g. Rose et al., 2014). As the name suggests, the challenge stems from the claim that cortical brain areas are necessary for consciousness, and since they are not present in the more controversial genera, they are incapable of pain experiences. So far, relatively few find the no-cortex-no-consciousness challenge persuasive, with replies either suggesting that cortical areas may not be required for consciousness in humans, and/or that it is an open question what sort of brain structures might support consciousness in other species (see e.g. responses to Key by Damasio & Damasio, 2016; Edelman, 2016; Merker, 2016; Seth, 2016). One possibility is that since the no-cortex-no-consciousness challenge is not deemed to be persuasive, allied critiques of behavioural indicators are too. However, as I argue below, there are legitimate and fairly serious concerns to be raised about them.

In what follows I use a simple interventionist approach to experimentation for assessing the validity of behavioural indicators of pain experiences (Craver, 2007; Kästner & Andersen, 2018; Woodward, 2003). Roughly, this works as follows. To investigate a target (e.g. putative pain experiences) one must be able to reliably causally intervene on it, and for it to have specific and observable causal effects. If an experimental paradigm does not successfully intervene on the target state, and/or if the wrong set of causal effects are observed, then any measurements based on these are not valid. The relationship between nociception, which refers to the non-conscious sensory processing of aversive stimuli, and

conscious pain experiences, is sufficiently complex for this to be a useful framework to structure the discussion.

I also make a possibly controversial assumption: that the nematode *C. elegans* is not conscious and does not have pain experiences. This may simply be question begging, and the evidence presented here may instead persuade some that *C. elegans* is a strong contender in the consciousness stakes. However, I take this assumption to be reasonably defensible. While consciousness is unlikely to be determined by how many neurons a species has in any straightforward way, *C. elegans* is rather at the low end of the scale with 302, and (arguably) little in the way of centralized information processing. Instead, *C. elegans* shows how an organism can display remarkable behavioural complexity with very little in the way of a complex central nervous system. This is one of the reasons why it continues to be used widely as a model organism in biology (for an overview of *C. elegans*, see e.g. Corsi et al., 2015). In this way it provides a useful contrast case between, on the one hand, the sometimes surprising behaviours that are plausibly driven by nociceptive processing and simple forms of non-conscious learning alone, and on the other, behaviours unique to pain experiences.

In what follows I apply this basic interventionist framework, and the contrast case of *C. elegans*, to the list of criteria for identifying pain experiences in the more controversial genera from Sneddon et al. (2014), which I take to be representative of the basic principles of reasoning used across much of this literature. Combined, these approaches show that many behavioural indicators of pain are invalid as they are currently described. Instead, I argue that where they are valid or can be made so, it is because of their relationship to

plausible theoretical claims about the features of consciousness in general. That is, valid behavioural indicators of pain are so in virtue of referring to capacities for performing complex forms of learning and/or complex decision making associated with consciousness. While highlighting the significant methodological challenges in pain research, I suggest throughout how existing indicators could be developed to be more theoretically robust, and identify evidence relevant to these. This shows that there are still good reasons to think that pain experiences are widespread across the animal kingdom.

A note: This paper is officially neutral about whether pain (vs nociception) is always conscious. I am primarily interested though in pain that is conscious. I therefore typically refer to 'pain experiences' to make that clear in the discussion below.

The paper proceeds as follows. In Section 2 I outline a very basic interventionist approach to experimentation and use this to differentiate between valid and reliable behavioural indicators of nociception and pain. In Section 3 I present the principles and criteria used to identify pain experiences from Sneddon et al. (2014), which I take to be fairly representative of those used in this literature. In Section 4 I analyse the validity of behavioural indicators of pain experiences, making recommendations about their development where relevant. Section 5 concludes.

2. Interventionism, nociception and pain

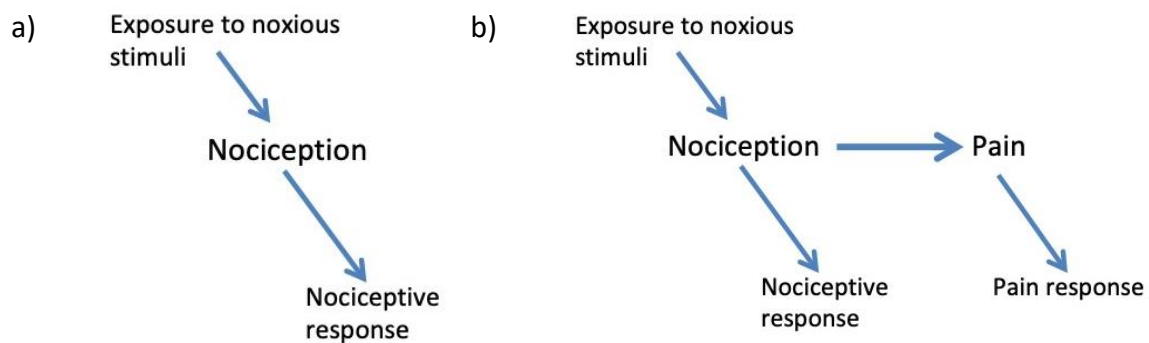
An interventionist approach to experimentation is a structured way of understanding causal relationships between experimental interventions, target states, and observed effects (Corsi

et al., 2015). In order for observed effects (e.g. animal behaviour) to be valid indicators of a target state (e.g. pain experiences), at least two things must obtain. First, the experimental intervention must have a causal effect on the target state. Second, the observed effects must be produced by a change in the target state. There are many more relevant criteria but these are the bare essentials, and are sufficient for this discussion (for more, see e.g. Craver, 2007; Kästner & Andersen, 2018; Woodward, 2003).

Figure 1 compares the causal structures of nociception and pain. Nociception (Figure 1a) refers to the non-conscious process of sensing and responding to (potentially) harmful stimuli. Nociceptors are nerve cells that are often sensitive to a range of noxious stimuli (chemical, heat, pressure, etc). Nociceptive signals are sent from these peripheral nerve cells to the spinal cord, into (in mammals) subcortical and cortical brain areas. Nociception can trigger behavioural responses like the withdrawal reflex, but (as outlined below) more complex responses too². In organisms capable of consciousness, nociceptive activity above particular thresholds often causes conscious pain experiences (Figure 1b). Plausibly, these pain experiences generate a different suite of behavioural responses to nociception alone (e.g. typically more long-term and flexible behaviours). There are of course exceptions to this. Some pain experiences are not caused by nociception (e.g. in chronic pain conditions), and some nociceptive activity does not cause pain experiences (e.g. if it is strongly modulated top-down). However, in the fairly simple experimental paradigms used in this literature, it is, I think, safe to assume that this captures the most relevant parts of the causal structure.

² I describe here how nociception is understood to work in mammals. Nociception may be realized differently across genera, or potentially in artificial systems.

Figure 1



Hopefully it is obvious that the way to tell whether an organism is capable of having pain experiences is to see if one can elicit behavioural responses that are unique to pain experiences. The easiest way to do this is to experimentally intervene directly on nociceptive processing, for example by exposing an organism to noxious stimuli which trigger nociception. If these interventions generate behavioural responses unique to pain experiences, then pain experiences are present, and the causal structure is as in Figure 1b). If only nociceptive responses are generated, then pain experiences are (probably) not, and the causal structure is as in Figure 1a). The obvious difficulty is in identifying which responses are purely nociceptive, and which are related to pain. There are broadly speaking two ways to do this.

One way to select behavioural indicators of pain experiences is to look at what markers of pain are used in e.g. mammals and birds, where we are fairly confident that pain experiences are present, and to transfer these across to more controversial genera. If a marker provides evidence of pain in mammals, then finding it in these genera should do too.

Something like this reasoning seems to inform the inclusion of a range of the criteria used in this literature (see e.g. introduction of Ashley et al. 2009).

On the face of it, this sounds safe, but there is a significant hazard here. This is that markers of pain in mammals are not widely used to show that mammals are *capable* of having pain experiences. It is fairly widely accepted that they are. Instead, markers of pain experiences are often used in experimental settings to track the presence or absence of pain, or track levels of pain, for example to investigate responses to particular kinds of noxious stimuli or injury, or to monitor animal welfare. That is, it is assumed *that* mammals are capable of pain experiences, and what is needed is a way of tracking *when* they are in pain. In mammals though, tracking purely nociceptive responses might be a perfectly reliable way of doing this. To explain: if one can assume that there is a strong correlation between levels of nociceptive activity and levels of pain (which might be a safe assumption for certain experimental contexts in certain species) then tracking nociceptive responses is a reliable way of tracking pain experiences.

This can be illustrated through the rat grimace scale (Sotocina et al., 2011; Waite et al., 2015). This is now fairly routinely used to track pain experiences in laboratory rats and uses changes in the shape of eyes, whiskers, nose and cheeks, among others, to score the intensity of pain experiences. This rating scale was based on the use of similar scales to score pain experiences in humans unable to verbally report (Lints-Martindale et al., 2007) and on the assumption that facial expressions related to emotional states are widely conserved across mammals. It is also based on the assumption that rats feel pain. While changes in whisker and eye shape are likely reliable ways of tracking the intensity of pain

experiences, by themselves are very poor evidence that rats are capable of pain experiences, as they are generated in a reflex-like manner. Reflexive responses are generally associated with nociceptive and so are deemed to be poor indicators of the capacity for pain. In this case, attempting to use a marker like this in controversial species to show that they are capable of pain would be to misuse it. One has to be fairly careful therefore in simply transferring behavioural indicators across from mammals: while they might very reliably signal the presence of pain experiences in mammals, they are often not valid ways of showing that an animal is capable of pain experiences at all.

A more direct way of identifying appropriate behavioural markers of pain is to therefore do some theoretical work, and as above, identify behavioural responses that are likely to be unique to pain experiences. This is hard: there is no general agreement on a fine-grained functional description of pain experiences. This is hardly surprising given that there is no general agreement on a fine-grained functional description of consciousness in general. Similarly, there is not, as far as I know, a widely agreed description of the limits of nociception: this too is still work in progress. Nevertheless, researchers working on animal pain often use a theoretical framework that they argue has the power to identify valid markers of pain experiences, discussed below.

3: A definition, two principles and a set of criteria

In this paper I treat Sneddon et al. (2014) as presenting a fairly representative view of the general principles used in research on pain experiences in controversial genera (i.e. anything outside mammals and birds). Here, the central definition of pain, based on its likely

function, is (from Sneddon, 2009) to allow animals to: “quickly learn to avoid the noxious stimulus and demonstrate sustained changes in behaviour that have a protective function to reduce further injury and pain, prevent the injury from recurring, and promote healing and recovery” (p. 338). From this, two principles are used to generate a set of criteria. One principle is that there is a ‘whole animal’ physiological or behavioural response to potentially painful stimulation. This is to rule out local bodily responses generated by nociceptive processing only (like withdrawal reflexes of individual limbs), and includes both short and long term responses. The second key principle is that there is a change in motivation in the animal. This includes immediate changes in the things that the animal is motivated to do, as well as longer-lasting changes in behaviour driven by learning. These principles are then used to identify more specific criteria that can be used in experimental designs to try to identify pain experiences.

The assumed function of pain that this definition and these principles are based on is that it enables individuals to avoid similar painful stimuli and to promote healing in the long-term. In contrast, nociception is typically presented as having only local and short-term behavioural effects such as withdrawal reflexes. However, nociception also drives non-local and longer-term behaviours, and can be used for simple forms of avoidance learning. In *C. elegans*, nociception causes a stereotypical withdrawal reflex away from noxious stimuli, but also simple whole body movements locomoting forwards, again away from noxious stimuli (Mohammadi et al., 2013). Nociception can also generate longer-term behaviours. When stimulated multiple times, the sensitivity of nociceptors increases, such that lower levels of stimulation are subsequently required to activate them and drive nociceptive responses. Of more interest, nociceptors also increase their sensitivity at sites of injury and

inflammation. This means that at injured sites, stimulation that would not usually cause withdrawal reflexes and other nociception responses now does: in effect this helps the organism to avoid further damaging an injured/inflamed bodily site. In addition, nociceptive processing fed to the spinal cord only can support simple but efficient forms of avoidance learning in rats (Grau et al., 2006), which enables learning to avoid noxious stimulation. In this case, on at least an uncharitable reading of the definition of pain give above, mere nociception fits the bill. Nociception alone can support forms of avoidance learning that enable organisms to avoid similar painful stimuli, and can drive behaviours that promote healing and recovery.

This is a fairly uncharitable reading of the definition of pain though: in using the definition one is supposed to rule out behaviours that are likely just simple nociceptive reflexes. However, as raised by Rose et al. (2014), it is not always entirely clear what these include. Presumably the forwards locomotion of *C. elegans* away from noxious stimuli does not count as a 'simple' reflex: it is a whole animal response, which requires a fair degree of motor co-ordination. Yet it is (plausibly) a nociceptive reflex, but an example of a complex one. Some reflexes require more co-ordination of muscle groups than others: the Galant reflex in human newborns involves the whole trunk curving sideways in response to stimulation next to the spine. In this case we have to say rather more about the degree of complexity we should expect of nociceptive vs. pain-based responses. And indeed, the way that pain research is actually carried out, in terms of the experimental paradigms used and the variables measured, is often far richer than the definition and criteria suggest. This richness relies on implicit assumptions about what the relevant forms of complexity are. As recently argued by Birch (2020), one cannot make claims about animal sentience with no

theoretical commitments whatever.

A common theme of these assumptions is that responses to painful experiences should exhibit features of 'higher level' cognition: decision making that is based on a range of integrated information, and flexible and long-lasting forms of learning. Unsurprisingly, these are taken to be markers of consciousness more generally (e.g. see discussion in Godfrey-Smith, 2017). In what follows I claim that these capacities are key in evaluating the validity of behavioural indicators of pain, and that putting these front and centre makes it possible to identify a much more robust set of criteria, and related behavioural indicators, for identifying pain experiences. While these criteria are unlikely to convince die-hard sceptics, they are better theoretically motivated than existing ones, and rely on a deeper analysis of the interactions between nociception, cognitive capacities, and consciousness.

To do this I analyse the list of criteria related to the principles from Sneddon et al. (2014) listed in Table 1, drawing on a simple interventionist approach to experimentation and the contrast case of *C. elegans*. After a brief explanation of each criteria, the analyses explore whether pain experiences are reliably targeted by the criteria (or whether the criteria more likely target nociception), and whether *C. elegans*, assumed to be incapable of pain experiences, nevertheless meets the criteria of pain experiences.

It is also assumed throughout, with Sneddon et al. (2014), that convincing evidence of pain experiences requires multiple indicators, rather than a single measure. As Sneddon et al. explain, "together [these indicators] represent an increasing level of complexity of responses to pain that go beyond simple and acute detection and reflex responses and

begin to demonstrate a level of behavioural complexity that would require some form of experience” (p. 209, *op cit*). This fits very well with the idea that the (often implicit) theoretical assumptions that motivate particular criteria do indeed track the idea that it is ‘behavioural complexity’ that is key to showing that an organism is capable of consciousness, and pain experiences in particular.

Principle 1: Whole-animal responses to potentially painful events differ from innocuous stimulation	Principle 2: Change in motivational behaviours after a potentially painful event
<p>Criteria:</p> <ol style="list-style-type: none"> 1. Possession of nociceptors, evidence of central processing of nociception involving areas that regulate motivated behaviour (including learning and fear) 2. Nociceptive action responsive to endogenous modulators 3. Nociception activates physiological responses linked to stress (e.g. changes in respiration, heart rate or hormonal levels) 4. Evidence that responses are not just a nociceptive withdrawal reflex 5. Long-term avoidance behaviour 6. Protective behaviours such as wound guarding, limping, rubbing, licking or excessive grooming 7. All of the above reduced by analgesia or local anaesthetics 	<p>Criteria:</p> <ol style="list-style-type: none"> 8. Self-administration of analgesia 9. Pay a cost to access analgesia 10. Pay selective attention to response to the noxious stimulus over other normally salient stimuli; animal does not respond appropriately to competing events 11. Conditioned place avoidance and long-lasting avoidance learning 12. Relief learning 13. Trade-offs between avoiding noxious stimuli and other motivational requirements 14. Evidence of paying a cost to avoid the noxious stimulus

Table 1. This is adapted (to get permission!) from Sneddon et al. (2014), Table 1.

4. The validity of behavioural indicators of pain

4.1 Criteria related to first key principle: Whole animal responses that differ from innocuous stimulation

As the first principle suggests, these criteria are related to showing that the animal does not only show local bodily nociceptive responses (e.g. limb-specific withdrawal reflexes), but that noxious stimulation has physical and behavioural effects on the whole animal. These are therefore not all purely behavioural indicators, though most of them are assessed via observed changes in behaviour.

The first criterion related to this principle (Criterion 1 from Table 1 above) is that nociceptive signals are centrally processed, in areas that regulate learning and fear. In some ways this seems to presuppose that the animal is conscious: presumably fear is a conscious experience like pain, and centralized processing is often associated with the capacity for consciousness. If this is right, then this criterion assumes exactly what is under debate. On another interpretation however, the criterion just demands that nociceptive processing cannot be purely local within the central nervous system, and must be able to support some forms of learning. This though does not provide a valid indicator of the presence of pain experiences. As reviewed throughout this paper, nociceptive processing in *C. elegans* occurs 'centrally', and supports learning. However, by assumption, *C. elegans* is not capable of pain experiences. So, while it is probably necessary that nociceptive signals are transmitted to 'central' areas of the central nervous system, it is not itself a valid indicator.

Criterion 2, that nociceptive processing is sensitive to endogenous modulators, is an interesting one. Here the idea is that nociceptive processing can be affected by neurotransmitters produced within the central nervous system, which in turn affects behavioural responses. Opioids are generally a reliable way to modulate nociceptive processing across animal taxa, and morphine has behavioural effects on *C. elegans* (Nieto-Fernandez et al., 2009). Nematodes administered with morphine show less avoidance behaviour than controls towards areas of noxiously high temperatures. As Mills et al. (2016) show, nociceptive processing in *C. elegans* is also modulated by endogenous (internally generated) opiates. That is, nociception in *C. elegans* satisfies this criterion. The widespread presence of endogenous antinociceptive systems across taxa is discussed in more depth later in this section. For now though, the simple contrast case provided by *C. elegans* shows that this criterion does not present a valid way of identifying pain experiences.

The same is true for Criterion 3, that nociception activates physiological responses related to stress, for example including changes to heart rate, respiration or hormone levels. While *C. elegans* does not have a specific respiratory or circulatory system, it does show a range of physiological stress responses to noxious stimuli, including noxiously high temperatures and toxic heavy metals. These responses include changes to signaling systems essential for food storage and growth, and may involve hormonal signaling via sterols (Gerisch et al., 2001; Rodriguez et al., 2013; Zhou et al., 2011). So, while *C. elegans* has rather different physiological stress responses compared to mammals, they are activated by nociception.

Evidence that the behaviour is not due to nociception alone (Criterion 4) is a prerequisite for

any valid indicator, and does not suggest a specific variable to be measured.

Things are a little more promising for long-term avoidance behaviour (Criterion 5). Avoiding noxious stimuli, particularly in the long term, seems indicative of learning and flexible behaviour. However, the behaviour of *C. elegans* again provides a useful contrast case.

Amano and Maruyama (2011) showed that these nematodes also do avoidance learning.

Here, nematodes learned to avoid an olfactory cue that had been paired with noxious acidic environments. These learned associations can drive avoidance behaviour for up to 24 hours: this is fairly impressive for an organism that typically lives for 2-3 weeks (see also Sasakura & Mori, 2013, for a review of learning and memory in *C. elegans*). Long-term avoidance learning in response to nociceptive stimuli therefore occurs in *C. elegans*, plausibly without pain experiences.

However, avoidance behaviour that displays more flexibility in learning and motor responses may be more likely to index pain experiences. For example, Dunlop et al. (2006) showed that goldfish learned different strengths of avoidance behaviour to different levels of electric shocks that were given in different spatial areas of a tank, and retained this information for at least 24 hours. It is important to identify exactly what about this is distinctive. As emphasized in Dunlop et al. (2006), goldfish show evidence of *spatial* memory in response to nociceptive stimuli. Goldfish learn to avoid specific locations that they associate with aversive stimuli. In the example above *C. elegans* does not do this; it instead simply moves from higher to lower concentrations of the olfactory cue, which does not require having any spatial representation of the environment.

Of relevance here, Barron and Klein (2016) have argued that having an egocentric spatial model of the environment is key to having a capacity for consciousness. They follow (Merker, 2005, 2007) in arguing that consciousness is supported by an internal integrated model of the organism, their needs, and the external environment. This system regulates decision making in relation to the priority of different needs, and the lay-out and resources available in the external environment. If this is a plausible model of consciousness, then this provides the reason for thinking that goldfish are conscious, and can have pain experiences, but *C. elegans* does not. Goldfish are capable of spatial memory and learning, an indicator of consciousness in general, while *C. elegans* does not. In this case a better criterion for identifying pain experiences is that an organism shows that they are capable of learning about noxious stimuli in a way that reflects features that are general indicators of consciousness, for example via spatial learning and memory.

Next up is Criterion 6: that the organism exhibits protective behaviour, including “wound guarding, limping, rubbing or licking” (see Table 1, Sneddon et al. 2014). This again looks promising, but needs unpacking. First is the observation made in Rose et al. (2014) that some of these behaviours are not clearly protective: a trout whose lips have been injected with acid is not protecting an injury by rubbing their lips on the side of a tank. It could be an attempt to dull pain sensations by ‘confusing’ the receptors in the injured area, or it could be a reflex-like response to rub off a noxious substance: it is not clear how this behaviour should be interpreted. Wound guarding and limping, or generally not using an affected body part, seems more like it. These behaviours suggest that the organism’s behaviour is being driven by ongoing pain sensations. However, again, there are other possible plausible explanations. As reviewed above, one of the properties of nociceptors is that with repeated

or prolonged stimulation, or at sites of injury or inflammation, their sensitivity dramatically increases. Nociceptive responses (e.g. withdrawal behaviours) are now produced for very low intensity stimulation of injured areas. In organisms capable of simple avoidance learning, this also means that nociception can itself drive organisms to 'protect' injured areas, potentially by learning to hold body parts in particular positions to avoid further stimulation, or avoiding using injured body parts for moving or grooming.

This criterion is therefore, I think, particularly hard to evaluate. Presumably there is some difference between the guarding behaviours related to purely nociceptive processing and those related to pain experiences, but pulling these apart is not straightforward. To do this, referring to more general cognitive abilities associated with consciousness seems like a possible move. Protective behaviour based on more complex forms of learning, prediction or planning about what sorts of stimuli or activities are likely to trigger strong nociceptive responses, and how to avoid them, seem more likely to indicate the presence of pain experiences than more local protective responses.

The last criterion on the list of whole animal responses to noxious stimuli (Criterion 7) is that all the above are reduced by administered analgesia and local anaesthetics. Here I focus on analgesia. Morphine and other opiate-based analgesics act on nociceptive pathways, so if these pathways feed into mechanisms of pain experiences, then they also act on pain. However, where nociceptive pathways do *not* feed into mechanisms of pain experiences, one would still expect morphine to affect purely nociceptive responses, as they do in *C. elegans*. For sensitivity to these kind of analgesics to be a valid indicator of pain experiences, one therefore needs to show that these analgesics affect behaviours that really are valid

indicators of pain experiences.

However, the structure of this criterion leads to a problem. The idea is that if one has found valid indicators of pain experiences, the behaviours associated with these criteria (i.e. the behaviours that organisms exhibit when they have pain experiences) should be reduced when the organism is given analgesics. The problem is that put like this, it is not very clear what behavioural sensitivity to administered analgesics adds in evidential terms. For this criterion to work, it requires that researchers have *already* identified a set of valid behavioural indicators of pain experiences.

There is another twist. This is that if an organism is capable of having pain experiences, but its pain responses were *not* sensitive to (standard) analgesics, this would be very surprising indeed. Analgesics typically target nociceptive pathways, and there is typically a strong association between the modulation of nociceptive processing and the modulation of pain experiences and associated pain responses. So, where an analgesic successfully modulates nociceptive processing, it would typically also modulate pain experiences. The twist is that organisms with nociceptive systems tend to have (as far as I can tell, *always* have) in-built antinociceptive systems. Endogenous antinociceptive systems inhibit or otherwise disrupt nociceptive processing when it is profitable to the organism to do so: they produce what are effectively analgesics. For example, in *C. elegans*, an antinociceptive opioid system modulates nociception when there is both noxious stimuli and food present: by inhibiting nociceptive processing, standard nociceptive responses (like withdrawal reflexes) are triggered less often, making the organism able to access food while noxious stimuli are present (Mills et al., 2016).

Most (plausibly all) organisms capable of pain experiences are capable of producing effective analgesics through antinociception. Where pain experiences are present then, their being sensitive to at least *some* analgesics is incredibly likely, given the background biology of antinociception. It follows from this that showing that pain responses are sensitive to analgesics is, yet again, of little evidential value: indeed this is entirely to be expected. So, while, on the face of it this looks like a fairly persuasive criterion for identifying the capacity for pain experiences, behavioural sensitivity to administered analgesics fails to do much work by itself.

In summary, many of the criteria associated with the first principle fail to be valid indicators of pain experiences, at least in creatures whose capacity for pain experiences is in doubt.

The criterion of central processing of nociception, relating to fear and learning (Criterion 1), either assumes too much (that the animal is indeed conscious, and capable of having valenced conscious states), or interpreted in a less demanding way, is not a valid indicator.

Nociception is processed 'centrally' in *C. elegans*, and supports avoidance learning.

Nociception in *C. elegans* is also sensitive to endogenous modulators, and activates physiological stress responses (Criteria 2 and 3). Criterion 4 is a prerequisite for *any* valid indicator, and does not suggest a specific variable to be measured. *C. elegans* also shows long-term avoidance behaviour (Criteria 5). However, avoidance behaviour in other genera may be tied to abilities related to consciousness in general, like spatial learning, or perhaps more complex forms of associative learning. Protective behaviours (Criterion 6) are, I think, hard to make a call on: some are not very obviously related to pain experiences and could be generated by nociception alone, but some types of protective behaviours, again related

to capacities related to consciousness in general, may be valid indicators. Behavioural sensitivity to analgesia (Criterion 7) was argued to be of little evidential value: this is almost guaranteed in animals that exhibit valid behavioural indicators for pain, on basic biological grounds.

4.2 Criteria related to the second key principle: Animal shows a change in motivation

The second principle expresses the idea that pain plays a role in learning and both short and long-term decision making relating to noxious stimuli. As such, some of the features related to this principle have already been covered in the discussion above. These include (from Criterion 11) that animals are capable of conditioned place avoidance (avoiding locations associated with noxious stimuli: possibly a valid indicator) and long-lasting avoidance learning more generally (not a valid indicator).

The first criterion to be considered for this second key principle is self-administration of analgesia (Criterion 8 from Table 1). This criterion (and others like it below) relates to behaviour in situations where analgesics are available to the animal, such that they can choose to access them, but are not directly administered. That is, it relates to learning and decision making. In species that continue to eat and drink while in potentially painful situations (e.g. chickens) there is evidence that following 'conditioning' for periods of being fed either only regular food or only food containing analgesics (these are colour coded), injured animals subsequently prefer foods that contain analgesics (Danbury et al., 2000). This is deemed to be evidence of the presence of pain experiences: animals undergoing continual pain experiences regulate their behaviour so that they can access pain relief, via

previously learned associations between specific types of food, locations or other cues, and pain relief. This needs some unpacking.

First, it is somewhat ambiguous in the original article (Danbury et al., 2000) whether the aim is to show that chickens are capable of having pain experiences at all, or whether, given that they can, that lameness in particular (common in chicken farming) causes pain and so demands action in animal welfare terms. This is important. As reviewed above, if the study is based on the premise that chickens are capable of pain experiences, then behaviours driven by nociceptive processing alone could be used as reliable indicators of pain. This is based on the further assumption that particular levels of nociceptive processing usually also cause pain experiences. In this case, transferring the same experimental paradigms to other genera requires further argument to show that the behavioural responses really are unique to pain experiences.

This argument might work as follows. Through learning these animals associate one colour coded food with lower levels of felt pain, and so they prefer eating it over food that contains no analgesics. Food preferences are the result of reward learning: injured animals have a learned association between analgesic laden food and a reward (less pain), and this guides their eating behaviour. The question is whether one should expect reward learning of this kind to work only where pain experiences are present, or whether similar reward learning would be expected where only nociception is present. For example, a downwards modulation of nociceptive processing may count as a reward in the central nervous system, whatever its source, and whether conscious experiences are present or not. And as above, sensitivity to analgesics, and the ability to do avoidance learning concerning nociceptive

activity, do not in themselves reliably indicate the presence of pain experiences. In this case, and in the absence of any straightforwardly parallel studies in *C. elegans* (as far as I know), claiming that this is a valid indicator of pain experiences is unclear, and needs more theoretical defense.

Next to be considered here, Sneddon et al. (2014) suggest that relief learning (Criterion 12) is a valid indicator of pain experiences. In relief learning organisms learn to prefer stimuli associated with the sudden absence of a (previously present) potentially painful stimuli. One interpretation of this is that it may signify a preference for 'relief' from the painful stimulus. Sneddon et al. (2014) argues that as relief learning requires "considerably higher processing" than reflex actions or nociception, it may signify the presence of pain experiences (p. 205).

Whether relief learning requires "considerably higher processing" than simple, non-conscious forms of learning is in fact not entirely clear. Relief learning is a form of backwards conditioning on noxious stimuli. In more standard forward conditioning paradigms, the conditioned stimulus (CS) (e.g. an odor) is presented before the unconditioned stimulus (US) (e.g. electric shock), and organisms learn to avoid the CS via its association with the US. In backward conditioning, the order is reversed: here the US (e.g. shock) is presented before the CS (e.g. odor). In relief learning in particular, organisms learn to now approach the CS.

Backwards conditioning is however somewhat messy: in a range of organisms, backwards conditioning sometimes leads to 'relief' learning, where the organism learns to approach

the CS, but it also sometimes results in the same response as forward conditioning, where the organism learns to avoid the CS (e.g. see reviews across genera in Andreatta et al., 2010; Felsenberg et al., 2014; Gerber et al., 2014; Green et al., 2019). Which response is more prevalent depends on a range of experimental factors. There has, to my knowledge, been little work on backwards conditioning in *C. elegans*, though work so far has not found evidence of relief learning (Amano & Maruyama, 2011).

On the one hand then, relief learning may just be a form of simple associative learning, different in some ways to regular forward conditioning, but perhaps not substantially so. *C. elegans* is perfectly capable of forward conditioning with aversive stimuli, so perhaps more experimentation would find evidence of relief learning in this organism too. In this case, relief learning is unlikely to be a valid indicator of pain experiences.

On the other hand, relief learning may be a form of associative learning that is only possible in creatures that are capable of consciousness. Trace conditioning, for example, where animals learn to associate two stimuli presented across a time interval, has been argued to be a form of learning that is unique to conscious organisms (Birch, 2020). Trace conditioning requires some sustained 'memory' of the first stimulus, perhaps indicative of (conscious) working memory. Similarly, relief learning is perhaps conceptually more complex than standard forward conditioning. Instead of associating an aversive stimulus with the presence of a neutral cue, here the association is between a cue and the *offset* of an aversive stimulus. As such, this may require some sort of tracking of temporal relationships that goes beyond that required in forward conditioning. If the ability to track and predict features of the external environment based on distinctly spatial and temporal features is

key to consciousness in general, then perhaps relief learning can only be done in animals that are capable of consciousness. In this case of course, relief learning would be a valid indicator of pain experiences, through its relation to more general features of consciousness.

Relief learning is still a fairly new research area, so more information on the neural mechanisms that supports it would be helpful in determining which option is more likely. One interesting line of evidence from neural and molecular research on relief learning in fruit flies (as reviewed in Gerber et al., 2014) suggests that the mechanisms regulating forward conditioning and relief learning are not radically different: they may use the same network of cells, but be the product of different dynamics of molecular signaling. Whether this difference is likely to track other differences between conscious and non-conscious organisms is an open question (though, on the face of it, seems unlikely). In this case, more empirical work on relief learning, and stronger theoretical arguments linking relief learning to consciousness are required before it can be seen as a valid indicator of pain experiences.

A criterion of pain experience that is often recognized to be of obvious validity is that animals show evidence of trade-offs in motivational requirements surrounding aversive stimuli. For brevity, I here include the related criteria of paying a cost to avoid noxious stimuli and paying a cost to access analgesia: all of these are essentially about trade-offs (this combines Criteria 9, 13 and 14). Examples of these trade-offs include trout trading between being near conspecifics and avoiding electric shocks (Dunlop et al., 2006), goldfish trading between accessing food and avoiding electric shocks (Millsopp & Laming, 2008), zebrafish trading off being in a barren (aversive) tank against accessing analgesics present in

the water (Sneddon, 2013), and hermit crabs trading off electric shocks in accordance with the 'value' of their shell (Appel & Elwood, 2009; Elwood & Appel, 2009).

Again, *C. elegans* is a party pooper, at least on straightforward readings of this criterion. The nematode shows trade off behaviour between the 'bad' of crossing a strip of aversive copper ions to the 'good' of approaching an attractive odour, depending on the concentration of both the aversive and attractive stimuli (Shinkai et al., 2011; see also Mills et al., 2016 reviewed above on antinociception). The nematode does this using very simple mechanisms of sensory integration across a handful of neurons. In a sense, this is hardly surprising behaviour: any organism that has different needs is going to need some system for deciding what to prioritise. So, it cannot be the mere existence of trade-offs in needs that does the work: the trade-offs need to be of a certain kind. Again, the general properties associated with consciousness are key: that decision making is based on a range of integrated information and that learning is flexible and long-lasting. These properties are not found to a great degree in the behaviour of *C. elegans*, but they might be in some of the other examples mentioned above.

Take the first case above. Here, trout were subject to the same experimental set up as the goldfish described in Section 4.1: electric shocks were delivered in particular spatial areas of a tank, which the fish learned to avoid (though the trout did not hold onto this information for long) (Dunlop et al., 2006). When presented with a conspecific next to the areas where shocks were present, trout (a shoaling fish) often remained in these areas, and so endured electric shocks while choosing to be close to a conspecific. As argued in the paper, this shows that pain avoidance in these species "is not purely a reflex action...If a fish was willing

to change this supposedly inelastic, stereotypical response to a noxious stimulus, it is highly probable that pain perception can occur in fish” (p. 269, *op cit*). Using a similar paradigm, Millsopp and Laming (2008) showed that goldfish trade off between the motivation to avoid different intensities of electric shocks and the motivation to eat. These authors are rather less convinced that this provides evidence of fish pain experiences, but note that “If a fish is willing to change this reflex response to a noxious stimulus, as shown here, it is possible that there is some sort of conscious decision making taking place” (p. 253).

One key assumption here is that actions that are not ‘pure’ or inelastic reflexes provide evidence of pain experiences, because elasticity in decision making requires conscious decision making. Rose and colleagues have heavily criticised this assumption over a number of publications (Rose, 2002, 2007; Rose et al., 2014) on the grounds that elastic behaviours can be performed (and often are) without awareness. The *C. elegans* example also shows that the assumption is straightforwardly false. Very simple neural mechanisms are capable of guiding actions in an elastic way in response to different concentrations of aversive and attractive stimuli. This is parallel to different intensities of electric shock and different levels of hunger in goldfish.

This is where the general properties associated with consciousness come back in.

Pretheoretically, it seems plausible to assume that some kinds of decision making can only be done if the neural machinery of consciousness is in place, for example when it involves a reasonable degree of integration and comparison of multiple sources of information.

Clearly, *C. elegans* can do decision making that requires some integration and comparison, but it can’t do much. Identifying if there are qualitative differences between the abilities of

C. elegans and goldfish in this regard would be telling. One possibility comes back to the ability of goldfish (and a lesser degree trout), but not *C. elegans*, to direct trade-off behaviour informed by spatial learning and memory. In this case, it is not the presence of trade-off behaviour *per se* that does the work, it is the evidence of having an integrated and spatial internal model of the environment, 'tagged' with the locations of aversive and attractive stimuli, that guides behaviour.

There is more to say too about the case of trade-off behaviour in hermit crabs. In Elwood and Appel (2009) hermit crabs given electric shocks via their shells were more likely to move into a new shell when one was presented 20s later, and those in low-value shells were particularly likely to move. Appel and Elwood (2009) found that shocked crabs were more likely to move into a new shell up to 1 day after the shocks. The authors discuss this in terms of forming a memory of the painful event that informed later behaviour. In many ways though, this behaviour is very similar to simple but long-lasting associative avoidance learning and trade-off behaviour found in *C. elegans*. The difference here is that the aversive and attractive stimuli (noxious stimulation and high quality shelter) are linked to specific shells. Given that having a high quality shell is clearly a significant motivator for hermit crabs in general (as reviewed in both papers), this behaviour may not be so surprising. In addition, the 'memory' of the painful event may simply consist of unconscious associative memory, rather than anything episodic, as is somewhat implied in the author's original discussion.

Later though, Magee and Elwood (2016) suggested that these findings may be an experimental artifact from where the electrical wires were positioned in the high and low value shells (the shocks in 'low value' shells may have effectively been higher than in 'high

value' shells). They developed an alternative paradigm that also shows trade-off behaviour, now between electric shock and predator avoidance. Here, hermit crabs were housed in the same shell types, and shocked when a variety of odours were present in the tank. Hermit crabs shocked when odours from predators were present were less likely than others to exit their shells. Again, this trade-off may or may not reflect the presence of pain experiences: it may be a simpler non-conscious decision based on the relative degrees of aversion to these two stimuli.

Instead, other behaviours discussed in these papers seems (to me) far more convincing of consciousness and the capacity for pain experiences in hermit crabs. When choosing new shells, hermit crabs perform complex and multi-sensory investigations of shells to identify one of adequate size, shape, strength and weight, and that is better than their current shell. They combine visual information with information gained from investigation of the external surfaces and inside of the shell with their legs and chelipeds (claws), and retain this information for up to 40 minutes. This behaviour displays far more complex abilities relating to integration and comparison of information than in the mere existence of trade-off behaviour between two aversive states. If the findings of Appelwood and Elwood (2009) are not in fact experimental artefacts, and if trade-off behaviour relating to shell choice is based these integrative decision making abilities, it is these abilities that makes it plausible that hermit crabs have pain experiences.

Further, Magee and Elwood (2016) report on other interesting effects. Hermit crabs shocked in high intensity odours of mussels (a prey food) are less likely to exit their shells, perhaps (according to the authors) because areas where mussels are concentrated are more

likely to have predators. Some hermit crabs exited their shell and then proceeded to investigate the inside of it. The crabs that exited their shells and moved far away were heavier than other crabs: the authors suggest that heavier crabs are more able to forcibly take shells from others, so exiting their shells entirely may be less risky. This again suggests rather more complex decision making than in the simple trade-off case: hermit crabs here may not merely be trading off between two valenced states, but a much broader range of states and factors. These potentially include various aspects of shell quality, predator threat assessed from a variety of stimuli, own size and general health, and perhaps more. This is a rather more complex matrix of information that needs to be integrated and compared, and plausibly goes well beyond what can be achieved by the machinery that *C. elegans* relies on. It is this ability to do complex decision making, based on a wide range multi-modal information, that plausibly does the work of showing that hermit crabs are conscious and capable of pain experiences, rather than the existence of simple trade-offs in relation to aversive stimuli.

The final criterion to be examined here is that organisms exposed to noxious stimuli selectively attend to their 'painful' state, rather than any other usually salient stimuli (Criterion 10). Here, I assume that paying attention to the wound site via protective behaviours (grooming etc) is covered by that criterion as discussed above. This criterion is more related to whether, when injured, organisms are less responsive to external stimuli that usually elicit specific responses (e.g. to new objects, predators). For example, Ashley et al. (2009) showed that rainbow trout treated with noxious stimuli (acid injected into lips) show a large drop in antipredator responses compared to control groups. This suggests that

nociceptive stimulation changes the attentional focus of the injured organism, indicating that they have pain experiences.

There is a complication here though. There is evidence from many genera, including in bony fish, that predator threats (among other cues) stimulate antinociceptive responses in the brainstem (as reviewed in Alves et al., 2013). As described earlier, these responses effectively block nociceptive processing and responses, and so allow other context-appropriate behaviours to be produced, such as anti-predator responses. In this case, one would *not* usually expect a huge drop in antipredator responses in injured fish:

antinociception works exactly to preserve antipredator behaviour. In this case, it is not entirely clear what was found in (Ashley et al., 2009). Perhaps these fish were noxiously stimulated to such an extent that standard antinociception responses were not triggered, or only had a minimal effect on behaviour, but this is not explicitly claimed in the paper.

Further, this shows that the criterion needs modification: perhaps one should not expect significant changes in behaviour regarding highly salient stimuli in animals given noxious stimulation, but only changes regarding less salient stimuli that less directly related to survival, and so where it is less likely that there are reliable and specific antinociceptive responses.

It is still though slightly unclear exactly what this criterion is getting at. One possibility is that this criterion is arguably another version of the criterion relating to trade-offs: animals trade off between responding to salient stimuli in typical ways (often related to survival), and avoiding the possibility of further noxious stimulation or exacerbating existing injuries. As above, this trade off is likely related to the degree of existing injury or level of nociceptive

activity. In this case though, the same comments about trade-offs can be made here. More persuasive evidence of pain experiences would need to show that trade-offs are made on the basis on the integration and comparison of multiple factors.

In sum, many of the criteria relating to the second principle do not provide obviously valid indicators of pain experiences either (again, at least in creatures whose capacity for pain is in doubt). Criterion 11 was covered in the discussion in Section 4.1. The ability to self-administer analgesia based on reward learning needs more theoretical defence, as does relief learning (Criteria 8 and 12). The remaining criteria (Criteria 9, 10, 13, and 14) have been argued to be variations on the general theme of being able to make trade-offs between different motivational states, sometimes relying on prior associative learning. Trade-offs relating to noxious stimulation are shown in *C. elegans*, as is associative learning. Instead, more complex trade-offs in decision making, based on the integration and comparison of multiple factors and stimuli are better able to differentiate between behaviours based on nociception and pain experiences. Again, these bring us back to abilities related to consciousness in general, rather than any specific to pain.

5. Conclusion

I have used a very basic interventionist approach to experimentation and the contrast case of *C. elegans* to show that many of the behavioural indicators used to identify pain experiences in controversial genera are unlikely to be valid in their current form. A summary of the analysis of these is in Table 2 below. As suggested throughout, the idea here is not to radically undermine claims about pain experiences in the more controversial genera. The

aim was instead to describe, where possible, behavioural indicators that are more likely to be robust, using the general strategy of allying them to general features of learning and decision making associated with consciousness. Indeed, a fair amount of the current literature can I think be reinterpreted according to more valid indicators of pain experiences, and still support the claim that these genera are capable of pain experiences.

Principle: Whole-animal responses to potentially painful events differ from innocuous stimulation	Valid indicator?	Principle: Change in motivational behaviours after a potentially painful event	Valid indicator?
<p>Criteria:</p> <p>1. Possession of nociceptors, evidence of central processing of nociception involving areas that regulate motivated behaviour (including learning and fear)</p> <p>2. Nociceptive action responsive to endogenous modulators</p> <p>3. Nociception activates physiological responses linked to stress (e.g. changes in respiration, heart rate or hormonal levels)</p> <p>4. Evidence that responses are not just a nociceptive withdrawal reflex</p> <p>5. Long-term avoidance behaviour</p> <p>6. Protective behaviours such as wound guarding, limping, rubbing, licking or excessive grooming</p>	<p>No</p> <p>No</p> <p>No</p> <p>N/A</p> <p>Depends</p> <p>Unclear</p>	<p>Criteria:</p> <p>8. Self-administration of analgesia</p> <p>9. Pay a cost to access analgesia</p> <p>10. Pay selective attention to response to the noxious stimulus over other normally salient stimuli; animal does not respond appropriately to competing events</p> <p>11. Conditioned place avoidance</p> <p>12. Relief learning</p> <p>13. Trade-offs between avoiding noxious stimuli and other motivational requirements</p> <p>14. Evidence of paying a cost to avoid the noxious stimulus</p>	<p>Unclear</p> <p>Depends</p> <p>Depends</p> <p>Yes</p> <p>Unclear</p> <p>Depends</p> <p>Depends</p>

7. All of the above reduced by analgesia or local anaesthetics	No additional value		
--	---------------------	--	--

Table 2. ‘Unclear’ refers to the need for further empirical work and theoretical defense of this criterion. ‘Depends’ refers to whether or not the criterion refers to general features of learning and decision making associated with consciousness. Where it does, it is likely to be valid indicator. Where it does not, it is unlikely to be valid.

This analysis also sheds some light on the strategy used in animal pain research that often seems to trade on the idea that indicators, evidence and conclusions about pain experiences can be somehow made independently of indicators, evidence and conclusions about consciousness more generally (the words ‘conscious’ and ‘consciousness’ are not mentioned anywhere in Sneddon et al. 2014). This contrasts with views from critics who sometimes claim that in order to show that an organism is capable of pain experiences, one must ‘first’ show that they are conscious at all. For example Rose et al. (2014) claim that: “It is considered axiomatic that pain depends on consciousness, so a demonstration that fishes can feel pain depends on showing that they are also conscious” (p. 118). This is intended to be an off-putting requirement, and it is sometimes explicitly claimed (by both sides) that it is impossible to ‘prove’ the presence of consciousness in any non-human animal (e.g. as reviewed in Elwood, 2019).

Whether or not this is simply a conservative dialectical move, it is peculiar on several fronts. First, if an animal can (consciously) experience pain, they are by definition capable of being

conscious. Evidence that an animal is capable of feeling pain is therefore evidence that the animal is capable of consciousness. If one has valid markers of pain experiences, one need not 'also' or 'first' show that the organism is conscious at all. Similarly, a capacity for consciousness of at least some kind is a requirement for a capacity for pain experiences. In this case, valid indicators of pain experiences cannot help but track the capacities and responses associated with consciousness in general. That is, research on pain experiences is research on consciousness. This fits neatly with the analysis above: researchers arguably already work with an implicit set of theoretical assumptions about what sorts of evidence are relevant to pain experiences, and where successful, these tend to track the theoretical assumptions made about consciousness in general.

More broadly, talking about consciousness does not commit one to making radically unsubstantiated claims of the sort not generally tolerated in science. Claims about pain experiences, or animal consciousness more generally, are like any other, and are generally treated as such by critics. Where arguments are made *against* claims of animal pain experiences, for example by positing plausible alternative explanations based on nociception, critics are staking a position about the truth value of these claims, or at the least stating what would need to be known or tested experimentally in order to assess them properly. That is, both proponents and critics do actually think there is a fact of the matter, and that empirical research can help us figure out what it is. This is very different from claiming that is scientifically impossible to ever know the truth value of claims about animal pain and animal consciousness.

On this basis, and given the analysis provided throughout the paper, I propose that the best way to make progress on questions around animal pain is to clearly ally them with questions of animal consciousness more generally, and to productively use conceptual and empirical work in both areas to develop more theoretically defensible criteria and experimental procedures. Most of the work in substantiating claims about animal pain is done in the theoretical arena, and it is where critics attack, often very reasonably so. This is not to say that existing claims about pain in controversial genera are likely to be wrong: above I outlined evidence not specific to pain that I think is already fairly persuasive. I also outlined which general capacities associated with consciousness seem more likely to do the work in providing persuasive evidence of pain experiences. These include spatial learning and memory, complex forms of associative learning, and complex trade-offs in decision making. Developing valid indicators based on these, and being explicit about the theoretical reasoning behind them, would be a productive way of moving forwards in this literature.

A final note: of course, this strategy might not work. It might well prove impossible to come to a general consensus on the capacities associated with consciousness (see e.g. Irvine, 2012, 2017 for a skeptical view on human visual consciousness science). In this case, other strategies will have to be used (see e.g. Shevlin, this volume, for an alternative). My hunch though is that similar issues around consensus building are likely to arise to any attempt to separate off sets of privileged capacities that are used to motivate ethical arguments about animal welfare. In that case, analyzing the capacities associated with consciousness seems like a reasonable place to start.

References

- Alves, F. L., Júnior, A. B., & Hoffmann, A. (2013). Antinociception in piauçu fish induced by exposure to the conspecific alarm substance. *Physiology & Behavior*, *110*, 58–62.
- Amano, H., & Maruyama, I. N. (2011). Aversive olfactory learning and associative long-term memory in *Caenorhabditis elegans*. *Learning & Memory*, *18*(10), 654–665.
- Andreatta, M., Mühlberger, A., Yarali, A., Gerber, B., & Pauli, P. (2010). A rift between implicit and explicit conditioned valence in human pain relief learning. *Proceedings of the Royal Society B: Biological Sciences*, *277*(1692), 2411–2416.
- Appel, M., & Elwood, R. W. (2009). Gender differences, responsiveness and memory of a potentially painful event in hermit crabs. *Animal Behaviour*, *78*(6), 1373–1379.
- Ashley, P. J., Ringrose, S., Edwards, K. L., Wallington, E., McCrohan, C. R., & Sneddon, L. U. (2009). Effect of noxious stimulation upon antipredator responses and dominance status in rainbow trout. *Animal Behaviour*, *77*(2), 403–410.
- Barron, A. B., & Klein, C. (2016). What insects can tell us about the origins of consciousness. *Proceedings of the National Academy of Sciences*, *113*(18), 4900–4908.
- Birch, J. (2020). *The Search for Invertebrate Consciousness* [Preprint]. <http://philsci-archive.pitt.edu/16931/>
- Corsi, A., Wightman, B., & Chalfie, M. (2015). A transparent window into biology: A primer on *Caenorhabditis elegans*. In The *C. elegans* Research Community (Ed.), *WormBook*. doi/10.1895/wormbook.1.177.1, <http://www.wormbook.org>
- Craver, C. (2007). *Explaining the brain: Mechanisms and the mosaic unity of neuroscience*. Oxford University Press.

- Damasio, A., & Damasio, H. (2016). Pain and other feelings in humans and animals. *Animal Sentience*, 1(3). <https://animalstudiesrepository.org/animsent/vol1/iss3/33>
- Danbury, T. C., Weeks, C. A., Waterman-Pearson, A. E., Kestin, S. C., & Chambers, J. P. (2000). Self-selection of the analgesic drug carprofen by lame broiler chickens. *Veterinary Record*, 146(11), 307–311.
- Dunlop, R., Millsopp, S., & Laming, P. (2006). Avoidance learning in goldfish (*Carassius auratus*) and trout (*Oncorhynchus mykiss*) and implications for pain perception. *Applied Animal Behaviour Science*, 97(2–4), 255–271.
- Edelman, D. (2016). Leaving the door open for fish pain: Evolutionary convergence and the utility of ‘just-so stories’. *Animal Sentience*, 1(3).
<https://animalstudiesrepository.org/animsent/vol1/iss3/36>
- Elwood, R. W. (2019). Discrimination between nociceptive reflexes and more complex responses consistent with pain in crustaceans. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 374(1785), 20190368.
<https://doi.org/10.1098/rstb.2019.0368>
- Elwood, R. W., & Appel, M. (2009). Pain experience in hermit crabs? *Animal Behaviour*, 77(5), 1243–1246.
- Felsenberg, J., Plath, J. A., Lorang, S., Morgenstern, L., & Eisenhardt, D. (2014). Short-and long-term memories formed upon backward conditioning in honeybees (*Apis mellifera*). *Learning & Memory*, 21(1), 37–45.
- Gerber, B., Yarali, A., Diegelmann, S., Wotjak, C. T., Pauli, P., & Fendt, M. (2014). Pain-relief learning in flies, rats, and man: Basic research and applied perspectives. *Learning & Memory*, 21(4), 232–252.

- Gerisch, B., Weitzel, C., Kober-Eisermann, C., Rottiers, V., & Antebi, A. (2001). A hormonal signaling pathway influencing *C. elegans* metabolism, reproductive development, and life span. *Developmental Cell*, 1(6), 841–851.
- Godfrey-Smith, P. (2017). The evolution of consciousness in phylogenetic context. K. Andrews and J. Beck, *The Routledge Handbook of Animals Minds*. New York: Routledge.
- Grau, J. W., Crown, E. D., Ferguson, A. R., Washburn, S. N., Hook, M. A., & Miranda, R. C. (2006). Instrumental learning within the spinal cord: Underlying mechanisms and implications for recovery after injury. *Behavioral and Cognitive Neuroscience Reviews*, 5(4), 191–239.
- Green, L. J., Luck, C. C., Gawronski, B., & Lipp, O. V. (2019). Contrast effects in backward evaluative conditioning: Exploring effects of affective relief/disappointment versus instructional information. *Emotion*.
- Irvine, E. (2012). *Consciousness as a scientific concept: A philosophy of science perspective* (Vol. 5). Springer Science & Business Media.
- Irvine, E. (2017). Explaining What? *Topoi*, 36(1), 95–106. <https://doi.org/10.1007/s11245-014-9273-4>
- Kästner, L., & Andersen, L. M. (2018). Intervening into mechanisms: Prospects and challenges. *Philosophy Compass*, 13(11), e12546. <https://doi.org/10.1111/phc3.12546>
- Key, B. (2016). Why fish do not feel pain. *Animal Sentience*, 1(3). <https://animalstudiesrepository.org/animations/vol1/iss3/1>
- Lints-Martindale, A. C., Hadjistavropoulos, T., Barber, B., & Gibson, S. J. (2007). A psychophysical investigation of the facial action coding system as an index of pain

- variability among older adults with and without Alzheimer's disease. *Pain Medicine*, 8(8), 678–689.
- Magee, B., & Elwood, R. W. (2016). Trade-offs between predator avoidance and electric shock avoidance in hermit crabs demonstrate a non-reflexive response to noxious stimuli consistent with prediction of pain. *Behavioural Processes*, 130, 31–35.
- Merker, B. (2005). The liabilities of mobility: A selection pressure for the transition to consciousness in animal evolution. *Consciousness and Cognition*, 14(1), 89–114.
- Merker, B. (2007). Consciousness without a cerebral cortex: A challenge for neuroscience and medicine. *Behavioral and Brain Sciences*, 30(1), 63–81.
- Merker, B. (2016). Drawing the line on pain. *Animal Sentience*, 1(3).
<https://animalstudiesrepository.org/animalsent/vol1/iss3/23>
- Mills, H., Ortega, A., Law, W., Hapiak, V., Summers, P., Clark, T., & Komuniecki, R. (2016). Opiates modulate noxious chemical nociception through a complex monoaminergic/peptidergic cascade. *Journal of Neuroscience*, 36(20), 5498–5508.
- Millsopp, S., & Laming, P. (2008). Trade-offs between feeding and shock avoidance in goldfish (*Carassius auratus*). *Applied Animal Behaviour Science*, 113(1–3), 247–254.
- Mohammadi, A., Byrne Rodgers, J., Kotera, I., & Ryu, W. S. (2013). Behavioral response of *Caenorhabditis elegans* to localized thermal stimuli. *BMC Neuroscience*, 14(1), 66.
<https://doi.org/10.1186/1471-2202-14-66>
- Nieto-Fernandez, F., Andrieux, S., Idrees, S., Bagnall, C., Pryor, S. C., & Sood, R. (2009). The effect of opioids and their antagonists on the nocifensive response of *Caenorhabditis elegans* to noxious thermal stimuli. *Invertebrate Neuroscience*, 9(3–4), 195–200.

- Rodriguez, M., Snoek, L. B., De Bono, M., & Kammenga, J. E. (2013). Worms under stress: *C. elegans* stress response and its relevance to complex human disease and aging. *Trends in Genetics, 29*(6), 367–374.
- Rose, J. D. (2002). The neurobehavioral nature of fishes and the question of awareness and pain. *Reviews in Fisheries Science, 10*(1), 1–38.
- Rose, J. D. (2007). Anthropomorphism and ‘mental welfare’ of fishes. *Diseases of Aquatic Organisms, 75*(2), 139–154.
- Rose, J. D., Arlinghaus, R., Cooke, S. J., Diggles, B. K., Sawynok, W., Stevens, E. D., & Wynne, C. D. (2014). Can fish really feel pain? *Fish and Fisheries, 15*(1), 97–133.
- Sasakura, H., & Mori, I. (2013). Behavioral plasticity, learning, and memory in *C. elegans*. *Current Opinion in Neurobiology, 23*(1), 92–99.
- Seth, A. (2016). Why fish pain cannot and should not be ruled out. *Animal Sentience, 1*(3).
<https://animalstudiesrepository.org/animalsent/vol1/iss3/14>
- Shinkai, Y., Yamamoto, Y., Fujiwara, M., Tabata, T., Murayama, T., Hirotsu, T., Ikeda, D. D., Tsunozaki, M., Iino, Y., & Bargmann, C. I. (2011). Behavioral choice between conflicting alternatives is regulated by a receptor guanylyl cyclase, GCY-28, and a receptor tyrosine kinase, SCD-2, in AIA interneurons of *Caenorhabditis elegans*. *Journal of Neuroscience, 31*(8), 3007–3015.
- Sneddon, L. U. (2009). Pain Perception in Fish: Indicators and Endpoints. *ILAR Journal, 50*(4), 338–342. <https://doi.org/10.1093/ilar.50.4.338>
- Sneddon, L. U. (2013). *Do painful sensations and fear exist in fish?*
- Sneddon, L. U., Elwood, R. W., Adamo, S. A., & Leach, M. C. (2014). Defining and assessing animal pain. *Animal Behaviour, 97*, 201–212.

- Sotocina, S. G., Sorge, R. E., Zaloum, A., Tuttle, A. H., Martin, L. J., Wieskopf, J. S., Mapplebeck, J. C., Wei, P., Zhan, S., & Zhang, S. (2011). The Rat Grimace Scale: A partially automated method for quantifying pain in the laboratory rat via facial expressions. *Molecular Pain*, 7, 1744–8069.
- Waite, M. E., Tomkovich, A., Quinn, T. L., Schumann, A. P., Dewberry, L. S., Totsch, S. K., & Sorge, R. E. (2015). Efficacy of Common Analgesics for Postsurgical Pain in Rats. *Journal of the American Association for Laboratory Animal Science*, 54(4), 420–425.
- Woodward, J. (2003). *Making things happen: A theory of causal explanation*. Oxford University Press.
- Zhou, K. I., Pincus, Z., & Slack, F. J. (2011). Longevity and stress in *Caenorhabditis elegans*. *Aging (Albany NY)*, 3(8), 733.