LETTRES
SORBONNE UNIVERSITÉ

UFR de Philosophie

Mémoire de Master II

Spécialité : Logique, Philosophie des sciences, de la connaissance et de l'esprit

Présenté et soutenu par :

Victor DURAN-LE PEUCH

2021

How can we know whether fish feel pain? Epistemology of the scientific study of fish sentience

Sous la direction de :

Mme Anouk BARBEROUSSE – Professeur, Sorbonne Université

Table of content

Introduction

Sentience refers to the ability to have subjective experiences, which corresponds to the phenomenal aspect of consciousness (Tye 2017: 10), more often when it comes to animal consciousness. Notoriously, Jeremy Bentham stressed about animals that what mattered was not "Can they reason? nor, Can they talk? but, Can they suffer?" (Bentham 1823). This intuition that what makes a living creature a moral patient is their capacity to feel and thus to suffer has become very common since Peter Singer published his flagship book *Animal Liberation* in 1975. It gradually came to compete with other features that were considered as the legitimate criteria for determining the moral status of living beings such as thought/rationality or personhood (DeGrazia 1997)[1]. Since then, arguments for the consideration of animal welfare in our ethical frameworks have been put forth within virtually all the main moral theories: utilitarianism (Singer 1980), deontology (Regan 1984), care ethics or more broadly feminist ethics (Donovan & Adams 1996), virtue ethics (Hursthouse 2011) but also within some theories pertaining to political science (Cochrane 2018). These arguments often presuppose that at least some animals such as mammals are sentient but leave it to the scientists to determine the full scope of sentient beings.

Hence, the concept of sentience is at the intersection of ethics and science[2]. Indeed, in the late 1960's a 'sentience turn' began to happen within the field of animal welfare science, which had previously been focused merely on stress as an indicator of welfare, but came to realize that understanding and studying sentience was an integral part of studying the welfare of animals (Duncan 2006: 13): the publication of Ruth Harrison's book *Animal Machines* (1964) created a small uproar in the UK and led to the conduction and publication of the "Brambell Report" in 1965 which introduced the 'Five Freedoms' as the fundamental aspects of animal welfare and stated that what the animals feel should be taken into consideration as well. This new interest in the animal experiences spread among scientists of the emerging of animal welfare science[3], of whom Marian Stamp Dawkins was an important figure with the publication of *Animal Suffering: The Science of*

[1] Today, most philosophers within the field of animal ethics seem to agree that sentience is both sufficient and necessary for having moral status. However Kagan (2019: 23) offers a compelling case that sentience is sufficient but not necessary since agency without sentience may also be sufficient for deserving moral consideration.

[2] Unlike 'qualia' which is mainly used in philosophy of mind, or 'consciousness' which has very broad-ranging uses across the interdisciplinary field of cognitive science.

[3] Most of them came from psychology or zoology, which is not surprising since animal welfare is still at the intersection of ethology, psychology (/neuroscience) and zoology (/biology).

Animal Welfare in 1980 (Duncan 2006: 14). A similar turn happened in the fields of zoology and ethology, with the publication of *The question of animal awareness* by zoologist Donald Griffin, who argued that animals are consciously aware (Wynne 2007: 130). This represented a complementary direction along with the cognitive turn that happened at the same time in psychology, which was under the grip of behaviorism before the 1960's. The name of a later book from Griffin, *Animal minds: Beyond cognition to consciousness* (2001), is very revealing of this defence of the study of consciousness as distinct from mere cognition.

But determining with certainty that an animal feels anything at all is actually a really difficult scientific question. There seems to be a scientific consensus that mammals and birds are sentient, but the evidence about the sentience of fish is more contested (DeGrazia 2020: 21). This was clearly shown by a meta-analysis on the papers published on animal sentience conducted by Helen Proctor, who concluded that most articles took sentience in mammals and birds for granted but that only a very low proportion of the article published concerned fish, and hypothesized that this was maybe due to a hesitation over fish sentience (Proctor et alia. 2013: 894). But the focal period of this meta-analysis was from 1990 and 2012, whereas the scientific study of fish sentience only really took off in the beginning of the 2000's, particularly with the discovery of nociceptors in fish (Sneddon 2003a&b). Numerous studies have since been done on fish, while a lot of epistemological debates have been taking place at the same time. There seems to be no robust consensus on which types of arguments would be acceptable to prove the existence of sentience, which poses the question of how a consensus about the sentience of a given animal can be reached at all if there is no prior consensus on the criteria that are sufficient for sentience nor on the method to assess those criteria. It seems that, based on the same empirical data, some scientists conclude in the plausibility of sentience (Sneddon 2019), while others argue to the contrary for what they consider is still a lack of evidence (Key 2016a). Hence, this scientific debate is both an empirical debate about the experiments that are led on fish, and an epistemological debate about the validity of the inferences that are drawn from the empirical findings in favor or against the plausibility of fish sentience. For that reason, I chose to focus on fish sentience in this dissertation to study how a scientific consensus can gradually form on the topic of sentience, and to try to assess which types of arguments are most prone to leading to progress on the epistemological side of the debate. The main goal of this work is to identify the types of arguments that most often lead to a stalemate in the discussions, and to propose a way out of the epistemological disagreements by focusing on the imperative of making scientific progress as a guiding principle.

I start by defining sentience and giving an analysis of the epistemological problems that plague its scientific study; this consists mainly in justifying that the attribution of sentience is underdetermined by the data. Second I show that as a result of this situation of underdetermination, most of the types of arguments used to infer sentience from the data are inconclusive and lead to a stalemate. Third, I argue that the stalemates arise from a foundationalist epistemology which needlessly leads to skeptical conclusions; as an alternative, I propose to adopt a coherentist framework and defend a process of 'epistemic iteration' (Chang 2004) within that framework, which I argue gives us a way out of the underdetermination.

I. Assessment of the epistemological challenges of evaluating sentience

1. Defining sentience, pain and welfare

A. Sentience

Sentience refers to the ability to have subjective experiences, which corresponds to the phenomenal aspect of consciousness, or to the "what it is like" to be something (Nagel 1974), consciousness itself being a multi-faceted concept without a unified nor unanimous definition. Birch et alia. (2020) distinguish between five dimensions of consciousness in order to map out the consciousness profiles of different types of animals: perceptual richness, evaluative richness, integration at a time, integration across time, and self-consciousness. None of these mesurable, comparable, objective dimensions of consciousness directly enlighten us on what is at stake when we wonder whether an animal is sentient: can it feel pain or pleasure? can it have positive and negative experiences? It seems perfectly conceivable for an animal not to be self-conscious, to have a bad memory, and not much perceptual richness but to still be able to feel things, and in particular to feel pain. It is prima facie that last subjective dimension that is most important from an ethical point of view because only then can what is being done to them have a moral importance *to them*

(and not merely by having indirect consequences on moral agents for example)[4]. Similarly, it would be judged immoral to torture a human being who had lost self-consciousness but could still feel pain. Consequently, sentience only depends on one aspect of consciousness, the phenomenal and subjective aspect.

The topic of phenomenal consciousness in non-human animals actually raises two distinct questions:

- The Distribution Question: Can we know which animals beside humans are conscious?
- The Phenomenological Question: Can we know what, if anything, the experiences of animals are like? (Allen & Trestman 2020: section 4)

Note that it is much more difficult to answer the phenomenological question than the distribution question: even for humans, it's much easier to know that they are experiencing something at all than to know exactly what they are experiencing. Logically an enquiry into sentience should always start with the distribution question anyway, since it makes no sense to ask what the experience of a certain animal is like if there is no experience. Thus, the question that is being debated about fish sentience for now is the distribution question, namely whether they feel anything at all and in particular pain. Maybe the phenomenological of what fish pain feels like and how different it may be from human pain will come if and once a robust consensus that fish feel pain emerges.

B. Pain and welfare

Many perceptual cognitive systems can be associated with a phenomenal experience. For example in most humans, it's the case for visual system but not for the vomeronasal system, which apparently responds to pheromones and affects human behaviour but is not associated with any phenomenal consciousness (Allen 2004: 630). A good case study to outline the difference between a mere cognitive state and the phenomenal aspect that can be associated with it is the phenomenon of blindsight, where patients can respond to visual stimuli and make systematically right 'guesses' about shapes or objects, but don't actually see anything in the sense that they have no visual

[4] Carruthers (1992) thus argues that we should treat animals correctly not because they are conscious but to fulfill a duty towards ourselves.

experience at all and claim they are blind. So the visual information gets to their brain and is processed, but the phenomenal aspect of it is absent.

The field of animal welfare science has been very focused on the study of pain, which is the conscious mental state that can be associated with nociception, which itself is only the unconscious signal that alerts the organism of noxious stimuli. The International Association for the Study of Pain (IASP) defines pain more precisely as "an unpleasant sensory and emotional experience associated with, or resembling that associated with, actual or potential tissue damage" (Terminology | International Association for the Study of Pain). Pain is only one type of a negative phenomenal conscious state, so it is only one aspect of the study of sentience. Arguably an organism could have conscious visual experiences but no pain, and on the contrary another could feel pain but only have an unconscious visual system. Since sentience is the capacity to feel anything at all, both these organisms would qualify as sentient. The reason that pain is more studied than any other conscious state in the field of animal welfare science is probably that, of all the states of suffering, it's the one responsible for the bigger reduction of welfare in animal agriculture (Duncan 2006: 15). But in principle, nothing should stop scientists from studying the other dimensions of sentience, and in particular some of them have called to focus not only on the suffering of animals but also on their positive affects, since the presence of pleasure adds much in the quality of life (Duncan 2006: 16; Proctor 2012: 636).

In line with what has been done by scientists, this work focuses on pain as a telltale of sentience, and I sometimes use pain and sentience interchangeably, while keeping in mind that other dimensions of sentience could and probably should be studied more thoroughly in the future. The difference between both terms becomes relevant again in Part III when talking about the issue of the semantic extension of concepts.

Is is also important to note that pain, although a very central element of welfare, is only one element of it. It is possible to study and try to improve the welfare of animals even when they are not experiencing phenomenally conscious states. Broom gives the example of an individual with a broken leg but asleep, or someone who is passed out on drugs, or affected by a disease without being aware of it (Broom 2016: 3). There is certainly a way of addressing part of the animal welfare

issues without enquiring about their experiences, for example by taking care of their injuries and giving them antibiotics to prevent diseases (even if we suppose that they don't cause them pain)[5].

But since in this work we are looking to know whether certain animals can suffer, let's examine the main epistemological problems that come up in the scientific study of sentience.

2. Scientifically addressing the epistemological problem of sentience

A. Rejection of 'first philosophy'

This work adopts a rather naturalist approach, in the Quinian sense of not adopting the stance of a 'First Philosophy', the illusion of benefitting from an extra-scientific, overarching stance upon how science should work. As Quine puts it: "Naturalism sees natural science as an inquiry into reality, fallible and corrigible but not answerable to any supra-scientific tribunal, and not in need of any justification beyond observation and the hypothetico-deductive method" (Quine 1981: 72). In this perspective, philosophy is considered as continuous with science and not as a final judge of the justification of scientists' beliefs and methods of inquiry. Mary Leng for example applies that approach to her philosophical study of mathematics, and defends the philosophical modesty that comes with it: "Sharing the concerns of scientists to provide a coherent theoretical systematization of our experiences, philosophers should approach the question of the justifiability of our theoretical claims using the same kinds of concepts, methods, and styles of reasoning as are accepted by scientists" (Leng 2010: 36). This doesn't mean that philosophers should take the assumptions of scientists at face value without questioning them, but they shouldn't think that their work has any prescriptive precedence over the way that scientists already think about those methodological problems[6]. In particular, if a philosophical theory entails that scientists are mistaken in holding certain beliefs, then the philosopher should probably refrain from calling for a revision of practices that are deemed successful by the scientists themselves.

A good example of 'first philosophy' in the case of animal sentience is the argument against the sentience of all non-human animals put forth by Carruthers based on a particular theory of

[5] See (Duncan 2004) for objections to this point and a defence of the view that feelings are the only thing that matters for animal welfare.

[6] As we will see, the scientists studying fish sentience make themselves abundant use of philosophical tools and arguments.

consciousness. He used to defend a higher-order theory of consciousness (HOT), according to which a "conscious state is one that gives rise to an activated second-order belief in their own existence" (Carruthers 1989: 508). And from that theory he inferred that, since animals don't have second-order beliefs, they don't experience anything. Apart from the fact that empirical findings seem to have shown since that at least some animals can have second-order beliefs[7], the very definition of the capacity to have experiences through higher-order beliefs is not self-evident but instead very much open to discussion[8]. Different theories are in competition to explain phenomenal consciousness (Block 2009) and there is simply too much uncertainty and long-standing disagreements around what the right theory of consciousness could be to be confident in drawing robust inferences from theories to attributions of sentience (Michel 2020: 796).

Instead of this top-down approach, animal welfare scientists have applied a more pragmatic bottom-up approach to try and test for sentience, without having to presuppose too much in terms of theory. It seems to be better suited for the study of sentience, at least at this stage of the scientific research:

> While it may seem natural to think that we must have a theory of what consciousness is before we try to determine whether other animals have it, this may in fact be putting the conceptual cart before the empirical horse. In the early stages of the scientific investigation of any phenomenon, putative samples must be identified by rough rules of thumb (or working definitions) rather than complete theories. Early scientists identified gold by contingent characteristics rather than its atomic essence, knowledge of which had to await thorough investigation of many putative examples — some of which turned out to be gold and some not. Likewise, at this stage of the game, perhaps the study of animal consciousness would benefit from the identification of animal traits worthy of further investigation, with no firm commitment to idea that all these examples will involve conscious experience. (Allen & Trestman 2020: section 4.8)

As a matter of fact, most scientists participating in the debate on fish sentience seem to share a relative degree of theoretical non-commitment about what sentience is, and try instead to build empirical experiments to better stake out its boundaries. In fact, they seem to be very pragmatic, in the sense of being in line with the philosophy of pragmatism. Hasok Chang identifies the key features of pragmatism: seeing knowledge as rooted in practice, insisting that concepts should be operable (that they should be put to use in concrete empirical work), recognizing a degree of

[7] For example, (Hare et alia. 2001) show that in some situations chimpanzees know what conspecifics have seen, and thus what they know, from which it is reasonable to infer that they can have second-order beliefs about their own states.

[8] See (Allen 2004: 634) and (Tye 2017: 18) for objections to Carruthers.

fallibilism and anti-skepticism as well as seeing science as continuous with concerns of life (Chang 2012: 197). I come back in part III on those elements of pragmatism.

Thus in this work, instead of trying to develop an overarching theory of sentience, I look directly at the scientists' practices, in particular their argumentative practices since I am mostly interested in the epistemological part of the scientific debate. I try to identify among those types of argumentation which ones are most prone to helping make empirical progress and reach a better theoretical systematization, which means assessing them by the scientists' own lights. Before doing that though, it is worth identifying the epistemological problems that come up with phenomenal consciousness, and try to distinguish among those problems which are relevant to assess whether fish feel pain.

B. The hard problem of consciousness

By comparison with the easy problems of consciousness -finding the neural mechanisms that explain certain cognitive functions associated with consciousness such as the integration of information or the reportability of mental states- Chalmers identified a hard problem of consciousness (Chalmers 1995). It is the problem of explaining why we have felt, subjective, qualitative experiences at all; in other words, the existence of sentience is itself a mystery that calls for an explanation. This is one of the toughest problems in philosophy, but luckily one does not need to address it in order to establish which animals are sentient. Indeed, let's imagine that in the distant future, we had a full characterization of which animals are sentient and of which neural mechanisms give rise to experiences; in this hypothetical situation, the hard problem could still not be solved since it would still be a mystery to explain how subjective experiences in 1st person can be created by objective physical brain functions. As Chalmers puts it:

> The problem persists even when the performance of all the relevant functions is explained. [...] This is not to say that experience has no function. Perhaps it will turn out to play an important cognitive role. But for any role it might play, there will be more to the explanation of experience than a simple explanation of the function. (Chalmers 1995: 202-206)

Since the hard problem is over and above the simpler problem of finding neural correlates that are both necessary and sufficient for phenomenal consciousness - which would allow for a systematic testing for sentience in all animals - it is not really a problem for scientists, at least not

those who are only looking to assess which animals are sentient. As previously said, this more modest endeavor doesn't require a full-fledged theory of sentience, even less an explanation for its very existence in the material world. Hence the hard problem is rightfully ignored in the scientific debate on fish sentience[9].

C. The problem of other minds and the resulting problem of detection

Whereas the hard problem seems almost insoluble but can easily be (at least temporarily) discarded in the scientific enterprise, another problem arises that is not without consequence for the scientists. It is the classic problem of other minds: because of the intrinsically subjective and private nature of phenomenal consciousness, it is in principle unobservable in third person. We only have indirect evidence of the existence of an inner life in other living beings. For most fellow humans, this indirect evidence is the testimony they can provide through a common language. Many humans like babies, comatose patients, people with certain mental illnesses, cannot express what they feel, but we can infer their mental experiences through even more indirect means than language, such as behaviour, neural activity in certain cortical regions, stress hormone levels, etc. Non-human animals are in the same case of not being able to testify of their experiences through a common language with us (Harnad 2016: 5).

The difficulty is however not completely intractable for all non-human animals, particularly for those which are closest to us phylogenetically speaking such as the primates. Indeed the phenomenon of blindsight for example, which distinguishes between conscious vision and unconscious processing of visual information, was first observed not in humans but in monkeys (Humphrey 1974). But it is true that the more we consider animals that are very different from humans, the more indirect means we have to employ in order to assess whether they are sentient. Fish are sufficiently different from us in terms of brain structures and behaviour that any certainty seems to be lost when it comes to their potential sentience.

Whereas the hard problem was an explanatory problem, the problem of other minds is an epistemological problem, that is a problem about the limits of our knowledge. Initially, it's meant as

[9] Except maybe for those who try to give a functional definition to sentience, and argue that the best explanation as to why sentience exists is meant to be an evolutionary explanation. More on the question of defining sentience functionally in Part I-3.

a skeptical challenge: strictly speaking, any certainty is lost as soon as we talk about other living beings than ourselves. But this isn't as big a problem for science as one might have thought, for two reasons: scientific explanations are only done in 3rd person so anything scientific that can be said about sentience is going to be knowledge of a different nature than the introspective knowledge we have of our own experiences; moreover science doesn't work with absolute certainty anyway. As Andrews puts it, "the skeptical problem should not be seen as a problem for science. Science is grounded on background assumptions we cannot prove, such as the existence of an external world and the existence of other minds." (Andrews 2020: 23) Since no amount of empirical evidence would be able to defuse the skeptical challenge, it simply falls outside of science.

So in the context of the scientific study of sentience the epistemological problem of the knowledge of other minds has to be rephrased into a methodological problem about how to come up with methods that allow us to be reasonably confident about our attributions of sentience. Let's call this methodological problem resulting from the problem of other minds, the problem of detection.

- Problem of detection: the problem of coming up with scientific methods to systematically detect phenomenal consciousness in other animals, both human and non-human, with reasonable confidence.

The problem of other minds was in a sense binary -I can only know my own experience, and cannot know for sure that others experience anything at all- whereas the problem of detection is one of degrees: the methods to systematically detect sentience become more and more difficult to come up with as we consider animals that are very different from us.

Once the implicit demand of certainty in the skeptical challenge of the problem of other minds is lifted, scientists are left free to try out imperfect ways of answering the problem of detection and then gradually refine them. The crux of the problem is that sentience cannot directly be observed, and is all the less observable as we consider animals that are different from us. So the way sentience can be scientifically conceived of is probably in a similar manner as an (as of yet) unobservable physical property, much in the same way as temperature was unobservable before numerical thermometers were developed and refined (Chang 2004: 87). Here observable is not meant in the van Fraassian sense of being observed with the unaided senses if it were present to us, but in a sense that takes into account the contingencies of scientific progress and what the tools of

science allow us to see more or less indirectly but with a "reliable determination from sensation" (Chang 2004: 86). According to that sense, in a hypothetical situation[10] whereby microbes would have been postulated as theoretical entities that explained a lot of observable macro phenomena before the invention of microscopes, then they would have become observable only once the microscopes were developed, refined to be reliant, and precise enough to see the microbes whose existence had until then only been hypothesized (Chang 2004: 85)[11].

It would probably be too strong to say that there is a widely shared physicalist assumption about sentience among scientists who study it, whereby sentience would be nothing more than one more cognitive function and shouldn't receive any preferential treatment. Many scientists recognize its very particular nature (Dawkins 2015: 26), but when it comes to empirical work they often eschew its supposed intrinsically private nature and treat it as something that has all the relevant characteristics of an unobservable physical property[12]: we can at least in some situations detect it with our senses -although only on ourselves for direct evidence- and it correlates with other physical properties. In particular, phenomenal consciousness is thought to be analysable within a causal chain of the form: physical cause —> phenomenally conscious state —> behaviour.

Let's stress the fact that this way of proceeding is rational: scientists have no other choice since there isn't a scientific way of studying what is intrinsically private; so they choose the pragmatic option of trying to analyse it like they would any physical property, assuming that sentience can somehow be detected although they don't know exactly how yet, and see if they manage to come up with empirical solutions to the problem of detection. Similarly, mathematicians treat numbers as if they existed, and this assumption -regardless of its truth, which is a hotly debated topic in philosophy of mathematics- allows them to do their job really well. So philosophical scrupules about phenomenal consciousness being defined as intrinsically private should not stand in the way of empirical inquiries.

3. The underdetermination of sentience by the scientific evidence

A. Finding a set of criteria for pain

[10] This fictional example is due to Maxwell (1962: 185–7)

[11] I will come back on this definition of observability as non dichotomous in part III.

[12] It doesn't have to be an actual belief that scientists have, but their practices are based on that implicit assumption.

Once sentience is constructed as a physical unobservable property, scientists can try to test for it indirectly by studying the correlations between sentience and the objective states of affairs that seem to be systematically associated with experiences. That means trying to come up with a list of criteria that could discriminate between cases where there is phenomenal experience with cases where there isn't. Let's focus right away on the case of pain, since most debates concentrate on that aspect of phenomenal consciousness.

The objective correlates of pain can be of different types:

1. Physiological: neuronal, organization of the nervous system, integration of different functional systems in the brain, hormone levels, temperature, heart rate, etc.
2. Behavioural: nocifensive behaviours, cries, grimaces, apathy, etc.
3. Other cognitive functions: emotions, intelligence, preferences, etc.

The physiological correlates are probably the most important ones because in a physicalist perspective the mechanisms of pain are assumed to lie in a certain organization of systems within our bodies. However this is a broad category and not every physiological correlate has the same value in terms of evidence for the presence of pain: the presence of nerves seems prima facia more important than a change in the hormonal balance. But this hierarchy is only the result of our current understanding of the mechanisms of pain, which tells us that stress hormones such as adrenaline are sometimes but not always associated with pain (Dawkins 2015: 8).

Behaviour is also a very broad-ranging category; it probably requires more interpretation to establish the correlations with pain, but is also more easily observable than internal physiological systems. Hence why behaviorism, an approach based almost exclusively on observable behaviour and which tended to exclude inner mental states from the phenomena that could be called upon in a scientific explanation, was dominant in both psychology and ethology until the cognitive turn of the 1960's. It's important to note here that technically speaking human testimony through their pain is also a kind of behaviour, admittedly a very elaborate one but ultimately it too is vulnerable to the skeptical challenge of other minds and inferring pain from verbal reports is an indirect means of knowledge (Dawkins 2015: 5). In consciousness science experiments, reports can sometimes be wrong or imprecise.

The other cognitive functions are actually an improper category because each function has physiological and behavioural components, but it is still interesting to consider the correlation with other functions as a whole, because it allows for another level of analysis, the functional level. Since pain is defined through a certain phenomenology -contrary to mere nociception-, it is not prima facie reducible to a set of functional properties. But it is clearly associated with some functional properties that have been identified, such as emotions, memory or preferences.

As an example, some scientists have developed methods to investigate the preferences of some animals. They gave the animals a choice between two possibilities and let them choose the option they preferred (Fraser & Nicol 2011). That allowed them to learn for example that when given a choice, hens unsurprisingly prefer outdoor spaces to battery cages but also that they prefer fine-gauge wire to heavy-gauge wire for the floor of their cage, which ran contrary to what was assumed until then (Dawkins 2015: 12). Methods were also developed on top of that to know how much more the options were preferred, by using operant learning[13]: the animals had to pay a cost, by having to push a lever a certain number of times for example, before obtaining what was tested for; you could then raise the cost to see if the animal was still willing to pay it for the reward (Patterson-Kane et alia. 2008). Preference is a cognitive state that doesn't have to be associated to conscious experience. But if the cost to pay is an electric shock for example, it can give you information about how pain (or mere nociception) is linked with preferences. Looking for these links with other cognitive functions allows for a better understanding of pain and may then allow to distinguish between nociception and pain in other animals: if it was found that in mammals only pain is associated with preferences, and not mere nociception for some reason, then observing a strong preference to avoid noxious stimuli in other animals such as fish would be interpreted as a fairly good indicator of pain and suffering. The same type of analyses can be made for the link between pain and memory, or pain and emotions, etc.

Once the objective correlates of pain are identified, scientists should ideally determine the causal relations between them and phenomenal consciousness. However, it's generally difficult to establish the exact causal links; in the meantime, scientists can try and work out criteria that, if not always causally connected to consciousness, can serve as useful heuristics to test for it. The point is that it must be possible to come up with an operable list of criteria that would allow the assessment

[13] Operant methodologies were especially developed by Duncan.

of sentience in an animal, without the need to have got to the bottom of the mechanisms of consciousness beforehand.

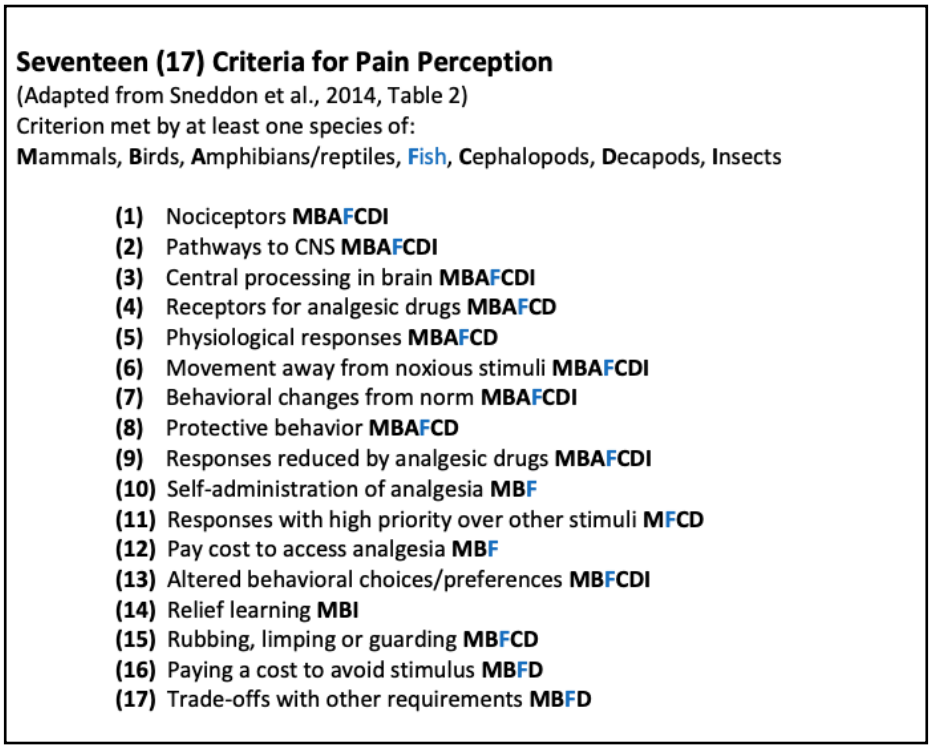
In humans, we have a clear enough picture of the sufficient conditions for pain[14]: the existence of nociceptors, of a nervous system to transmit the information from the nerve endings to the brain, a processing centre in the brain which receives the information (the thalamus), itself functionally connected to a set of other brain structures (the limbic system and several other neocortical structures).

Several remarks can be made about the conditions that are found to be sufficient for pain in humans:

- each condition taken separately is not sufficient for pain, only the set of all conditions is;
- some conditions may not be formulated precisely enough, and thus overly general conditions (such as 'the presence of a neocortex') may not be necessary whereas once broken down and precise enough some of the sub-conditions may be (such as certain properties like signal amplification or signal integration (Seth 2016: 1)); this means that some sub-conditions might be subtracted from the set of conditions, and the new set would still be sufficient for pain. The formulation of the conditions is thus of paramount importance in order to assess which conditions really are necessary;
- a set with slightly different conditions may also be sufficient (with another kind of nociceptors, or a different organization of the nervous system or of the neuronal structures) to realize sentience. This is the argument from multiple realizability, which I will come back to in greater details.

How can scientists compare the conditions that are jointly sufficient for sentience in humans with the correlates that they find in fish? Fish have nociceptors, a nervous system and a processing centre (Sneddon 2011) but don't have a neocortex like mammals. However, research has shown that they meet many of the criteria that are considered to be a sign of pain. (Sneddon et alia. 2014) came up with a list of seventeen criteria to try and "triangulate" (Walters 2018: 2) the likely existence of a pain state, which then allows for comparison between different types of species:

[14] We'll see in the next section that determining whether a condition is necessary for pain raises the problem of underdetermination.



List of 17 criteria for pain perception, taken from (Walters 2018), initially adapted from (Sneddon et alia. 2014)

To determine which conditions are really necessary for pain, scientists need to rely not only on the experiments that strive to establish the causal links, but also on a series of other elements in order to mount arguments for or against the importance of a given condition.

Here are the most important of those elements:

- **functional analysis**
- **phylogenetic proximity**
- **plausible evolutionary story**

As we've already seen, what role phenomenal consciousness plays and why it emerged are a mystery -that's the hard problem. So there is no consensual functional analysis of sentience as there is for short-term memory. The most promising attempts to define consciousness functionally is probably through its etiology, that is how it came to appear and how it brought additional fitness to certain creatures so as to get selected in evolution (Black 2020: 9); doing so implies once again

sidestepping the hard problem[15] and applying a physicalist assumption to consciousness. If we had a functional definition of sentience, it would certainly be much easier to assess whether a certain animal is sentient because we could check whether that function was useful in the animal's environment, whether similar animals needed that function, etc. Arguments of phylogenetic proximity would also come in if we had a rough idea of when consciousness developed in evolution.

But such a functional definition does not yet exist for consciousness, and scientists chose not to wait for it before trying to come up with criteria for pain. In the meantime, the absence of a fully developed functional analysis doesn't exclude the use of functional hypotheses and scientists still rely on a rough functional understanding of conscious states.

Phylogenetic proximity can be useful to determine whether a trait is shared by two species because it tells you how far back you have to go to find their common ancestor. Since there is an evolutionary continuity, the closer two species are on the phylogenetic tree, the likelier they are to share many traits. For example, chimpanzees share many traits with humans because they are the mammal with the most recent common ancestor with us. Hence it is very probable that they feel pain too, since statistically there is very little chance that the trait 'feeling pain', or 'phenomenal consciousness' in general, developed only for humans after this common ancestor. Conversely, if birds were thought not to feel pain, it would be very improbable that fish would because the common ancestor between fish and humans is also the ancestor of birds, and it is very rare that a trait would develop, disappear and then reappear further down the phylogenetic tree.

These previous considerations can be combined with a plausible evolutionary story. It consists in trying to account for the development of a trait based on the fitness it brought relatively to the environment of the creatures in question at the time. It requires to have at least a partial functional definition of the trait, in order to make conjectures about its additional fitness. For example, if we have good reason to think that pain is interesting for creatures because it allows for a certain behavioural flexibility, the observation of nocifensive behaviour or of learning mechanisms to avoid noxious stimuli will be interpreted as good signs that the animal feels pain. The evolutionary story helps to refine the analysis of the objective correlates of pain, since they are

[15] It consists in doing as if there was no explanatory gap and directly ask how certain kinds of mental states could bring an evolutionary fitness and capacities that were interesting for some creatures. It is an open question whether providing a full functional analysis of consciousness would lay the hard problem to rest.

always interpreted in a certain light, based on our functional assumptions on pain. A good example of that is the debate about whether one should expect fish to become agitated or on the contrary to stop moving if they felt pain. Key suggests that scientists should focus on the suppression of normal behaviour rather than on responsive behaviour due to noxious stimuli, because the latter could just be a reflex whereas the former is a better sign that pain would actually be felt: "This has led to a call for animal studies of pain to assess pain-suppressed rather than pain-elicited behaviours" (Key 2016a: 15). But maybe the suppression of normal behaviour makes less sense for species that are not social like humans, since being apathetic might render them more vulnerable to predators.

This is only an example of the type of reasoning involved when assessing which criteria are the most important, and which should be discarded. It shows that the conditions are certainly not just a checklist which gives you the certainty that there is pain if all the boxes are ticked. The conditions have complicated interactions between each other, which require using external considerations and usually elicit a lot of debates.

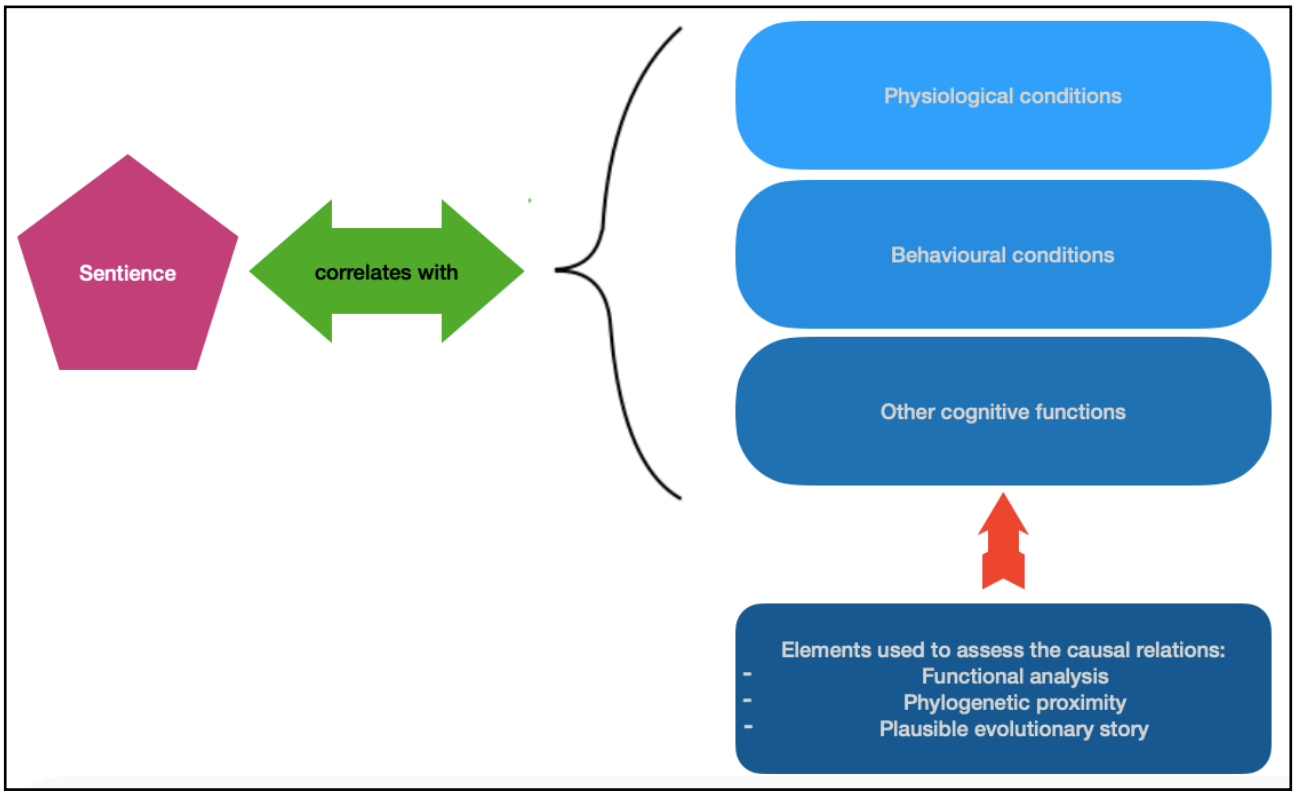


Diagram of the objective correlates of sentience and the additional considerations to assess them

As soon as we have a set with slightly different conditions than the one corresponding to human sentience, which is virtually almost the case when we consider non-human animals, how can scientists determine whether that set is sufficient to realize sentience? It seems as though there isn't enough agreement on the rules to determine which conditions should be retained, and as though the inference from a set of conditions to sentience can always be contested. And indeed in the case of fish pain it does appear that given the same set of conditions present in fish, some conclude that they feel pain (Sneddon 2019), whereas others that they don't (Key 2016a). When the same empirical data allows for multiple theories or assertions, there's a good chance that this is a classic situation of underdetermination. Let's examine why that is indeed the case for fish pain.

B. Underdetermination

The reason that it is so difficult to come up with a definitive list of criteria for pain is that the attribution of sentience is underdetermined by the evidence. The fact is that if we knew exactly which species are sentient, it would make it possible to work out the list of criteria that are necessary for sentience, by keeping track of the conditions that are present in all those species but not in non-sentient species. But of course knowing which species are sentient is precisely the distribution problem that we are trying to answer, which would be feasible if we had a precise definition of what sentience is and how to detect it, but we don't have either at the moment. There is thus a circularity between the criteria for sentience and the distribution problem. Technically, the attribution of sentience is underdetermined because there are multiple sets of (list of criteria + list of sentient species) that are all compatible with the evidence.

This results from the lack of direct observability of sentience in non-human species. But it's important to stress here that underdetermination is a rather common difficulty in science. For example, various theories of the atomic-molecular constitution of water were compatible with the evidence at the time -including the proposition by Dalton that water was HO- with a circularity this time between molecular formulas and atomic weights. Hasok Chang identified no less than five competing systems of chemistry that were compatible with the evidence at the time, and it took almost fifty years to reach the consensus that water is H2O (Chang 2012: 133). How much the situation is underdetermined depends partly on the amount of evidence since some sets might not be compatible with newly found evidence. Another way out of underdetermination is through a new

theoretical breakthrough, which was the case for water (Chang 2012: 173) and which could be the case for sentience if a functional analysis based on its etiology was bolstered.

So the situation of underdetermination is a consequence of the problem of other minds, but it could be resolved in a similar way as other underdetermination situations were for physical properties or entities that were initially unobservable[16], such as temperature. Unlike the debate about the molecular formula of water which was solved by theoretical advancement, the underdetermination of the measurement of temperature was somewhat resolved before a unified theory of temperature came to be adopted (Chang 2004: 155). I will argue in part III that this could also be the case for sentience: in other words, scientists can find a way out of the underdetermination without having to wait for a fully developed functional analysis of sentience.

Let us characterize even more precisely the underdetermination of the list of criteria for phenomenal consciousness. Mathias Michel offers the following analysis:

> We cannot justify necessary conditions for consciousness in a non-circular way. To put this argument formally:
> (1) To find a criterion that demarcates between plausible and implausible cases of multiple realizations of consciousness, one must provide a necessary condition for consciousness.
> (2) To know whether a condition C is necessary for consciousness, one must test hypotheses of the form: for all entities, if E is conscious, E satisfies C.
> (3) In order to test these hypotheses, one both needs to know whether the tested entity E satisfies C, and whether E is conscious or not.
> (4) However, we do not know if E is conscious or not, since it is precisely what we try to assess.
> (5) Therefore, we cannot find a criterion that demarcates between plausible and implausible cases of multiple realizations of consciousness. (Michel 2019: 14)

According to him, the underdetermination only concerns the necessary conditions. That's because Michel assumes a framework whereby structure determines function. That means that when you have a full description of a certain physical structure you can in principle deduce which function it realizes. But the converse isn't true: function doesn't determine a particular structure, hence multiple structures can, in principle at least, realize the same function. For the matter at stake here, that translates into the fact that if we find conditions that we know are sufficient for consciousness in some beings, then we can deduce the existence of consciousness. But if a condition that we deem necessary for some beings is lacking, then we cannot deduce the absence of

[16] Still assuming that sentience should be treated as an unobservable physical property in a scientific perspective.

consciousness since consciousness could be realized by other structural conditions. This is the multiple realizability argument that stems from the functionalist tradition (Michel 2019: 6)[17].

Much of the debate around fish sentience hinges on that issue of multiple realizability. Indeed, fish have most of the criteria considered sufficient for pain in humans, except for particular brain structures that are found in the neocortex and which they lack. The crux of the issue is to know whether that condition really is necessary for pain, which is what defenders of the 'no cortex, no cry' argument hold (Michel 2019: 2), or if different brain structures could also realize pain in fish, which is what defenders of the multiple realizability of pain argue for.

Michel fears that as soon as you accept the principle of multiple realizability, you lose the possibility of finding necessary conditions for pain because in principle another structural condition could realize the same function. It then becomes impossible to present evidence that fish don't feel pain, because someone may systematically argue that pain is realized differently in fish. Consequently, you also lose the ability to even determine a lower boundary to sentience, since any organism might actually realize sentience with different physical structures. Here we stumble again upon the circularity between the list of conditions for sentience and the list of creatures which are sentient, and hence this leads to underdetermination.

I will argue in part III that there are actually some ways worth exploring that would allow to move past the underdetermination. In particular, there may be ways of distinguishing between plausible and implausible cases of multiple realizability which don't imply being able to identify necessary conditions with absolute certainty. That would refute the first premisse of Michel's argument which states that providing a necessary condition is the only way of distinguishing between plausible and implausible cases. But we can already give temporary and intuitive reasons for doubting this argument: it seems that the scientific debate about fish sentience is not really an argument about multiple realizability in general, but rather about whether the brain structures in fish are homologous enough to those of mammals to reasonably infer that they can also realize pain. Scientists who defend that they are homologous don't seem committed to saying that any structure could realize pain, nor that no lower boundary for sentience can be found in the phylogenetic tree. I

[17] Note that multiple realizability can apply to states, properties or functions. So depending on what we're focusing on, we can say that phenomenally conscious states are multiple realizable, or that the property of sentience is, or that the function which sentience supposedly performs is.

will defend a distinction between a narrow and a wide multiple realizability, and try to show that you can have the first without the other.

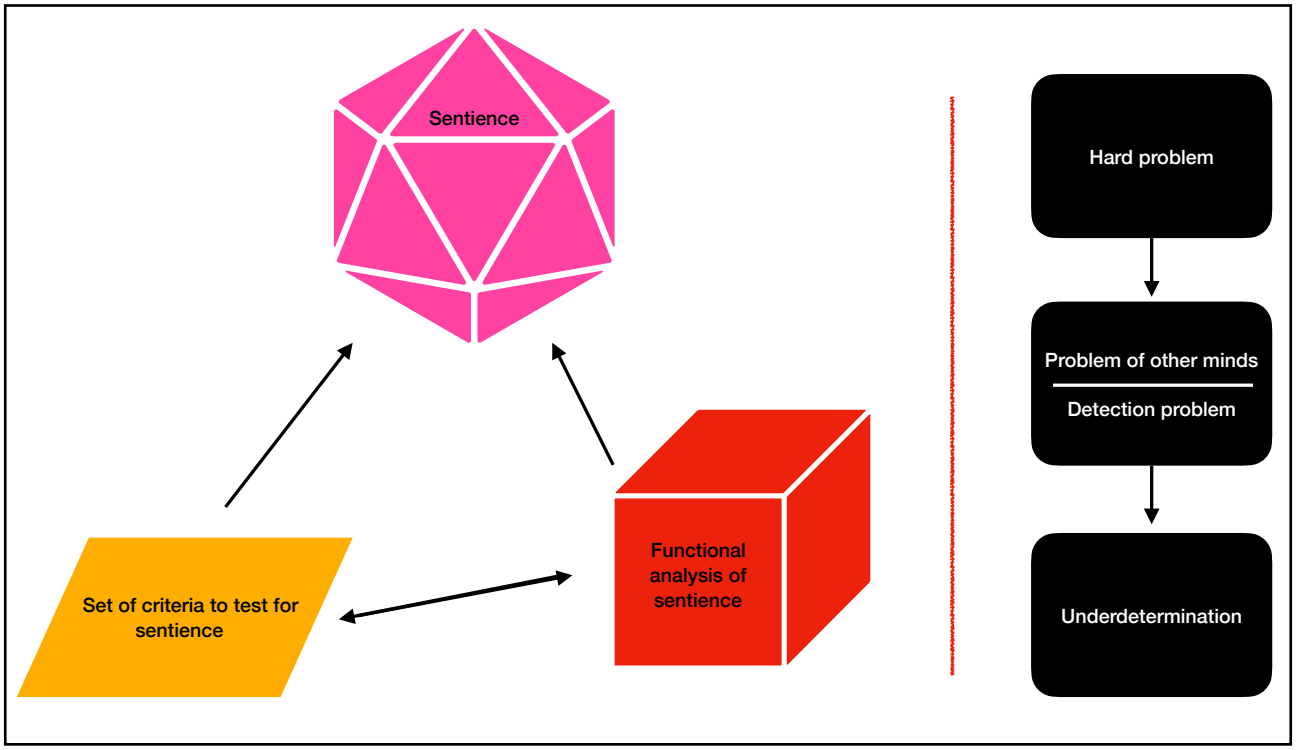
C. Assessment of the epistemological difficulties so far

A bunch of difficulties come bundled together when we start thinking about phenomenal consciousness. However, once distinguished the three problems of underdetermination, the problem of other minds (or its scientific rephrasing the problem of detection), and the hard problem, the task at hand turns out to be less daunting than anticipated. Indeed, answering the distribution problem of sentience has virtually nothing to do with the hard problem of consciousness since we could come to have a good empirical understanding of sentience, and be able detect it or even to measure without even coming close to solving the mystery that it represents for philosophers. A good comparison would be mathematics: we can learn a lot about mathematical objects, and how to measure them or use them in models without ever solving the mystery that the very existence of mathematics represents or getting to know whether mathematics are abstract objects or concrete physical ones.

The problem of other minds was analysed for what it is: a skeptical challenge that cannot be dealt with through scientific means. Scientists will by default treat sentience as a physical property, and will accept less than absolute certainty, just as we commonsensically accept that other people have minds without strictly speaking having access to them. It just means that one way of identifying sentience, although very reliable, has an extremely limited scope; but science can just resort to other means, and this is the whole issue with the detection problem, that of coming up with such indirect methods and procedures that would allow a reasonable degree of confidence in the attributions of sentience. So this problem is one of degree which can in principle be met by scientists.

The real issue that seems to stand in the way of solving the distribution question of sentience is that of underdetermination. I argue that there are two main ways of dealing with this issue, depending which conception of sentience we adopt to analyse it. Either we can try to develop a functional analysis of it and work out a way to operationalize it, but this requires a certain degree of theoretical grounding for sentience, and we've seen that there is no consensus on the theories of phenomenal consciousness. So another way to try to analyse sentience is by trying to come up with

a list of criteria that would serve as a reasonable approximation for the detection of sentience. This bottom-up approach is certainly less neat than a functional analysis, and almost looks like a 'homemade solution' to the problem. But its relative lack of ambition compared to the functional analysis probably makes it the best candidate to make gradual progress through scientific experiments. Hence, I offer the following schematic to sum up the levels of abstraction of the problems surrounding sentience and the corresponding levels of the conceptions of sentience that can be adopted to tackle these problems:



Schematic of the epistemological problems around sentience, and the conceptions of sentience that can be adopted to deal with them

Finding a functional definition of sentience would certainly allow for its operationalization and thus for the elaboration of an operable set of criteria within a detection procedure. But working on an approximate set of criteria may also foster interesting theoretical hypotheses that might lead to a functional definition, or can serve as a testing arena for functional hypotheses. So it appears that both ways are very complementary and should be pursued concomitantly in order to come to a better understanding of what sentience actually is, and maybe one day unravel its mystery.

Nevertheless, it seems that both ways are plagued with underdetermination. Indeed, what poses a problem for the necessary conditions of the set of criteria poses as much problem for a functional definition of sentience. What matters for this work is that it is prima facie possible to

make headway on the side of the set of criteria without the need of a prior functional definition, since this is the way I'll be following along with the ethologists and biologists studying sentience.

In order to break through the ceiling of underdetermination, scientists try to come up with epistemological principles or types of reasonings that can arbitrate the debate. This will be the topic of part II, but before that I would like to dispel right away one specific principle which could be proposed: the precautionary principle. The way the principle is applied to animal consciousness, the idea is that whenever the evidence is inconclusive, we should give the animal the benefit of the doubt and assume that it is sentient out of precaution (Birch 2017: 1). But this principle is meant as a guide to action for political or ethical decision based on the uncertainty; what it is not however is a scientific principle as to how scientists should proceed to better address the problems surrounding the study of sentience. In other words, in this work we want to know if fish really feel pain, or whether we have reasonably good reasons to assume it, and not whether we should assume it for extra-scientific reasons. That isn't to say of course that scientific work should be completely independent of ethical of political decisions, because how scientists treat the animals in their experiments implies moral decisions for which it might be relevant then to apply a precautionary principle. But it doesn't entail that the demandingness for robust evidence should be lowered.

II. Types of arguments that fail to break away from underdetermination

Following the perspective set out in the section on the rejection of 'first philosophy', I try in this part to take stock of the most common arguments marshaled by scientists more or less explicitly to argue for or against an attribution of sentience. This descriptive dimension is combined with a normative assessment of whether these arguments allow to resolve the situation of underdetermination. This is not done in a spirit of telling scientists how they should think about these issues or even less how they should do their work, but rather with the goal to show which arguments lead to stalemates between opposing views, which hinders empirical progress and calls for another way of maybe settling the debate, by the scientists's own lights.

1. Anthropomorphism or anthropocentrism

A. The risks of anthropomorphic biases

Defenders of conservatism about animal consciousness -conservatism being the position which refutes the attribution of sentience for a given animal (Murray 2020)- often accuse the other side of anthropomorphism, that is of wrongfully projecting human characteristics upon animals. This is seen as a pre-scientific bias[18] which leads to mistakes in the interpretation and explanation of animal behaviour. For example, it would be a mistake to interpret the smile of dolphins or of chimpanzees in the same way as a human smile: the mouth of dolphins is shaped in a way that we humans spontaneously interpret as a smile, even when they are actually suffering, whereas for chimpanzees showing your teeth amounts to threatening another individual (Evers 2019: 21). In this case, attributing a similar mental state as ours based on the similarity of behaviour is indeed wrong-headed and pre-scientific. As for the case which interests us here, some argue that attributing pain to animals when we observe pain-like behaviour is one more example of the tendency humans have of anthropomorphizing animals, and it should be resisted (Key 2016: 2).

Wynne furthermore denounces the risk of committing a 'nominalist fallacy', that is the belief that naming something explains it (Wynne 2007: 125). But as Burghardt argues, the labelling part of animal behaviour is only a first phase in the work of a ethologist, which only lays the ground for further analysis. It does not explain in itself but makes it much easier to include the behaviour observed into an explanation, which is precisely the point of a label (Burghardt 2007: 136). But of course labels have an influence on the further interpretations of scientists, and calling a certain behaviour 'pain' might ultimately be unwarranted even if it made the empirical work easier. (Rose et al. 2014: 100) denounce the irreflective use of the label 'pain' in many scientific experiments when nothing more than nociception has been proven, which only causes confusion between nociception and pain further down the line.

B. The risks of anthropocentric biases

[18] This comes with a skepticism towards folk psychology (Evers 2019: 21).

This criticism of anthropomorphism actually stems from the behaviorist tradition within psychology and ethology and the debate around anthropomorphism is a long-standing one. But behaviorism has come under fire since the 1960's, and this has allowed a lot of discoveries about the mental lives of animals. Frans de Waal argues that the behaviorist avoidance of anthropomorphism has mostly hindered empirical progress in the field of ethology (de Waal 2016: 55). Several ethologists have in fact defended the use of a form of anthropomorphism, distinct from the pre-scientific 'naïve anthropomorphism' (Burghardt 2007), as a relevant scientific tool: Gordon Burghardt has defended using 'critical anthropomorphism' which he defines as a "heuristic method to formulate research agendas that result in publicly verifiable data that move our understanding of behaviour forward" (Burghardt, 1991, p. 86); Marc Bekoff (2000) has defended a 'biocentric anthropomorphism' and Frans de Waal (1997) an 'animal-centered anthropomorphism'. What they argue is that some of the labels they use, such as 'laughter' or 'anger' or 'pain', aren't only useful labels that allow for shortcuts but can sometimes be misleading: sometimes it just would not even make sense to use different labels just because they apply to non-human behaviour. And that is a reflection that comes from their practice as scientists who are on the ground and learn how to notice the relevant differences between human and non-human behaviour, not from uninformed naïve anthropomorphic tendencies that everyone naturally tends to have[19].

Those scientists who rely on anthropomorphism thus denounce what they see as an opposite bias in the defenders of conservatism about sentience, namely anthropocentrism or 'anthropodenial' (de Waal 2016: 22). According to them, many mistakes result from the fact that almost all the concepts and experiments are thought relatively to a human standard. It becomes almost trivial to deny non-human animals the characteristics that were defined according to the particular way in which they manifest themselves in humans: as a result, the tests we think of to determine whether a certain characteristic is shared by another animal are tailored to humans and oftentimes very unadapted for non humans. And this comes from a lack of imagination, because scientists are often too centered on humans to put themselves in the animal's shoes as much as they should.

A good example of this is the multiple failings of the mirror test, initially developed for chimpanzees by Gordon Gallup in the 1970's to test for self-consciousness. The animal would be anesthetized and then marked with an odorless color on a body part that the animal can't normally see, and then would be presented with a mirror; if they noticed the mark and/or tried to scratch it, it

[19] For more details about the different types of anthropomorphism and how they can legitimately be put to use by scientists, see (Evers 2019).

was considered a good sign that the animal recognizes itself in the mirror and sees the reflected image as an image of itself. Chimpanzees successfully passed that test, but gorillas failed it. But it turned out that was only because gorillas try to avoid direct gazes towards them, because they find it aversive, so they had no motivation to interact with the mirror. So failure to take into consideration the typical behaviour of the two different species led ethologists to the wrong conclusion for gorillas (Andrews 2011: 482). Similarly, elephants were also deemed lacking self-consciousness, just because the first experiment was ill-conceived and the mirror was too small for the size of the elephant and placed outside its cage (de Waal 2016: 17). Moreover, in the case of aquatic animals it isn't clear at all what failure of the mirror test really implies, since they cannot touch their bodies when they see a mark on them, or maybe they have no motivation to remove the mark; and there is the fact that smell is a much more important sense than vision for many of them because odors are much more reliable underwater. However, it was recently shown that a fish passed the mirror test (Kohda et alia. 2019), fact on which I will come back in part III.

C. An inconclusive argument for fish sentience

The jury seems to be still out about which bias, the anthropomorphic or the anthropocentric one, has been more detrimental to the scientific enterprise of better understanding animal behaviour, cognition and sentience. But one thing to note is that the risk of anthropomorphism may be greater for species that are phylogenetically close to us, their ressemblance with us making anthropomorphic assumption more spontaneous, whereas on the contrary the risk of anthropocentrism seems greater for species distant from us humans. It may thus be wise to be more particularly careful of anthropocentric attitudes when it comes to fish.

But although anthropomorphism can be reclaimed as a valid methodological principle, as a valuable tool to form hypotheses, it can at best count as a heuristic and not technically speaking as a form of argument. Heuristics can lead you to make better and maybe broader empirical findings. That is why Kristin Andrews calls for a recognition that every scientist comes to her work with her own specific biases, which can lead her to specific types of discoveries. Consequently, the normative principle that allows for most empirical progress is a pluralist principle: a variety of people with differing biases should be encouraged to get into the scientific arena and they should be judged not on their biases but on the results that these biases lead them to (Andrews 2020: 41). She also points out that research has to start with "naturalistic observations and folk

expertise" (Andrews 2011: 490), and that this is a strength that enables to better devise the experiments, not a weakness. This is particularly egregious for psychologists making experiments with children: they know they have to play with them and earn their trust in order for the children to get interested in the tasks they have to perform. There is no reason it should be any different for experiments with non-human animals (Andrews 2011: 476). Likewise, she identifies five harms that can result from ignoring consciousness such as missing out on certain topics of study -no one would try to study what animals feel if they didn't assume that they feel something- or not noticing relevant variables -chimpanzees perform certain tasks poorly not because they are not capable of doing them, but on the contrary just because they are bored, but the experimenter has to assume they can feel bored in order to realize this (de Waal 2016: 34)[20]. Hence, at least some scientist working with animals should assume they are conscious, or else that could lead to many missed opportunities and wrong interpretations for their empirical research (Andrews 2020: 20).

However, the considerations so far have only concerned the context of discovery, in terms of what can elicit the most promising empirical results. But it doesn't seem that the analysis through the lens of anthropomorphism/anthropocentrism brings much to the debate about fish sentience in terms of justification or proof. The inference from the empirical data to an attribution (or not) of sentience only comes in a second time in the scientific work. And anthropomorphism/ anthropocentrism seem very relevant to the production of the empirical data and less so for the later inference. That inference can hardly be driven by a pluralist principle as for the biases of the experimenters, since the conclusion should be either that fish are sentient or that they aren't. At that level of reasoning, the accusations of anthropomorphism or anthropocentrism -which are still commonly thrown in the debates about animal cognition and sentience- are not conclusive for the discussion. And in particular, they don't give a way out of the underdetermination.

2. Morgan's canon as a principle of parsimony

Regularly invoked in the study of animal cognition is Morgan's canon, which derives from a principle of parsimony in the explanation. It states: "In no case is an animal activity to be interpreted in terms of higher psychological processes if it can be fairly interpreted in terms of

[20] It may look like begging the question, but by making certain assumptions scientists may find clues that they were right in assuming so that they wouldn't have found otherwise. Hence why this is a heuristic and not a model of valid argumentation.

processes which stand lower in the scale of psychological evolution and development" (Fitzpatrick 2008: 224). It often comes conflated with the criticism of anthropomorphism, because Morgan's canon was formulated in the time when anthropomorphic attitudes were being questioned within psychology and ethology (Wynne 2007: 129), but I think both can be analyzed separately. As we've seen, anthropomorphism is either a bias or a heuristic depending on how we value it; Morgan's canon is a more precise methodological principle. One could thus reject anthropomorphism without subscribing to Morgan's canon, which means that the latter needs independent positive justification. Another difference is that, whereas the lens anthropomorphism/anthropocentrism was mostly used as a heuristic in the context of discovery, Morgan's canon is also employed as a justification for inferences about animal cognition and consciousness:

> On the most widespread reading of Morgan's Canon in the modern literature then, Morgan's Canon amounts to an *inference principle* that can be used to decide between competing explanations of animal behaviour. This reading has had a far-reaching influence in debates in animal psychology: time and again theorists have used Morgan's Canon to justify substantive claims about the nature of animal minds. (Fitzpatrick 2008: 229)

At first glance, the canon seems quite reasonable, since parsimony in the explanation is widespread as a scientific value. But it actually presents a number of problems, the main of which being that more general simplicity considerations don't necessarily support Morgan's canon. The crux of the issue is that the canon is based on only one type of simplicity, which can easily conflict with other types. Fitzpatrick lists five of them (2008: 230) but I choose to only focus on one of them here, which is best expressed by de Waal: "The pursuit of cognitive parsimony often conflicts with evolutionary parsimony" (de Waal 2016: 43). In other words, evolution doesn't necessarily follow the most economical ways because it is based on random mutations for the most part. This is why we are now able to devise programs that perform cognitive functions much better than humans and with much more simplicity than the neuronal connexions in a human brain; it can be interesting to develop the simplest possible algorithms to perform the same cognitive functions as we do but this isn't what will explain how a human brain effectively works. Of course considerations of simplicity will have a role in evolutionary explanations because lack of simplicity creates a cost in fitness which is less likely to be selected than a better alternative; but natural selection doesn't apply on an infinity of possibilities, only on those which were made possible by random mutations.

Furthermore, since we accept that humans have cognitive capacities and conscious states, supposing that other animals don't have them implies that those traits would have developed only after their common ancestor, which itself calls for an explanation because of the common

assumption of gradualism in evolution. As de Waal puts it: “Evolutionarily speaking, it would be a true miracle if we had the fancy cognition that we believe we have while our fellow animals had none of it.” (de Waal 2016: 43). So when thinking in evolutionary terms, very few hypotheses come with no explanatory cost and it becomes less obvious which default position we should adopt if we want to respect a simplicity principle. In fact, Morgan himself wouldn’t have agreed with what his successors have sometimes made of his canon: he added a caveat to his principle by explicitly stating that there is nothing wrong with attributing complex cognitive capacities once they have been proven to exist in an animal (de Waal 2016: 42).

So Morgan’s canon in itself doesn’t give us a principled way of drawing the inference from the evidence to an attribution or not of sentience.

3. Null hypothesis and burden of proof

The question of anthropomorphism and Morgan’s canon were really local debates within ethology and psychology. There is however a more general way of looking at the issue while still relying on a principle of parsimony in the explanation. Key argues that in the debate about fish pain, the ‘null hypothesis’ is that they don’t feel pain (Key 2016b). In a statistical model, the null hypothesis is the hypothesis that there is no relationship between two phenomena under investigation (Fidler & Wilcox 2021: section 1.2). Another way to define it is to say that the null hypothesis is that chance alone is responsible for the link between both phenomena. It is then possible to try and test the null hypothesis and from that test, two types of errors can occur. A type I error -‘false positive’- means wrongfully rejecting the null hypothesis, that is to consider that both phenomena are causally linked when they actually aren’t. A type II error -‘false negative’- means wrongfully accepting the null hypothesis when it would have been rejected because both phenoemena are actually causally linked. For Key, concluding that fish feel pain runs the risk of committing a type I error.

Nevertheless, these definitions of the null hypothesis don't apply very well to the case of sentience, because it is not a question of determining whether two identified phenomena are causally connected (which would refute the null hypothesis), but whether a presumed phenomenon -pain or phenomenal consciousness- is what explains an observed phenomenon -pain-like behaviour. The study of sentience is simply not based on a statistical model, because it would

require that sentience be an identifiable phenomenon, independently of its objective correlates; we saw this is one of the epistemological difficulties with sentience.

So I'll assume that Key is using the concept of 'null hypothesis' as an analogous way of designating a default hypothesis, the one that would be the weaker conclusion, or the most parsimonious. Another way of saying this is that the burden of proof falls on those who wish to defend that fish do feel pain, who have to 'falsify' the null hypothesis according to Key (2016b: 4). The use of Popper's falsificationist vocabulary by Key reminds us that it is much more difficult to prove the absence of pain than to prove that it is present; in this perspective, the hypothesis that they don't feel pain is thus more falsifiable because you can admittedly come up with empirical data that would refute it. So initially, it makes a lot of sense to defend the idea that the burden of proof has to be taken by the scientists who strive to prove that fish feel pain.

But the very problem is that there isn't any meta-consensus on the amount of evidence needed to successfully 'falsify' the null hypothesis, this question being precisely what is being debated in the no cortex, no cry argument for instance. Once there is mounting evidence of the presence of many criteria for pain -as there is for fish is one accepts the list that (Sneddon et alia. 2014) developed- it doesn't remain clear that the default assumption for scientists should still be that fish don't feel pain.

Actually, talk of a default hypothesis or of a burden of proof seems too binary to correctly represent the dynamic process in which scientific evidence slowly piles up and gets debated in the scientific arena. Especially now that at least twenty years of research on fish pain has been done, claiming that the burden of proof is on the side of claiming that fish feel pain is unhelpful to settle the underdetermination situation.

4. Analogical reasoning and Inference to the best explanation

A. Analogical reasoning

The attribution of conscious states to certain animals relies first and foremost on an analogy with humans: animals that are similar enough to humans can be presumed to share some of the traits which are known to exist in humans. In the case of sentience, scientists are forced to resort to analogy precisely because it is not directly observable for non-human animals. So they assess the

relevant similarities between species, systematically compare the conditions for sentience in humans with what they can observe for other animals, and then use an analogical argument for drawing the inference from the empirical data to the attribution or not of sentience.

An analogical argument has the following form (Bartha 2019: section 2.2):

1. *S* is similar to *T* in certain (known) respects.
2. *S* has some further feature *Q*.
3. Therefore, *T* also has the feature *Q*, or some feature *Q** similar to *Q*

So they infer an unknown similarity from other known similarities. This is an ampliative argument, meaning that the conclusion isn't certain even if the premisses are true but it is supported with varying degrees of strength, just like induction.

Analogy is useful to form a first judgement, or work as a heuristic to form scientific hypotheses, but it soon reaches its limits. Indeed, an analogical argument can in principle always be rejected by claiming that the similarities between S and T are not relevantly connected to Q. In the case of phenomenal consciousness, analogy doesn't offer a way out of underdetermination since determining which similarities are relevant to attribute sentience is the very thing under study. So it would be difficult to find an independent justification for criteria that would determine whether an analogy is valid in the case of sentience. Analogy can be a useful tool, and is virtually ubiquitous in the reasonings of scientists on the issue (Dawkins 2001: 22), but it has to be integrated in a more general argument in order to have any hope of settling the debate. This is explicitly how Tye conceives of the standard argument made for the attribution of sentience:

> The argument I have given—that the hypothesis that you have beliefs and desires is justified since it affords the best explanation of your behavior—is an example of inference to the best explanation. It is not an inductive argument by analogy. An analogy is effectively made, of course, in the hypothesis—that you are like me in having beliefs and desires—but that analogy does not figure in an argument by analogy. (Tye 2017: 55)

B. Inference to the best explanation

By default, one would assume that the same effects have the same causes. This rather intuitive principle was explicated by Isaac Newton: "The causes assigned to natural effects of the same kind must be, as far as possible, the same" (quoted in Tye 2017: 72). Hence, by applying

Newton's rule we can infer from the similar pain-like behaviour we observe in humans and fish that the cause for humans, the feeling of pain, has to be the same for fish since it provides the best explanation for it. According to Tye, this reasoning is nothing else than an inference to the best explanation (IBE) relying on the analogy between pain-like behaviour in humans and fish: assuming that common effects have a common cause is the best available explanation *unless* there is evidence which undermines that inference and makes an alternative explanation more appealing (Tye 2017: 74). The 'unless' is important because it means that one can come up with a defeater of the inference to the best explanation.

In the case of fish pain, the absence of a neocortex seems to be such a defeater (Key 2016b). But Tye argues that it is actually more costly to suppose that fish undergo 'ersatz pain', that is a state that has all the same causes and effects but doesn't feel like anything at all (so mere nociception and not actual pain), since it doesn't offer an explanation of the difference of causes between similar effects in humans and fish:

> It appears, then, that the hypothesis that you undergo ersatz pain is not as good an explanation of the facts as the hypothesis that you experience pain just as I do, for the former hypothesis generates a further explanatory puzzle with respect to me that the latter hypothesis does not. This puzzle is created by what was supposed to be a virtue of the ersatz pain hypothesis over the pain hypothesis, namely, its relative simplicity. We now see that this "simplicity" comes at a real explanatory cost. So, all things considered, the pain hypothesis is the best explanation after all. (Tye 2017: 58)

The way Tye defines 'ersatz pain' entails that it has exactly the same causes and effects as actual pain, but doesn't have a phenomenal dimension. Given that definition, defending the existence of 'ersatz pain' would entail that actual pain is epiphenomenal since it doesn't differ from 'ersatz pain' in the causal chain. That would indeed represent an auxiliary assumption that would create more puzzles than it would solve, and thus pain would obviously be a better explanation than 'ersatz pain'. But Tye is here making a straw man of Key's argument: Key never assumed that the causes and effects of the pain-like behaviour in fish were identical with that of humans. They are at best similar, and introducing the concept of 'ersatz pain' is too strong and makes it easy to refute.

There are actually three problems with Tye's argument. The first is that determining what are 'similar effects' relies on an analogy which can, as we saw in the last section, be rejected. Key claims for example that the causes of the pain-like behaviour in humans and fish are relevantly dissimilar for the analogy to be rejected. This isn't settled by including it in an IBE -and this is the second problem- since what counts as a defeater is also underdetermined. So resorting to Newton's

rule combined with the possibility of a defeater doubly reintroduces the problem where they were supposed to offer an independent justification, and thus begs the question (Murray 2020: 167). The third problem is that what counts as the 'best' explanation relies on explanatory virtues such as simplicity or unificatory power; but we've seen in the section on Morgan's canon that simplicity has different dimensions which can be mutually conflicting, and so can the various explanatory virtues. So there might be substantial disagreement as what counts as the best explanation in the debate, independently of the first two difficulties; which means you can't rely on the best explanation to settle the difficulties with the analogy or the defeater. The underdetermination plagues the whole argument.

More fundamentally, it seems that what goes awry in Tye's argument is that his reasoning is very binary. It assumes that there's a default position, and that this can be overturned by one specific defeater. So applying Newton's rule is just like reversing Morgan's canon, or the null hypothesis, and runs into the same difficulties as that of settling where the burden of proof lies. This binary thinking looks unfit to deal with the complexity of the formation of a consensus on the issue of fish pain. Note that the problem is not IBE in general but the specific binary way in which Tye uses it by relying on Newton's rule and defeaters. For example, Lipton's characterization of IBE (Lipton 2004) isn't binary at all and on the contrary encourages the multiplication of 'contrastive questions' -actively trying to come up with slightly different effects or causes- in order to select explanatory causes by a process of 'causal triangulation' (inspired by Mill's Method of Difference). Hence, in Lipton's characterization of what he defends as the best versions of IBE, there is no default hypothesis which can be defeated, but rather a gradual process towards identifying better and better causal explanations until reaching a certain degree of confidence (Lipton 2004: 42). Lipton's version of IBE seems much better suited to describe how the scientific debate on fish pain progresses.

None of the types of arguments assessed in this part have been shown to provide a way out of the underdetermination that plagues the debate on fish pain. In his conclusion, Michel draws a pessimistic conclusion about the possibility of a scientific consensus given the current evidence available: "If one does not want to go beyond the current neuroscientific evidence, one should thus remain agnostic on the issue of consciousness in animals such as fish, unless we discover that they realize sufficient conditions for consciousness." (Michel 2019: 21). I argue in part III that this is being overly pessimistic, not by defending a different type of argument which may succeed where

all the others have failed, but rather by defending a different epistemological framework altogether and presenting an overarching process which can be combined with those arguments. Indeed, I think that at least some of these arguments[21] fail because they are embedded within epistemological assumptions that must be questioned if we want to avoid a stalemate in the debate.

III. Coherentism and epistemic iteration: the way out of underdetermination?

The types of arguments presented in part II don't unlock the debate on fish pain. The reason for that is that they fail to tackle the underdetermination which was analysed in part I. I will argue in this part that the arguments are only lacking if they are used in a foundationalist epistemological framework, and that there is way of moving forward in the debate by adopting a coherentist framework instead. I will first defend that framework and draw the consequences it has on the way the debate has been phrased so far; second, I will present the process of epistemic iteration which this new framework allows and show how it can give us a way out of underdetermination; finally, I will analyse how this process can be applied for the case of fish sentience.

1. Moving past the underdetermination: the imperative of progress

A. Empirical progress as a pragmatic goal

Scientific discussions around fish pain often result in a stalemate, in the sense that two polarized parties -for of against fish pain- fail to come to a consensus; their arguments are opposed and they start arguing past one another. We are left with inconclusive arguments about anthropomorphism against anthropocentrism, or about who has the burden of proof, or which explanation is the best, all in a binary way that seems to lead nowhere. In such a situation, what should govern the scientific work is probably less which argument is susceptible to lead to a

[21] Not all the arguments fail for the same reason; in particular anthropomorphism is not very relevant for the context of proof regardless of the epistemological framework.

consensus (because a winner-takes-all situation in that context is unlikely to happen), but rather how to avoid deadlocks and how to allow for empirical progress despite the disagreements.

It would thus seem reasonable to identify and defuse the claims that purport to put an end to the debate in one direction or the other, in particular when those claims don't allow for new empirical input to count. Those can be arguments that set an impossibly high bar for new evidence to be worth considering. I think we should be able to say that a new experiment showing a sign of nociception in fish has a consequence on the debate about pain even if it is not a full proof of the existence of pain. There isn't gonna be one decisive experiment to prove pain in fish; rather the consensus may emerge from the accumulation of indirect evidence. The discovery that fish are willing to pay a cost in order to get access to an analgesic when they have been subjected to noxious stimuli (Sneddon 2012) should be considered, if not sufficient to prove pain, at least relevant in the debate and an interesting empirical result. However, the conservatives in this debate often insist on the difference between nociception and pain to conclude that such experiments prove nothing about pain (Rose et al. 2012: 101 & Key 2015: 150). This reasoning is not favorable to empirical research. The accumulation of signs of nociception can make scientists question what they considered heretofore as a defeater for pain; at least the possibility should be left open, and more empirical research encouraged.

But the conservative side seems to consider that no amount of indirect evidence for pain will be enough if there is even only one defeater. The problem with defeaters is that they work in an all-or-nothing way: the argument either holds completely or is equally defeated whether there is one defeater or many of them. But this assumption leads to complete stalemates when scientists don't agree on what counts as a defeater, and doesn't elicit more research.

The rejection of indirect evidence for pain might stem from a failure to recognize that because of the problem of other minds, any certainty is lost once we try to address phenomenal consciousness in third person. All you can ask for in the study of sentience is reasonably convincing evidence, not ironclad evidence; and the conclusions are structurally going to be drawn against a backdrop of high uncertainty. Donald Griffin cautions against what he dubbed 'paralytic perfectionism', which corresponds to the way certain scientists demand absolute certainty before they can accept animal sentience (Proctor 2013: 883).

The other pitfall would be to recognize that uncertainty but then conclude from it that one should remain agnostic on the issue. As was defended in part I and as will be shown in this part,

science has ways to proceed even in such a situation of uncertainty. One should not use the uncertainty that necessarily results from the first two epistemological problems of sentience -both the hard problem and the problem of other minds- against the scientific attempts to address the third one, the underdetermination. That is because the best way to address the underdetermination is not by bulletproof arguments but by a slow process of building up the evidence. Chang deplores that many philosophers fall into what he calls 'the certainty trap', because asking for certainty requires to defeat radical skepticism -and let's not forget that the problem of other minds is a skeptical challenge-, which is an almost impossible enterprise: "skeptical scrutiny is useful when it explodes overblown claims of certainty, but it cannot be a positive program of work" (Chang 2012: 243).

A way to both accept the uncertainty and not fall into an agnostic solution is to be resolutely pragmatic. We saw in part I that the key features Hasok Chang attributes to pragmatism are: seeing knowledge as rooted in practice, insisting that concepts should be operable (that they should be put to use in concrete empirical work), recognizing a degree of fallibilism and anti-skepticism as well as seeing science as continuous with concerns of life (Chang 2012: 197). The aspect of fallibilism and anti-skepticism is certainly the one that is most important in the study of fish sentience, and it is the failure to abide by it that leads to the various deadlocks that were analysed in part II.

B. Moving from foundationalism to coherentism

The main motive for skepticism in the debate on fish sentience is the impossibility of resolving the circularity between which creatures are sentient and what are the criteria for sentience by coming back to safe grounds, with beliefs that would be either self-evident or firmly justified. This is a foundationalist vision of justification, whereby one constructs one's system of beliefs in a hierarchical way by building upon the basic ones. As we saw, when it comes to sentience we have firm beliefs to begin with: our own phenomenal consciousness. But as soon as we want to start building other beliefs on top of it, the solidity of the structure begins to crumble since you cannot derive much from this belief with any certainty. The construction of our system of beliefs concerning the attribution of sentience should thus be based on another vision of justification, namely a coherentist one. Coherentism is what gives us a pragmatic way out of skepticism. We no longer have derivation from basic beliefs upon which the whole system depends, but rather an

initially sketchy construction of mutually dependent beliefs which have to be consistent as a whole and are still connected, although loosely, to more firmly justified beliefs. Justification is only found in the coherence of elements which lack ultimate justification in themselves (Chang 2004: 221).

Interestingly, this coherentist way of constructing the attribution of sentience is not only relevant for non-humans: it has long been relied upon implicitly for humans as well. Indeed, we cannot always get report on people's conscious states because not every human is able to make such reports; moreover, they may lie, or not give reliable testimony on their own mental states. More or less operationalized criteria for pain have been developed nonetheless for young babies, comatose patients, people with certain mental illnesses such as autism, or patients with brain trauma, etc. Sentience is a question that cuts across species differences, since there are also questions about who is sentient among human individuals: one debate in bioethics is to determine for example when fetuses develop sentience, and others debates regard people with severe mental disabilities. For these human cases, we do not accept uncertainty as a reason not to do our best to come up with criteria to be able to attribute pain or sentience. And we don't require absolute certainty before accepting the hypothesis that a baby feels pain. So scientists are entitled to ask: why should we accept the fallibility of conclusions and the use of coherentist reasoning for humans and not for non-humans?

An important consequence of a move towards a coherentist framework is that there is no right answer as to when the process of investigating sentience should stop: there isn't one decisive moment when we can declare that fish feel pain. But the confidence slowly piles up, until it becomes more and more costly to negate that fish feel pain without having to deny other better established claims.

The second consequence of the move to a coherentist framework is the need to change our conception of what observability is. In my exposition of the problem of other minds, I chose to follow Chang's definition of observability as what can be reliably determined from sensation, instead of the van Fraassian sense of being observed with the unaided senses if it were present to us. Chang's definition has the advantage of taking into consideration the contingencies of scientific progress, and offering a vision of observability that very much suits the coherentist framework that I am arguing for. It could be defined by the slogan: "observability is an achievement" (Chang 2004: 86). In the case of sentience, what scientists are trying to do is to go from a situation where sentience cannot be observed to one where it can be better observed, that is better determined from

sensation by reliable processes. This observability is neither dichotomous nor completely continuous, but is rather reached by stages. Chang analyses four of those stages for the case of temperature:

- Stage 1 is the bodily sensation of hot and cold
- Stage 2 is the use of thermoscopes using the expansion of fluids to measure differences in temperature precisely, but thermoscopes don't give us any absolute temperatures, only relative ones with an arbitrary unit (in other words, thermoscopes are only ordinal and not cardinal)
- Stage 3a is the development of thermometers based on the freezing and boiling of water as fixed points
- Stage 3b is the development of numerical thermometers as above, with the boiling point replaced by the steam point because it is found to be more reliable (Chang 2004: 47)

So observability is a question of more or less, but improving observability is most often done in jumps. And this is precisely why that conception of observability is particularly suited for our new coherentist framework: what matters is not so much being able to fully determine who is sentient (although ultimately this is the goal of course), but that there be an enhancement of the reliability to do so. Reaching new, more precise and reliable stages of observability is already an achievement even if it doesn't allow for a fully operationalized procedure to attribute sentience.

C. Consequences for the 'no cortex, no cry' argument and underdetermination

The move from a foundationalist to a coherentist framework sheds a new light on Michel's analysis of the 'no cortex, no cry' argument and of the situation of underdetermination. Let's go back to Michel's characterization of the circularity in the debate about fish pain:

> Providing us with good reasons for thinking [that pain is actually multiply-realized] would require finding pain in types of entities with brains that are sufficiently different from the brains of mammals. But, once again, finding any evidence in favor of the multiple realizability of pain will be impossible without running into a circular argument. Indeed, in order to test whether pain is multiply realized in a creature that is sufficiently different from us, we must know whether that entity has conscious experiences of pain or not. Without this information, we cannot validate nor invalidate the hypothesis that pain is multiply realized (Michel 2019: 16)

The circularity that leads to underdetermination results from the lack of knowledge that an entity has conscious experiences. But asking for *knowledge* instead of for instance *reasonably strong reasons for* is a foundationalist requirement: Michel demands to get back on safe grounds and thus discards any other less certain way to assert that a creature is phenomenally conscious.

This amounts to preemptively using the skeptical challenge of other minds against any attempt to address the underdetermination situation, which is nothing else than ‘paralytic perfectionism’ that we analyzed as one of the pitfalls of a failure to accept fallibilism and that leads to deadlocks. Again, we don’t ask for certainty for humans, so why should we in the debate for fish?

Michel’s worry about multiple realizability was primarily that once you accept the general principle, no amount of evidence can then be presented against a claim of multiple realizability since you cannot prove the absence of consciousness (you may just have an absence of evidence towards the claim of the presence of consciousness). In other words, if pain is multiply realizable, there is no way to prove that rocks don’t feel pain; because anybody could in principle defend that rocks realize pain differently than sentient animals do. According to Michel, there is no principled way to distinguish between plausible and implausible cases of multiple realizability. The first premisse in Michel’s argument which allowed him to come to that conclusion was “To find a criterion that demarcates between plausible and implausible cases of multiple realizations of consciousness, one must provide a necessary condition for consciousness” (Michel 2019: 14). This is certainly one robust way to do it, but Michel doesn’t justify that it is the only way to do it.

In fact, everyone should agree that it is much more plausible that fish feel pain than rock do, even though we don’t have access to either one’s alleged conscious states with certainty, so Michel’s claim is counter-intuitive. His claim only makes sense in a foundationalist framework where you have to build your hierarchy of beliefs on safe grounds. However, the plausibility of a presumed case of multiple realizability can be assessed by the indirect ways that I presented in part I-3-A, by identifying more and more correlates of pain and comparing it with what can be observed in non-human animals, with the additional help of functional analysis, phylogenetic proximity and an evolutionary story. Although these criteria are not perfect, plausibility is not certainty, and this is how a coherentist framework can somewhat defuse Michel’s pessimistic argument.

He writes: “Coming up with a criterion that satisfies one’s pre-theoretical intuitions about what is conscious and what is not is just as good as stipulating that some animals are conscious, and others are not, without arguing in favor of one’s claim.” (Michel 2019: 13). Yet, the uncertainty around the list of criteria for sentience should not lead one to think that those criteria are adopted in an arbitrary manner. Scientists weigh carefully the criteria they put forth, which are then debated in the community; and some criteria are recognized as more important than others. So what Michel describes does not seem to correspond to the way the debate around fish pain is actually conducted.

The precise way in which gradual progress can be made in the determination of sentience, while avoiding the reliance on arbitrary assumptions will become clearer in part II-2.

In this part II-1, I have made two negative claims: that arguments which don't allow fallibilism should be discarded for pragmatic reasons and that more generally such arguments are based on a foundationalist framework which is not suited to allow for empirical progress in the debate on fish sentience. I have made one positive claim: a coherentist framework is better suited for it. However, coherentism is only the epistemological framework, and it remains to be shown exactly how it is possible to make headway in the debate by adopting it. This is what I attempt to do in Part II-2 by defending the process of epistemic iteration, which I argue gives us good criteria to distinguish between plausible and implausible cases of multiple-realizability.

2. Epistemic iteration and semantic extension

A. Minimalist framework of principles to enable epistemic iteration in the case of sentience

Chang defines epistemic iteration as "a process in which successive stages of knowledge, each building on the preceding one, are created in order to enhance the achievement of certain epistemic goals" (Chang 2004: 45). The concept is inspired from mathematical iteration, but it differs from it in an important way: you apply mathematical iteration to approach a solution that is known, so you know you will get to the truth; on the contrary, you have no guarantee that what you get from the application of epistemic iteration is truth. If we had a way to know whether applying epistemic iteration would gradually get us closer to the truth, it would mean that we had a way of assessing truth independently of the process itself and thus would not have needed it in the first place (Chang 2004: 227); in particular this seems to the case for the attribution of fish sentience, as we've seen already. So the assessment of epistemic iteration should not be made in a foundationalist manner by asking whether it leads to justified conclusions, but should rather be made in a coherentist way. Chang defends that "the real potential of coherentism can be seen only when we take it as a philosophy of progress, rather than justification" (Chang 2004: 224). Thus the way of assessing the process of epistemic iteration is relatively to the various epistemic values that are

indicative of progress. The traditional ones are simplicity, support by more general theories, ability to predict previously unknown phenomena, credibility relative to background knowledge, consistency, scope, etc. (Chang 2004: 227).

Interestingly, this process appears to be very similar to Lipton's characterization of inference to the best explanation. The process of 'causal triangulation' plays a central role in refining the inference to the best explanation could accurately be described as an iterative process: it consists in coming up with 'foils', situations that are very similarly but only slightly different than the 'facts' under study, and gradually eliminating putative causes that are in the shared part of antecedents of fact and foil, as what explains that the fact happened rather than the foil must be in the part of antecedents that is different between both (Lipton 2004: 73). Lipton also identifies the same problem as epistemic iteration with the circularity of defining the 'best' explanation as the true one (the 'actual best explanation'): the whole point of the inference is to give us a way of identifying the 'symptoms' of truth when we can't know it. If the best explanation is the true one, then the inference to the best explanation is trivially valid, because it cannot be assessed by comparison with the truth. As Lipton puts it: "according to inference to the best explanation then, we do not infer the best actual explanation; rather we infer that the best of the available potential explanations is an actual explanation" (Lipton 2004: 58). And that inference can in principle be shown to be invalid if it turns out that what defines the best available explanation does not lead to the actual best explanation. So the way to assess which explanation is the best for Lipton also turns out to be through epistemic values, although this time restricted to the explanatory ones (Lipton 2004: 121). Hence for future research, it might be worth pursuing a way out of underdetermination along the lines of an IBE rather than epistemic iteration.

But for now, I propose a minimalist framework of normative principles that render possible the process of epistemic iteration for the study of sentience:

- **comparability** between different species[22]

This condition is necessary to even start the research: you have to assume that sentience will have identical or similar enough physical conditions across species to allow comparisons between them. If you accept that there is no way to compare fish pain (if it exists) and human pain, then you can't even start arguing for the existence of pain in fish. This is an implicit assumption in the very

[22] The principle of comparability is inspired by Chang (2004: 89), but with a slightly different definition, since Chang is dealing with the goal of *quantifying* temperature whereas here the goal is only to *detect* sentience.

practice of scientists: as soon as they start coming up with criteria that allow trans-species comparisons, such as the list of seventeen criteria from (Sneddon et alia. 2014), they presume that comparability is possible.

Making the assumption of comparability doesn't mean that it will necessarily turn out to actually be the case. The assumption can be disproved by the research based on the assumption: this is an advantage of coherentism that even the temporary premisses can be modified if they are not consistent with the rest of the net of beliefs. That offsets the pessimistic claim that there is no way to distinguish between plausible and implausible cases of multiple realizability: there is a way but it isn't based on any prior deductive argument but rather on what is allowed by the empirical research and what isn't. The criterion becomes whether comparability can be made to work at all in the case under study. But this alone would give too much leeway and allow for almost any species to be declared comparable if it wasn't combined with the second principle.

- **imperative of coherence** between the procedures and criteria used to detect sentience

If a procedure or a set of criteria is considered a good enough approximation to detect sentience in a context (for a species in particular), then it should also be considered good enough in another context even if the prior assumptions about sentience are different. So for example even if we presume more easily that babies are sentient than fish are, the procedures and criteria considered sufficient to accept sentience in babies should also be accepted for fish. This comes back to the recognition that detection procedures for sentience don't have to be perfect, but the bar shouldn't be set higher for non-humans than fore some humans unless there is a good reason: "If we can determine what infants and humans with disorders of consciousness feel in some cases, then we have, in principle, methods to make justified claims about animal experiences. To defend the claim that comparative psychology should not involve itself with animal consciousness requires showing that there is a relevant difference in kind between the sciences of psychology or anthropology when they are focused on human subjects and when they are focused on animal subjects. Language cannot be the difference, because while language is not present in humans with disorders of consciousness and prelinguistic infants, scientists have managed to study and treat humans who don't use language" (Andrews 2020: 27).

- **a narrow (as opposed to wide) multiple realizability**

I think scientists in the debate agree that comparisons don't have to be absolutely pristine, and that some sort of multiple realizability is possible. This is almost an implication of the very

definition of comparability: in order for two things to be compared, they have to be at least slightly different. Even the staunchest detractors of fish pain accept that multiple realizability is highly likely at some level of analysis (Brown & Key 2021: 10), because otherwise we wouldn't be able to claim that we share psychological experiences with other species at all. So multiple realizability needs to be thought of at the right level of analysis in order not to be trivialized. As Brown and Key put it: "When either the structural level is too granular or the functional level too indiscriminate, multiple realizability is assured—and its argumentative value is exhausted" (Brown & Key 2021: 11).

But there are demonstrated cases of multiple realization which do not look trivial: for example, Lynne Sneddon has pointed out that in the nervous system of trouts, A-delta fibers act in a similar way as mammalian C fibers, reacting to a variety of noxious stimuli (Sneddon 2003a). Which level of analysis is most relevant to assess multiple realizability is thus on a continuum, and if you can always stretch it by considering very implausible cases like rock pain for instance. However, the reason such cases are implausible is because we lose comparability and/or coherence, and you could only defend multiple realization by overlooking the evidence of the conditions of pain in humans and mammals more generally. That means that, contra Michel, we do have a way to distinguish between plausible and implausible cases of multiple realization. Hence, I propose to use comparability and coherence as criteria to define an acceptable narrow multiple realizability:

- narrow multiple realization: plausible cases that respect the principles of comparability and coherence
- wide multiple realization: implausible cases -but not impossible- where you cannot apply those principles and thus lose the possibility of using epistemic iteration to assess them

It should be noted here that I am not excluding at all the possibility of defending wide multiple realization, but only arguing that doing so would imply resorting to a different process than epistemic iteration. This actually brings out a limit of this process, since epistemic iteration could not for instance detect 'Martian pain' (Lewis 1980: 229), that is pain which would have completely different physical conditions; but I don't think this is as much of a problem for pragmatic scientists as it is for philosophers[23]. A more disturbing fact is that epistemic iteration may not be suited to deal with cases of alleged convergent evolution either. If one wants to argue that fish developed the function of pain not through relevantly similar structures with mammals, but rather with different structures that converged nonetheless towards developing the function of pain (for instance Segner

[23] Although it might be a problem for scientists thinking about issues such as the potential sentience of artificial intelligence, but not for the debates surrounding animal consciousness since animals are the product of evolution.

2016: 2), then there might not be enough comparability to deal with it through the process of epistemic iteration. A more fully-developed functional analysis will probably prove better suited for dealing with such cases. But comparability being relative to empirical findings -since observability is reached by stages in a coherentist framework-, the scope of comparability may very well be broadened in the future, and it is not to exclude that what counts as a narrow multiple realization may eventually be able to deal with cases of relatively close convergent evolution.

At any rate, within the framework I propose it is very clear that relying on multiple realization is not equivalent to saying that any structure could produce pain, which thwarts Michel's argument.

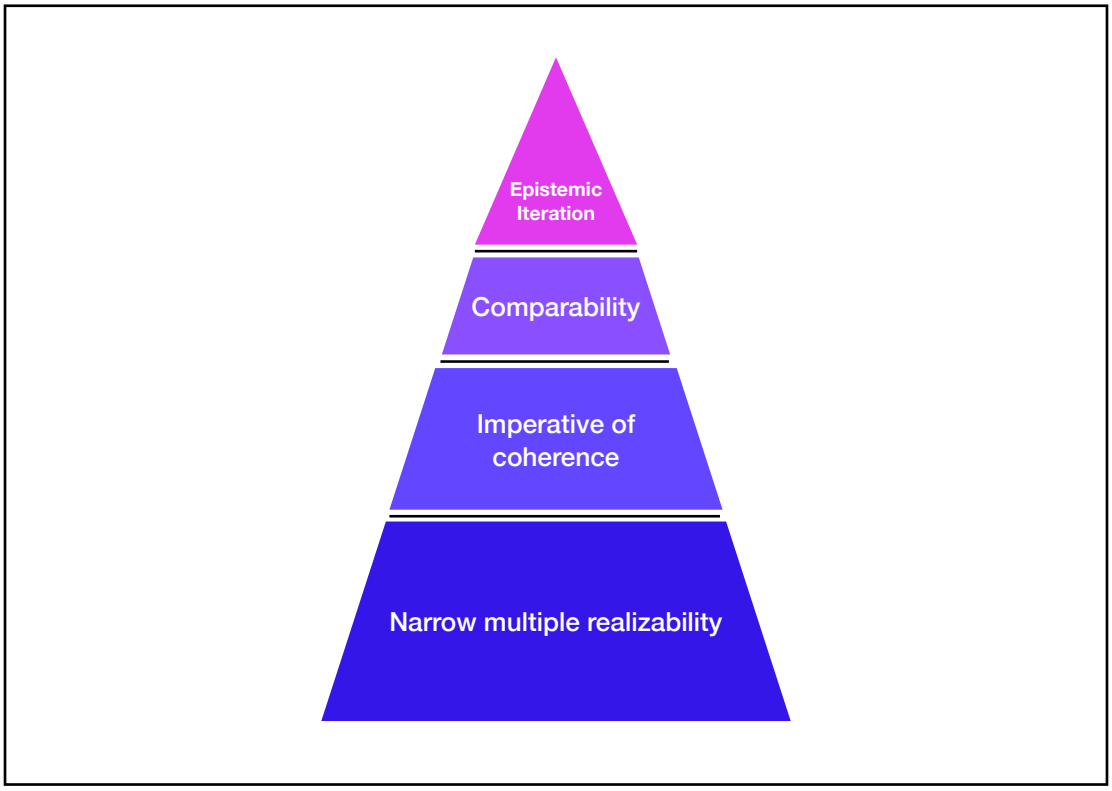


Diagram of the minimal framework of principles to enable epistemic iteration

Although the principle of coherence is rather consensual, one may object that this framework begs the question by simply assuming comparability without justifying it. It is indeed a pragmatic assumption that scientists need to do in order to apply this particular epistemic process, in the same way that it was already a pragmatic assumption to treat sentience as an unobservable property; this is a move which coherentism allows to do much more easily than foundationalism after all. We can adopt temporary assumptions that may very well have to be corrected or replaced if they are no longer consistent with the empirical findings that their adoption allowed to dig. What is more, as far as pragmatic assumptions go, presuming that you can find homologies across species

is not the most contentious one, particularly because it is already broadly accepted for mammals and birds (in that sense it is almost a call for consistency).

B. From a fixed list of criteria to a dynamic process of epistemic iteration

The consequence of narrow multiple realizability is that scientists have to accept that some manifestations of pain in humans may not be present in fish even if they feel pain (different nocifensive behaviour for example). So the attribution of sentience might be accepted even in the absence of a limited number of criteria that are deemed necessary for humans, even if the presence of those criteria would be interpreted as a good sign of pain since they are sufficient for pain in humans. There is thus a dissymmetry between the sufficient and the necessary aspect of each condition which results from the combination of narrow multiple realizability and the imperative of coherence.

This entails that the possibility of establishing a final list of criteria to detect sentience should probably be doubted, at least until sentience is fully operationalized, which we have seen is far from being the case at the moment. The set of conditions is a useful way of conceptualizing the epistemological difficulties of the detecting procedures of sentience and to outline the situation of underdetermination, but it is not the most accurate way of depicting how scientists reason in a coherentist framework. The best way to think about the set of conditions is not as a systematic procedure to detect sentience, but rather as a sort of improved[24] "Turing test" (Woodruff 2018: 12), useful for the development of a common framework to evaluate pain systematically across species boundaries, by striving to reach more comparability.

This new way of looking at the set of criteria explains why the interpretation of the empirical results are never straightforward and spur many debates about whether a given piece of evidence supports an attribution of sentience. The set is more of a complex net of intertwined signs that are interpretable in the light of other experiments and previous conclusions that have been made. To give but one example, the behavioural elements that are in the set of observed conditions in fish would be interpreted very differently if the phenomenon of blindsight had never been observed; blindsight doesn't completely discard the strength of observing nocifensive behaviour in an animal but it does severely weaken the inference from behaviour to sentience in the absence of other conditions that would support it.

[24] Improved because it doesn't rely exclusively on behavioural responses contrary to the Turing test.

The way epistemic iteration works is by starting with the (however thin) evidential base we have and gradually extend it by relying on the principles presented above. For the case of sentience, the initial evidential base would be human reports of their conscious states, or their pain. There are then two modes of progress:

- **enrichment**, in which the initially affirmed system is not negated but refined, resulting in the enhancement of some of its epistemic virtues
- **self-correction**, in which the initially affirmed system is actually altered in its content as a result of inquiry based on itself (Chang 2004: 228)

In the context of detection of sentience, I take enrichment to be the discovery in animals that are not presumed sentient of new conditions that are considered as indicative of sentience, or the refinement of those conditions for animals who are already considered sentient (even with no certainty). This allows to better and better identify conditions of sentience until the attribution can be made with some level of confidence. I've already given a few examples of discoveries of conditions for sentience in fish, such as nociceptors or willingness to pay a cost. An example of a refinement of the already accepted conditions would be for instance the discovery that different submodalites of pain, such as sharp-pricking pain and dull-burning pain, are mapped in different subregions of the singular cortex (Key 2016b: 12), which could allow for more precise observations in non-human animals.

Self-corrections would consist in the modification of the accepted conditions of sentience, or a reinterpretation of their relative importance, because of new observations either in sentient beings or in animals that are not yet accepted as sentient, by application of the imperative of coherence. One of the main types of self-corrections is when a condition that was deemed to be necessary for phenomenal consciousness is shown not to be. This is the case for instance in 'no report' paradigms where regions in human brains that were thought to be associated with conscious states turn out not to be activated when the participants are not required to report those states (Seth 2016: 2); that result is very striking because those conditions were part of the initial evidential base, the one which was supposed to be the most grounded and justified, so it has important consequences on the way we try to detect consciousness in non human animals.

A very eloquent manifestation of that self-correction comes from a recent change of definition of the very concept 'pain' from the International Association for the Study of Pain. As their website reads:

> A central change in the new definition, compared to the 1979 version, is replacing terminology that relied upon a person's ability to describe the experience to qualify as pain. The old definition read: "An unpleasant sensory and emotional experience associated with actual or potential tissue damage, or described in terms of such damage." This wording was interpreted as excluding infants, elderly people, and others – even animals -- who could not verbally articulate their pain, said Dr. Jeffrey Mogil, Director of the Alan Edwards Center for Research on Pain, McGill University and member of the Task Force. (Terminology | International Association for the Study of Pain)

One could thus speculate that it was the combination of 'no report' paradigms and of the growing consensus on the attribution of sentience for humans who can't express their conscious states and non human animals such as mammals and birds that prompted them to change the definition by dropping a condition that was no longer deemed necessary.

Another particularly stimulating example of self-correction coming this time from the study of non human animals is the conception of self-awareness[25]: the fact that a fish passed the mirror test (Kohda et alia. 2019) is very unexpected and has repercussions on the way we have been thinking about it until now: either one has to accept that fish are self-conscious, or one has to reassess the inference from passing the test to being self-conscious if one is not willing to bite the bullet that fish are self-conscious. A third way might be to argue that the observed behaviour of fish in front of the mirror should not be counted as passing the test, but that involves reassessing the interpretation of all the previous mark tests when those types of behaviour were deemed sufficient to pass (Kohda et alia. 2019: 1). That could also encourage ethologists to devise tests that are more adapted to the species: the fact that a fish may pass the test and not dogs calls for the reassessment of how the test was devised for dogs -if we want to cling to the intuition that dogs have higher cognitive and conscious capacities than fish. Maybe a better test for them should involve an olfactory mark, since dogs which are animals relying much more on their sense of smell than their vision.

Either way, this result is very well-suited to show how the study of fish can force to reconsider a great variety of formerly established conclusions and methods, and eventually foster progress throughout the studies of animal consciousness. This shows how powerful epistemic iteration, through the combinaison of comparability and the imperative of coherence, can be in eliciting progress in science, by gradually adding constraints on what can consistently be claimed.

[25] Keeping in mind that self-awareness is distinct from phenomenal consciousness, I still present this example because it is very revealing of the power of self-correction. There is after all no reason that epistemic iteration wouldn't also be suited to deal with other non phenomenal aspects of consciousness, even if it isn't the focus of this work.

C. The resulting process of semantic extension

As was established in part I, sentience is not a concept that is either fully defined nor operationalized. But it turns out not to hinder empirical progress, because the meaning of sentience is broader than its operational meaning (it it had one). This goes against what Chang calls Bridgman's reductive doctrine of meaning which postulates that the meaning of a concept is completely reducible to its methods of measurement. But in fact, if the meaning of sentience had been completely reducible to how it could successfully be measured, so only in humans in the first place, there would have been no possibility to extend the concept to other animals[26]. Indeed there would be no 'leeway' left since what would define the concept would be nothing else than the limited methods of measurement within a limited domain of application. If that doctrine of meaning was really to be followed to its consequences, you would technically create a new concept each time you try to extend it, because the way of measuring it would be different (Chang 2004: 150). But the fact is that we can understand the meaning of sentience without needing to measure it because the meaning is broader, and hence we have a way of determining whether an extension of the concept still corresponds to its meaning. This is why we can imagine the possibility of 'fish pain' even if we have no way to measure it directly, because we know that what we mean when we say pain is something relevantly similar to what humans feel in certain situations, and we can call it fish pain even if we suppose that it would be qualitatively different from human pain. And it probably is if it exists[27]; after all we do talk of 'vision' for cats even though we know that it must feel drastically different from human vision, but that difference does not justify creating a new concept for each species.

Discussing whether sentience should be attributed to fish boils down to determining whether the concept of sentience can legitimately be extended to fish; so the application of the process of epistemic iteration can lead to a successful semantic extension. The way semantic extension of a concept works is that we start with a secure net of uses giving it stable meaning in a restricted domain of circumstances. The extension then consists in giving it a secure net of uses credibly

[26] Another drawback of that conception is that if the measurement method defines the concept, there is no point is asking whether the measurement method is valid (Chang 2004: 151).

[27] This goes back to the distinction between the distribution problem and the epistemological problem.

linked to the earlier net, in an adjacent domain (Chang 2004: 150). The concept of sentience is strongly grounded in its ethical use, which sets clear limits to the extension of the concept. The reason the distinction between nociception and pain is so important is that it makes a morally relevant difference: we think it's bad to be in pain, and wrong to inflict pain on someone, which is not prima facie the case for mere nociception. So we know what we mean by sentience because in a sense it means 'the thing that makes you able to feel good things and bad things' in an ethical context, even when we don't know how to detect it. And the importance of that ethical net of uses that gives stability to the concept is shown by the fact that the extension of sentience leads to hotly debated topics in bioethics, such as animal ethics or the question of the moral status of fetuses, or anencephalic babies for instance[28]. The extension of sentience is thus grounded in the relevance of sentience in ethical debates.

Now there are multiple ways of extending a concept. Theoretical semantic extension is the most unifying and ambitious one, but it requires a strong background theory for the concept, which is still lacking in the case of sentience. We can thus resort to operational extensions, which correspond to an extension based on the physical operations that detect it. The validity of the extension is then assessed relatively to the other aspects of the concept's meaning (because again the meaning of the concept is broader than its operational meaning). In particular, we can check for two conditions to be satisfied:

- **Conformity:** If the concept possesses any pre-existing meaning in the new domain, the new standard should conform to that meaning
- **Overlap:** If the original standard and the new standard have an overlapping domain of application, they should yield measurement results that are consistent with each other (Chang 2004: 152).

So the extension should stay consistent with the uses of sentience in the human context, and in particular with ethical uses, and the condition of overlap corresponds to the application of the imperative of coherence between comparable criteria and procedures across species in the process of epistemic iteration.

An interesting fact is that the semantic extension of one aspect of sentience is in principle sufficient to extend the concept of sentience altogether. We've been focusing a lot on pain in this

[28] On these bioethical issues, see (Holland 2016).

work because scientists themselves have been very much focused on it, but the successful extension of any other dimension of sentience would allow the attribution of sentience, which is probably a good reason to be optimistic about the future progress of the study of sentience. Hence why many welfare scientists have been calling for an increase in the empirical research on positive aspects of phenomenal consciousness as well as negative ones (Duncan 2006: 16; Proctor 2012: 636).

3. Application of epistemic iteration to the study of fish sentience

Some examples of enrichment and self-corrections have already been given in the previous part, but we can try to draw a general image of how exactly epistemic iteration can be applied to the study of fish sentience. First, one important thing to acknowledge is that the current debate depends very much on previous iterations of the iterative process as a whole, that is the process at the scale of animal sentience in general. One way of retracing some of the iterative steps for sentience -as a heuristic to understand how epistemic iteration can be applied and not at all as an accurate historical picture of the sequence of empirical discoveries- could be the following overly-simplified sequence:

- the conditions for pain are slowly refined in humans by examining the consequences of lesions or manipulations of certain areas of the brain through experiments relying on reports (enrichment through refinement of the conditions)
- these conditions are extrapolated to be compared with observations in patients who can't report their pain (enrichment through semantic extension)
- the conditions in mammals are established by comparison with the criteria accepted for human patients who can't report pain (enrichment through semantic extension)
- certain complex behavioural adaptive responses to pain are considered as good indirect signs of pain in mammals, such as self-administration of analgesics and willingness to pay a cost in order to gain access to analgesics (enrichment through refinement of the conditions)[29]
- the same behavioural responses are observed in birds, along with other conditions that are considered good signs of pain for mammals (enrichment through semantic extension)[30]
- the discovery of a peripheral system (which isn't connected to the brain) that monitors behaviour relative to the responses to noxious stimuli is found in frogs, which either raises the bar of the complexity of the behaviour that has to be shown in order to prove that it is not just a product of

[29] (Gao et alia. 2004)

[30] (Danbury et alia. 1997)

this peripheral system or calls for the combination of the behavioural signs with an observed modification in the central nervous system (self-correction)[31]

This process of epistemic iteration leads to a never-ending sequence of refinements and self-corrections which gradually allow for more comparability with fish and then potentially other animals, while allowing a gradual semantic extension of the concept of sentience and an improvement in the observability of sentience. More specifically, some arguments in favor of fish pain are made possible only thanks to previous enrichments or self-corrections from studies on the sentience of other animals than fish. For instance, the consensus that birds can realize pain even though they have a brain quite different from mammals gives more credence to the argument that brain regions in fish might also be analogous to the regions in human brains that realize pain (Tye 2019: 125). So in broad strokes, if birds are assumed to be sentient, scientists conclude that a narrow multiple realizability is possible at that level, and then comparability with fish brains combined with the imperative of coherence get you closer towards the conclusion that fish are also sentient. Application of the principle of comparability explains why sentience in mammals would be recognized before that of birds and fish, but once the conditions are refined or corrected, new comparisons can arise with animals that are more distant from humans.

A good sign that the process of epistemic iteration is well-suited to analyse the debate about fish sentience[32] is that at least some scientists seem to not only accept the principle of comparability, coherence and narrow multiple realizability, but also to actually use arguments from consistency that rely on the previous iterative steps. For example, Lynne Sneddon and Matthew Leach rebut some critics by stressing that they refuse to accept criteria for fish that would be considered convincing for birds or even for mammals:

> Rose et al. have been criticised for not stating that on their account birds, too, could not possibly experience pain — in fact, no animal other than primates could — because birds, like fish, have only a singly laminated cortex. [...] Prolonged behavioural responses to painful stimulation in fish can be seen for at least hours and even days later. Furthermore, providing analgesia substantially reduces these adverse changes in behaviour in fish: this same evidence would be accepted as evidence of pain in mammals. (Sneddon & Leach 2016: 1-2)

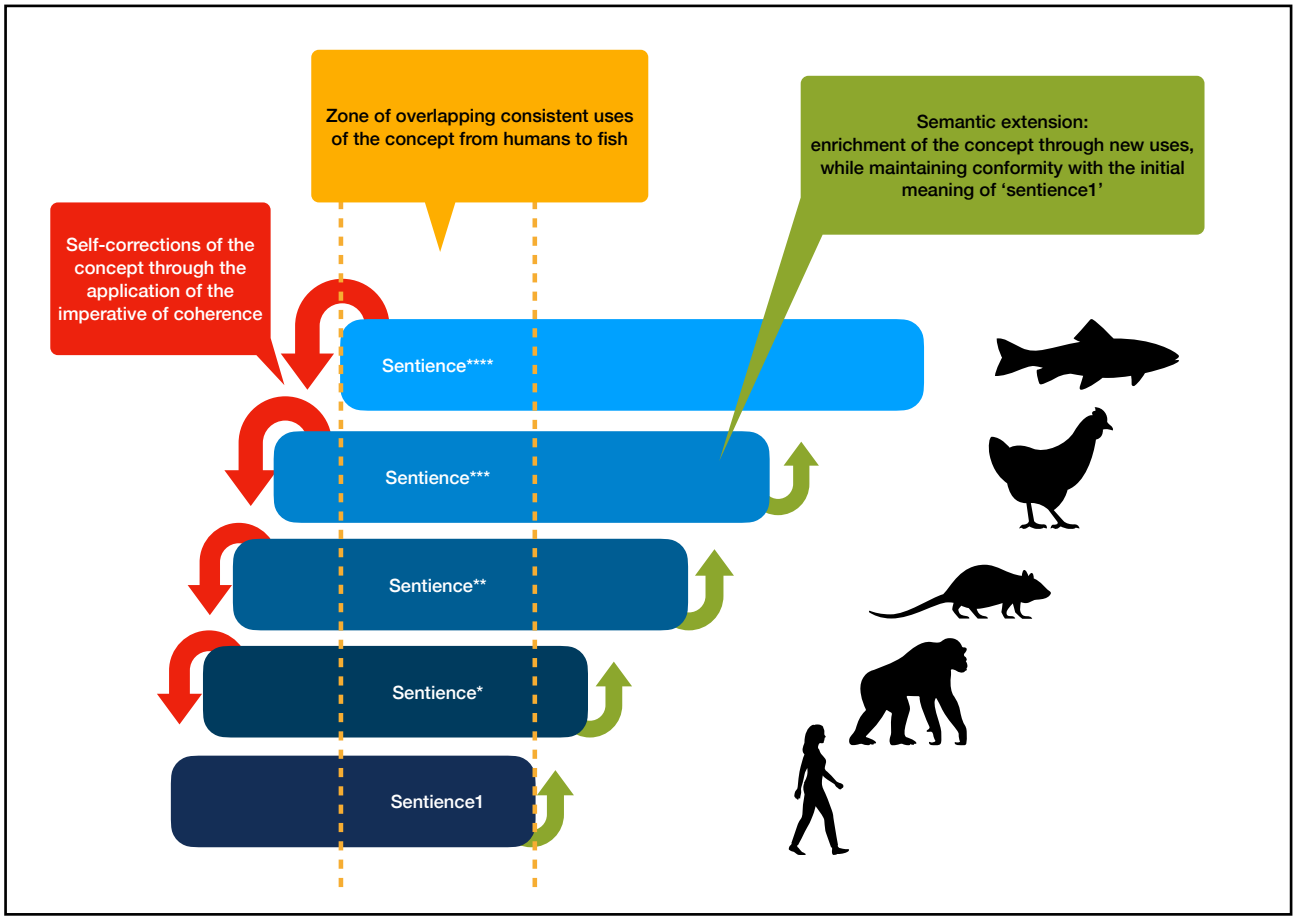
And again in the same article:

[31] (Klein 2017)

[32] For overviews of the empirical findings on fish, see Braithwaite (2010) on the specific question of pain, Balcombe (2016) for the both cognition and consciousness) and chapter 6 of Tye (2019).

> Key also argues that although we do not fully understand how birds process pain, he is willing to accept that pain can occur in birds; however, because we do not know enough about how fish process pain, they cannot experience pain. (Sneddon & Leach 2016: 2)

Here is a schematic which sums up the various elements of the application of epistemic iteration to the study of fish sentience, and the way it both simultaneously depends on and feeds back into the study of sentience in other animals:



Schematic of the application of epistemic iteration to the study of fish sentience

Here are some remarks on the limits of this schematic though. First, there could well be consistent uses of sentience between fish and humans which are not shared by birds, so it's only out of convenience and pedagogy that the categories are arranged that way, although the arrangement is meant to make some sense in terms of mutual comparability. Second, fish are actually not a proper taxon in the same way that mammals are: "It's a folk-biological term that does not correspond precisely to any monophyletic taxonomic group. This can be appreciated by noting that a coelacanth is more closely related to a human than to a tuna, or that a tuna is more closely related to a human

than it is to a shark." (Allen & Trestman 2020: section 6.5). This has important consequences on the directions in which the enrichments and self-corrections can occur, since the leading factor is comparability; and each of the experiments that have been referred to in my work were usually about a specific species of fish. So the conclusion that 'fish feel pain' may be meaningless at such a level of generality and it might very well be proven for certain types of fish and not for others. For now, the debate has focused on whether *at least one fish* could feel pain, but if and once that becomes more consensual, as that seems to be the case, the research question of the debate will probably shift towards looking for *which specific fish do feel pain*.

Conclusion

I isolated the underdetermination problem for sentience from two others problems, the hard problem and the skeptical challenge of other minds, showing that scientists should only or mainly concern themselves with the underdetermination one. I showed that arguments which don't allow for fallibilism should be discarded for pragmatic reasons because they lead to stalemates in the scientific debate, and that the reason for that is their underlying foundationalist framework which is not suited to allow for empirical progress in the debate on fish sentience. I defended instead that a coherentist framework is better suited for it, and that within that framework epistemic iteration is a promising system to tackle the situation of underdetermination, and foster empirical progress.

In the end, the possibility of an emerging consensus on fish sentience without a meta-consensus on a fully worked-out detection procedure seems possible by application of the process of epistemic iteration, which relies on rather minimal principles on which most pragmatic scientists should be able to agree. It allows for a certain level of disagreement on the criteria for sentience while also possibly leading to gradual progress by making some claims more and more costly. The arguments assessed in part II can still have heuristic value or even some degree of justificatory value, but they are only compatible with epistemic iteration insofar as they allow for comparability, coherence and narrow multiple realizability. The arguments looked inconclusive taken in a foundationalist framework, but I showed that if used within a coherentist framework, they might actually lead to a slowly emerging consensus.

Eventually, it can be believed that the building up of new evidence and the increase in comparability across species will force the emergence of a consensus, simply because it might become impossible to negate that fish are sentient without having to deny statements that are well-established in the scientific community, and very few will be willing to pay that cost. However, although it seems that the actual situation is converging towards that state for fish sentience because we can observe progress in the conditions, a consensus is made possible but never guaranteed by epistemic iteration. Indeed, it still depends on empirical discoveries and the process might get stuck for some reason. A long-term blockade of the system might mean that one of its working principles should be revised or that the system fails altogether (Chang 2004: 227). Not to mention that the consequence of accepting coherentism is that no step is ever considered final, and the process of epistemic iteration in principle never ends. It thus only gives us a potential way out of underdetermination.

I see three avenues which the scientific study of sentience could pursue in order to move forward in the future:

- more **epistemic iteration** through the integration of insects in the scope of comparability (Klein & Barron 2016); although there seems to be a widely shared ontological assumption that sentience is a binary property (either you are sentient or not), another advantage of epistemic iteration might be to open avenues for a continuous vision of sentience, where minimal levels of phenomenal consciousness might be detected in certain creatures such as insects.
- a precise **quantification**[33] of sentience in humans and the development of a precise scale that would be common to humans and (at least some) non-human sentient animals. For work in that direction, see Heather Browning's doctoral thesis (2019).
- a better **theoretical grounding** that would allow for a full-fledged functional analysis and its operationalization. Semantic extensions of sentience would then become theoretical ones.

[33] As opposed to mere detection, which was the focus of this work.

Acknowledgments

I am very thankful to my supervisor Professor Anouk Barberousse for believing in me and supporting me during my two years of master's degree, for her demanding and stimulating teachings as well as for the balance of autonomy and encouragement she has given me whenever I needed one or the other.

I am also grateful to Doctor Lynne Sneddon for her useful insight on some of the issues raised in this dissertation.

I thank Sophie Evers for sending me her master's dissertation.

I have greatly benefitted from the teachings of the professors at the University of York where I spent my second year of master's as an Erasmus student, particularly from Professor Stephen Holland's module on bioethics which inspired me the topic of this dissertation.

References

- Allen, C. (2004). “Animal Pain”. *Noûs*, *38*(4), 617–643.
- Allen, C., & Trestman, M. (2020). “Animal Consciousness”. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2020). Metaphysics Research Lab, Stanford University.
- Andrews, K. (2011). “Beyond Anthropomorphism: Attributing Psychological Properties to Animals”. In T. L. B. R. G. Frey (Ed.), *Oxford Handbook of Animal Ethics* (pp. 469–494). Oxford University Press.
- Andrews, K. (2020). *How to Study Animal Minds*. Cambridge University Press.
- Arbilly, M., & Lotem, A. (2017). “Constructive anthropomorphism: A functional evolutionary approach to the study of human-like cognitive mechanisms in animals”. *Proceedings of the Royal Society B: Biological Sciences*, *284*(1865), 20171616.
- Balcombe, J. (2016). *What a Fish Knows: The Inner Lives of Our Underwater Cousins*. Oneworld Publications.
- Bartha, P. (2019). “Analogy and Analogical Reasoning”. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2019). Metaphysics Research Lab, Stanford University.
- Beck, J. (2017). *The Routledge Handbook of Philosophy of Animal Minds*. Routledge.
- Bekoff, M. (2000). “Animal Emotions: Exploring Passionate Natures: Current interdisciplinary research provides compelling evidence that many animals experience such emotions as joy, fear, love, despair, and grief—we are not alone”. *BioScience*, *50*(10), 861–870.
- Bentham, J. (1823). *Introduction to the principles of morals and legislation*. Clarendon Press.
- Birch, J. (2017). “Animal Sentience and the Precautionary Principle”. *Animal Sentience*, *2*, 16–1.
- Birch, J., Schnell, A. K., & Clayton, N. S. (2020). “Dimensions of Animal Consciousness”. *Trends in Cognitive Sciences*, *24*(10), 789–801.
- Black, D. (2020). “Analyzing the Etiological Functions of Consciousness”. *Phenomenology and the Cognitive Sciences*, *1*, 1–26.
- Block, N. (2009). “Comparing the major theories of consciousness”. In *The cognitive neurosciences, 4th ed* (pp. 1111–1122). Massachusetts Institute of Technology.
- Braithwaite, V. (2010). *Do Fish Feel Pain?* OUP Oxford.

- Broom, D. M. (2016). “Sentience and animal welfare: New thoughts and controversies”. *Animal Sentience*, *1*, 2016.057.
- Broom, D. M. (2011). “A History of Animal Welfare Science”. *Acta Biotheoretica*, *59*(2), 121–137.
- Broom, D. M. (2014). *Sentience and Animal Welfare*. CABI.
- Brown, C. (2015). “Fish intelligence, sentience and ethics”. *Animal Cognition*, *18*(1), 1–17.
- Brown, D. J., & Key, B. (2021). “Is absence of evidence of pain ever evidence of absence?” *Synthese*.
- Browning, H. (2019). *If I Could Talk to the Animals: Measuring Subjective Animal Welfare* [Thesis, The Australian National University].
- Burghardt, G. M. (1991). Cognitive ethology and critical anthropomorphism: A snake with two heads and hognose snakes that play dead. In C. A. Ristau (Ed.), *Cognitive ethology: The minds of other animals: Essays in honor of Donald R. Griffin* (pp. 53–90). Lawrence Erlbaum Associates.
- Burghardt, G. M. (2007). “Critical anthropomorphism, uncritical anthropocentrism, and naïve nominalism”. *Comparative Cognition & Behavior Reviews*, *2*(1), 136–138.
- Carruthers, G., Carls-Diamante, S., Huang, L., Rosen, M., & Schier, E. (2019). “How to operationalise consciousness”. *Australian Journal of Psychology*, *71*(4), 390–410.
- Carruthers, P. (1989). “Brute Experience”. *Journal of Philosophy*, *86*(May), 258–269.
- Carruthers, P. (1992). *The Animals Issue: Moral Theory in Practice*. Cambridge University Press.
- Chalmers, D. (1995). “Facing Up to the Problem of Consciousness”. *Journal of Consciousness Studies*, *2*(3), 200–219.
- Chang, H. (2004). *Inventing Temperature: Measurement and Scientific Progress*. Oup Usa.
- Chang, H. (2012). *Is Water H2O?: Evidence, Realism and Pluralism*. Springer Netherlands.
- Cochrane, A. (2018). *Sentientist Politics: A Theory of Global Inter-Species Justice*. Oxford University Press.
- Danbury, T. C., Chambers, J. P., Weeks, C. A., Waterman, A. E., & Kestin, S. C. (1997). “Self-selection of analgesic drugs by broiler chickens”. *BSAP Occasional Publication*, 20, 126–128.
- Dawkins, M. S. (1980). *Animal Suffering: The Science of Animal Welfare*. Chapman and Hall.
- Dawkins, M. S. (2015). “Animal Welfare and the Paradox of Animal Consciousness”. In M. Naguib, H. J. Brockmann, J. C. Mitani, L. W. Simmons, L. Barrett, S. Healy, & P. J. B. Slater (Eds.), *Advances in the Study of Behavior* (Vol. 47, pp. 5–38). Academic Press.
- Dawkins, M. S. (2001). “Who Needs Consciousness?” *Animal Welfare Supplement*, *10*, 19–29.

- Dawkins, M. S. (2006). "The Scientific Basis for Assessing Suffering in Animals". In P. Singer (Ed.), *In defense of animals: The second wave* (pp. 26–39). Blackwell Publishing.
- Dawkins, M. S. (2008). "The Science of Animal Suffering". *Ethology*, *114*(10), 937–945.
- DeGrazia, D. (1997). "Great Apes, Dolphins, and the Concept of Personhood". *The Southern Journal of Philosophy*, *35*(3), 301–320.
- DeGrazia, D. (2020). "Sentience and Consciousness as Bases for Attributing Interests and Moral Status: Considering the Evidence and Speculating Slightly Beyond". In L. S. M. Johnson, A. Fenton, & A. Shriver (Eds.), *Neuroethics and Nonhuman Animals* (pp. 17–31). Springer International Publishing.
- Donovan, J., & Adams, C. J. (Eds.). (1996). *Beyond Animal Rights: A Feminist Caring Ethic for the Treatment of Animals*. Continuum Intl Pub Group.
- Duncan, I. J. H. (2004). "A Concept of Welfare Based on Feelings". In G. J. Benson, & B. E. Rollin (Eds.), *The Well-Being of Farm Animals: Challenges and Solutions* (Blackwell, pp. 85–101).
- Duncan, I. J. H. (2006). "The changing concept of animal sentience". *Applied Animal Behaviour Science*, *100*(1), 11–19.
- Evers, S. (2019). *Behavioural studies of humans and non-human primates: Bidirectional influences* [Master's dissertation]. Sorbonne Université / KU Leuven.
- Fidler, F., & Wilcox, J. (2021). "Reproducibility of Scientific Results". In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2021).
- Fischer, B. (2019). *The Routledge Handbook of Animal Ethics*. Routledge.
- Fitzpatrick, S. (2008). "Doing Away with Morgan's Canon". *Mind and Language*, *23*(2), 224–246.
- Fraser, D., & Nicol, C. J. (2011). "Preference and motivation research". In M. C. Appleby, J. A. Mench, I. A. S. Olsson, & B. O. Hughes (Eds.), *Animal welfare* (2nd ed., pp. 183–199). CABI.
- Gao, Y.-J., Ren, W.-H., Zhang, Y.-Q., & Zhao, Z.-Q. (2004). "Contributions of the anterior cingulate cortex and amygdala to pain- and fear-conditioned place avoidance in rats". *Pain*, *110*(1), 343–353.
- Griffin, D. R. (2001). *Animal Minds – Beyond Cognition to Consciousness*. University of Chicago Press.
- Griffin, D. R. (1976). *The question of animal awareness: Evolutionary continuity of mental experience*. Rockefeller University Press.

- Hare, B., Call, J., & Tomasello, M. (2001). “Do chimpanzees know what conspecifics know?” *Animal Behaviour*, *61*(1), 139–151.
- Harnad, S. (2016). “Animal sentience: The other-minds problem”. *Animal Sentience*, *1*(1).
- Harrison, P. (1991). “Do Animals Feel Pain?” *Philosophy*, *66*(255), 25–40.
- Harrison, R. (1964). *Animal Machines*. Vincent Stuart.
- Holland, S. (2016). *Bioethics: A Philosophical Introduction* (2nd edition). Polity Press.
- Humphrey, N. (1974). “Vision in a Monkey Without Striate Cortex: A Case Study”. *Perception*, *3*, 241–255.
- Hursthouse, R. (2011). “Virtue Ethics and the Treatment of Animals”. In T. Beauchamp & R. G. Frey (Eds.), *The Oxford Handbook of Animal Ethics* (pp. 119–143). Oxford university Press.
- Kagan, S. (2019). *How to Count Animals, More or Less*. Oxford University Press.
- Key, B. (2015). “Fish Do Not Feel Pain and its Implications for Understanding Phenomenal Consciousness”. *Biology and Philosophy*, *30*(2), 149–165.
- Key, B. (2016a). “Why fish do not feel pain”. *Animal Sentience*, *1*(3(1)).
- Key, B. (2016b). “Falsifying the null hypothesis that “fish do not feel pain””. *Animal Sentience*, *1*(3(39)).
- Klein, A. (2017). “The curious case of the decapitated frog: On experiment and philosophy”. *British Journal for the History of Philosophy*, *26*(5), 890–917.
- Klein, C., & Barron, A. (2016). “Insects have the capacity for subjective experience”. *Animal Sentience*, *1*(9).
- Kohda, M., Hotta, T., Takeyama, T., Awata, S., Tanaka, H., Asai, J., & Jordan, A. L. (2019). “If a fish can pass the mark test, what are the implications for consciousness and self-awareness testing in animals?” *PLOS Biology*, *17*(2), e3000021.
- Leng, M. (2010). *Mathematics and Reality*. OUP Oxford.
- Lewis, D. (1980). “Mad Pain and Martian Pain”. In N. Block (Ed.), *Readings in the Philosophy of Psychology* (pp. 216–222). Harvard University Press.
- Lipton, P. (2004). *Inference to the Best Explanation* (2nd edition). Routledge.
- Maxwell, G. (1962). “The Ontological Status of Theoretical Entities”. In H. Feigl & G. Maxwell (Eds.), *Scientific Explanation, Space, and Time: Minnesota Studies in the Philosophy of Science* (pp. 181–192). University of Minnesota Press.
- Michel, M. (2019). “Fish and Microchips: On Fish Pain and Multiple Realization”. *Philosophical Studies*, *176*(9), 2411–2428.

- Michel, M. (2020). “Consciousness Science Underdetermined: A Short History of Endless Debates”. *Ergo: An Open Access Journal of Philosophy*, *6*.
- Murray, S. (2020). “A Case for Conservatism About Animal Consciousness”. *Journal of Consciousness Studies*, *27*(9–10), 163–185.
- Nagel, T. (1974). “What is It Like to Be a Bat?” *Philosophical Review*, *83*(October), 435–450.
- Nordgreen, J., Garner, J. P., Janczak, A. M., Ranheim, B., Muir, W. M., & Horsberg, T. E. (2009). “Thermonociception in fish: Effects of two different doses of morphine on thermal threshold and post-test behaviour in goldfish (Carassius auratus)”. *Applied Animal Behaviour Science*, *119*(1), 101–107.
- Patterson-Kane, E., Elmore, M., & Pajor, E. A. (2008). “Operant animal welfare: Productive approaches and persistent difficulties”. *Animal Welfare*, *17*, 139–148.
- Proctor, H. S. (2012). “Animal Sentience: Where Are We and Where Are We Heading?” *Animals*, *2*, 628–639.
- Proctor, H. S., Carder, G., & Cornish, A. R. (2013). “Searching for Animal Sentience: A Systematic Review of the Scientific Literature”. *Animals: An Open Access Journal from MDPI*, *3*(3), 882–906.
- Quine, W. V. O. (1981). “Five milestones of empiricism”. In *Theories and Things* (Harvard University Press, pp. 67–72).
- Regan, T. (2004). *The Case for Animal Rights* (First Edition, Updated with a New Preface). University of California Press.
- Rose, J. D., Arlinghaus, R., Cooke, S. J., Diggles, B. K., Sawynok, W., Stevens, E. D., & Wynne, C. D. L. (2014). “Can fish really feel pain?” *Fish and Fisheries*, *15*(1), 97–133.
- Segner, H. (2016). “Why babies do not feel pain, or: How structure-derived functional interpretations can go wrong”. *Animal Sentience*, *1*(3).
- Seth, A. (2016). “Why fish pain cannot and should not be ruled out”. *Animal Sentience*, *1*(3).
- Shriver, A. J. (2020). “The Role of Neuroscience in Precise, Precautionary, and Probabilistic Accounts of Sentience”. In L. S. M. Johnson, A. Fenton, & A. Shriver (Eds.), *Neuroethics and Nonhuman Animals*. Springer.
- Singer, P. (2011). *Practical Ethics*. Cambridge University Press.
- Singer, P. (2015). *Animal Liberation*. Random House.
- Sneddon, L. U. (2003a). “Trigeminal somatosensory innervation of the head of a teleost fish with particular reference to nociception”. *Brain Research*, *972*(1–2), 44–52.

- Sneddon, L. U. (2003b). “The evidence for pain in fish: The use of morphine as an analgesic”. *Applied Animal Behaviour Science*, *83*(2), 153–162.
- Sneddon, L. U. (2011). “Pain Perception in Fish”. *Journal of Consciousness Studies*, *18*(9–10), 9–10.
- Sneddon, L. U. (2012). “Clinical Anesthesia and Analgesia in Fish”. *Journal of Exotic Pet Medicine*, *21*(1), 32–43.
- Sneddon, L. U. (2019). “Evolution of nociception and pain: Evidence from fish models”. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *374*(1785), 20190290.
- Sneddon, L. U., Braithwaite, V. A., & Gentle, M. J. (2003). “Do fishes have nociceptors? Evidence for the evolution of a vertebrate sensory system”. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *270*(1520), 1115–1121.
- Sneddon, L. U., Elwood, R. W., Adamo, S. A., & Leach, M. C. (2014). “Defining and assessing animal pain”. *Animal Behaviour*, *97*, 201–212.
- Sneddon, L. U., & Leach, M. (2016). “Anthropomorphic denial of fish pain”. *Animal Sentience*, *1*(3).
- Terminology | International Association for the Study of Pain. (n.d.). *International Association for the Study of Pain (IASP)*. Retrieved August 28, 2021, from https://www.iasp-pain.org/resources/terminology/
- Tye, M. (2017). *Tense Bees and Shell-Shocked Crabs: Are Animals Conscious?* (1st edition). OUP USA.
- Waal, F. B. M. de. (1999). “Anthropomorphism and Anthropodenial: Consistency in Our Thinking About Humans and Other Animals”. *Philosophical Topics*, *27*(1), 255–280.
- Waal, F. B. M. de. (2016). *Are We Smart Enough to Know How Smart Animals Are?* (1st edition). WW Norton & Co.
- Walters, E. (2018). “Defining pain and painful sentience in animals”. *Animal Sentience*, *3*(21).
- Woodruff, M. (2018). “Sentience in fishes: More on the evidence”. *Animal Sentience*, *2*(13).
- Wynne, C. D. L. (2007). “What are animals? Why anthropomorphism is still not a scientific approach to behavior”. *Comparative Cognition & Behavior Reviews*, *2*(1), 125–135.