

# *Boundaries of Consideration*

**Peter Godfrey-Smith**

University of Sydney

Whitehead Lecture 2, 2023, Harvard University

## ***1. Introduction***

## ***2. The Sentience Principle***

## ***3. Sentience in Non-Utilitarian Views***

## ***4. Experience and Agency***

## ***5. A Positive View***

## ***1. Introduction***

The first lecture in this sequence was a discussion of the biology of felt experience, with a focus on beginnings and liminal forms – the ground floor, the sub-basement.<sup>1</sup> I argued for a view of the biological basis of felt experience, and its distribution. More accurately, we reached at the end of that lecture a cluster of views that are hard to pull apart. But a broad distribution of experience, compared to many other positions, is seen in all the views in that cluster. I think we can reject a restriction of sentience to mammals or vertebrates, by way of the idea that only these animals have the right kind of brain. There is a lot of *neural multiple realizability* of experience-relevant traits. Behavioral complexity evolved on several distinct lines within animals, after earlier branchings from common ancestors. It evolved in contexts provided by different kinds of bodies, lifestyles, and brain architectures; those different bodies and architectures all allowed

---

<sup>1</sup> This is a revised version of the second of two Whitehead Lectures given at Harvard University in April, 2023. The first lecture, "Limits of Sentience," like this one, can be found here: <https://petergodfreysmith.com/philosophy/mind> (at the top of the page). The text of the first lecture is close to the talk as given. This second one has been rewritten in places, and augmented, in response to discussion after the lecture and during the visit. These changes, which are near the end, are indicated either in the main text or in footnotes. I am very grateful to the Philosophy Department at Harvard for inviting me to visit and give these talks.

acute sensing, complex behavior, and learning. We encounter several invertebrate animals with *conspicuous apparent marks* of felt experience, especially among cephalopods and crustaceans.

The setting in the first lecture also included a commitment to *gradualism* about the history of experience, and graded presence as likely today – a view without sharp lines, or a light that turns on. With that last move, we reach a tangle of issues, both substantive and terminological, about whether something is a *version* of felt experience or something else that is a bit *like* felt experience. I expect that imperfect posing of the issue to be replaced with the aid of a finer-grained vocabulary that is yet to come. But I tend towards a view in which what is conspicuous in the three "special" evolutionary lines exists in fainter but real forms in other groups. That is as far as we got.

Here is the main topic of this second lecture: it is common to see a coordination between sentience and some sort of moral considerability – "moral status" or "moral standing," as many discussions now have it. It's common to think something like this: all and only sentient systems are morally considerable, or deserve moral concern. This idea is sometimes called *sentientism*.<sup>2</sup> If one believes this, it is obviously important to work out who is sentient. A broadening to insects or plants, or a narrowing to mammals, will have a lot of consequences. The main topic today is how this principle coordinating sentience and moral concern looks in relation to the ideas in the philosophy of mind developed in the other lecture. I'll first discuss these principles in a general way, and then look at their role within some recent discussions of ethics and animals – especially the place of sentience within some *non*-utilitarian treatments, in books by Christine Korsgaard and Martha Nussbaum. Along the way, I'll fill out the philosophy of mind side, tying up some loose ends from before. Towards the end I'll sketch a positive view on these matters.

## ***2. The Sentience Principle***

I'll use this formulation of the principle I'll be discussing:

*Sentience Principle:* All and only sentient beings have interests worthy of moral consideration.

---

<sup>2</sup> The term "sentientism" goes back to at least the early 1980s. See Johnson, "Animal liberation versus the land ethic," *Env. Ethics*, 1981. Thanks to Lori Gruen for help with the background here and other issues.

There is the *all* side and the *only* side. Both sides raise questions. The phrase "worthy of" is intended to be weaker than "demand" or "require." Some issues I won't consider include applications of the principle to beings liable to become sentient, even if they are not so now, and I won't discuss decisions about which beings will come to exist, as opposed to the interests of those who do exist; I will keep things simple in those ways.

I take the sentience principle to be familiar within utilitarianism. It is seen in Bentham's famous passage about the right question to ask about who deserves moral consideration.<sup>3</sup> Don't ask of creatures: can they reason, or can they talk? Ask: can they suffer? Singer in *Practical Ethics* expresses the idea clearly:

[T]he limit of sentience (using the term as convenient, if not strictly accurate, shorthand for the capacity to suffer or experience enjoyment or happiness) is the only defensible boundary of concern for the interests of others.<sup>4</sup>

Differences between versions of utilitarianism – between preference and hedonic versions – do arise here, but I won't be talking about utilitarian views much, and when relevant, it will be hedonic, or at least experience-centered, versions that are on the table.

Eventually, I'll look at some issues around the *status* of the principle. There's a simple way the principle can be understood: We have a property – *sentience* – and another property – *moral status* – and these have a tight connection. In some recent discussions in philosophy of mind, the principle tends to be discussed like that. Moral status is taken to be a real property that we can have intuitions about. Sentience is also a real property, usually discussed as an on/off matter (where this is compatible with what I called partial gradualism – there's a hop onto the escalator, and a gradient, or many gradients, from there). The pattern of the discussion tends to be moral realist and intuitionist, whether this reflects a real commitment or not. I will leave these questions about status open for now, and come back to them later.

Sometimes the term "sentience" is used for the capacity to have any kind of felt experience. It can also be used more narrowly, for experience that includes pain and pleasure (perhaps along with other "evaluative" feelings). *Narrow sentientism* is then the

---

<sup>3</sup> *Introduction to the Principles of Morals and Legislation*, 1789.

<sup>4</sup> In all my block quotes, underlining indicates emphasis added.

The context there is a discussion of *equal* consideration, but Singer also endorses the basic view.

view that a being has interests worthy of moral consideration if and only if it has a capacity for felt experience, where this includes pleasure and pain. *Broad sentientism* holds that a being has interests worthy of moral consideration if and only if it has a capacity for felt experience, whether or not it can feel pleasure and pain.<sup>5</sup> A discussion has arisen around some thought-experiments, due to David Chalmers, featuring imaginary "Vulcans" who are conscious but not sentient in the narrower sense.<sup>6</sup> These imagined agents are supposed to be about as complex as people and behave in fairly normal ways, but with this "hole" in their experiential lives. I am doubtful of the thought-experiment in that form, but in neurally simpler animals, where it's not a matter of preserving a human-like behavioral profile, the possibility is worth thinking about.

For a time, I wondered whether insects might be in a situation like this.<sup>7</sup> When we think of the perceptual abilities and the demands of flight, in insects such as bees and flies, some form of *perceptual* experience looks likely. But various reports suggested that insects might have a gap in their experience in relation to pain – perhaps not in relation to all evaluation-related capacities (later I will discuss mood-like states), but in relation to acute pain from bodily damage.<sup>8</sup> Insect researchers sometimes said that heat, not bodily damage (which is what had been studied), is the aversive stimulus we should look at. And last year, a hole in our understanding and this apparent hole in plausible experiential profiles for (some) insects was filled by a paper from the lab of Lars Chittka, with first author Matilda Gibbons.<sup>9</sup> They found the same kind of evaluative trade-offs in bumblebees that have been seen as good evidence for felt pain in their crustacean relatives (and with some further complexity). The bees avoided heat in a food source area

---

<sup>5</sup> See also Browning and Birch, "Animal sentience," *Philosophy Compass*, 2022

<sup>6</sup> See his *Reality+*, 2022.

<sup>7</sup> See "Varieties of subjectivity," *Philosophy of Science*, 2020.

<sup>8</sup> Eisemann et al, "Do insects feel pain? — A biological view," *Experientia*, 1984. See also Groening et al., "In search of evidence for the experience of pain in honeybees: A self-administration study," *Scientific Reports*, 2017. Walter Veit, in his dissertation "Health, Agency, and the Evolution of Consciousness" (2022) has corrected an error in some of my earlier discussions here (e.g. in *Metazoa*) – it is not clear that no insect has been observed grooming a wound site. Eisemann et al. note a possible case in cockroaches.

<sup>9</sup> Gibbons et al., "Motivational trade-offs and modulation of nociception in bumblebees," *PNAS* 2022: "Bumblebees avoided noxiously heated feeders less when these dispensed higher sucrose concentrations than unheated feeders. Unlike trade-offs described in other invertebrates., this trade-off relied on associative memories, rather than direct experience of the stimuli." For more on bees, see Chittka's book *The Mind of a Bee*, 2022.

in a way finely calibrated with the superiority of the heated over an unheated food area. This is the sort of thing that had not been seen previously, in a way that was puzzling. Bees are sophisticated learners, even showing cultural learning (learning by copying their fellows). A "gap" for pain seemed surprising. Now there is evidence for no gap. This is not decisive, and should not be extrapolated to all insects. But I use this work, and other work from the Chittka lab, to justify the idea mentioned in the first lecture, that insects are no longer the frontier case for animals where there is decent evidence of experience. Together with some of the fly work I mentioned then, insects are now in the group where quite a lot can be said on the *yes* side.

This is significant; insects have often been used as "foil" cases in ethical discussions. Look after the cows; don't worry about the insects. Peter Carruthers once argued that not having to worry about insects is a constraint on ethical sensible views, hence a potential problem for utilitarianism.<sup>10</sup> Insects are often advocated now as a protein source. This would involve massive numbers. And more generally, their interests and human interests tend not to be aligned, to put it mildly.

An in-principle possibility of a gap in evaluative experience remains, even if insects don't occupy it (or if some insects do). All choices of action have a fact-related side and an evaluative, or preference-related, side: what do I want to get, and how do I get it in these circumstances?<sup>11</sup> Suppose we have a "threshold model" relating cognition and felt experience, one motivated in part by the recognition that very simple forms of cognition are found just about everywhere in cellular life. Then even if all actions at least implicitly involve a factual side and evaluative side, might there be an organism who is over-threshold on the perceptual side and not on the other? And might we find others who have the opposite combination? I take this possibility less empirically seriously than I used to, but it's part of the landscape. Below, when I talk about "sentience" I will generally assume we're talking about experiential profiles that do include an evaluative side.

---

<sup>10</sup> Carruthers, "Invertebrate Minds: A Challenge for Ethical Theory", 2007: "it is a fixed point for me that invertebrates make no direct claims on us, despite possessing minds in the sense that makes sympathy and moral concern possible. Invertebrates believe things, want things, and make simple plans, and they are capable of having their plans thwarted and their desires frustrated. But it isn't wrong to take no account of their suffering. Indeed, I would regard the contrary belief as a serious moral perversion."

<sup>11</sup> This is the structure of an expected utility model. See "Varieties of subjectivity" for more on the topics of this paragraph.

### ***3. Sentience in Non-Utilitarian Views***

Most philosophers now do not see utilitarianism as viable as a general ethical theory, especially because of its disregard for "the distinction between persons" (Rawls), and considerations of distribution. I'll sometimes talk about approaches I'll call *welfarist*, where the sentience principle is endorsed and experience is all that matters, but there's no summing over individuals and we consider the good of each individual on its own. That is not much of an outline of a view, as we have nothing on the table so far about conflicts and equity. It's a piece of various possible views.

I'll spend more time on some views that are further from utilitarianism – views based more on the individual as *agent*, as *chooser*. These views started life as frameworks aimed at the human case, and have been modified to deal with nonhuman animals. When those modifications and extensions were done, considerations of sentience ended up back on the table. The views I'll discuss are those in Korsgaard's *Fellow Creatures* (2018), and Nussbaum's *Justice for Animals* (2023).

Korsgaard uses a Kantian framework, drawing also on Aristotle. Agents of certain kinds *choose what is good for them*. We, as agents ourselves, make our own choices of what is good, and see these choices as defensible and as a basis for claims on others not to interfere with us. As a result of this, we find ourselves bound also to respect the choices of others. Their goods, too, become effectively *absolute* – recognizable and motivating from all perspectives.

I have criticized this central argument of *Fellow Creatures*, the attempt to leverage our own sense of ourselves as agents into a concern for others, in an essay in *Aeon*.<sup>12</sup> I won't revisit those issues here, but will instead be looking at the role of sentience.

---

<sup>12</sup> See "Philosophers and Other Animals," *Aeon*, 2021, with extra comments in "Further Thoughts on Fellow Creatures" (on my website). Very quickly: Korsgaard says that the way we take our own motives for action as adequate involves a commitment to the idea that what we seek is "absolutely good." Something is absolutely good when it can be recognized as good by everyone. If something is recognizable as good by everyone, it seems we all have reason to pursue it. But there are two ways something can be recognized as good by everyone. It might be recognized, by everyone, as good for anyone who is in a situation just like mine (not: for me individually, but anyone in a situation like mine). That does not mean it is recognized as good in another sense, where it becomes part of a "shared" good that others have reason to pursue.

Officially in *Fellow Creatures*, agency is what matters – making choices – but the relevant kind of agency requires sentience. The book opens as follows:

In this book I defend the claim that we human beings are obligated to treat all sentient animals, that is, all animals who have subjective experiences that are pleasant or painful, as what Kant called "ends in themselves," at least in one sense of that notion.

Another passage:

The organisms we are concerned with when we think about whether we have duties to animals are sentient beings who perceive the world in valenced ways and act accordingly. This is the feature of organic life that I have argued places an organism in the morally interesting category of having a final good.<sup>13</sup>

Martha Nussbaum's *Justice for Animals* came out earlier this year (2023). Here, she modifies the "Capabilities Approach" that she has helped develop, for the human case and especially in the context of development economics, with Amartya Sen and others. The new book is extension of that approach to the animal case (reworking a treatment that was part of her earlier Tanner Lectures).

Injustice in general, for Nussbaum, is a matter of *wrongful thwarting* – wrongful interference with action and projects. Animals, like humans, lead lives of "significant striving." It is unjust to thwart their forms of striving.

[T]he general intuition should be emerging more clearly: injustice centrally involves *significant striving* blocked by not just *harm* but also *wrongful thwarting*, whether negligent or deliberate.<sup>14</sup>

The basic goal is that all animals would have the opportunity to live lives compatible with their dignity and striving, up to a reasonable threshold level of protection.

---

<sup>13</sup> Another: "*The Absolute Goodness of Goodness-For*: It is absolutely good, good-for us all, that every sentient creature get the things that are good-for her, and avoid the things that are bad-for her."

<sup>14</sup> More detail: "our everyday pre-philosophical idea of injustice, which involves, I think, the idea that someone is striving to get something reasonably significant, and has been blocked by someone else - wrongfully, whether by malice or by negligence. (Ch. 1). Nussbaum avoids the term "moral status" here, as that category is broader and the book is about justice in particular.

But this does not apply to all animals – only to those who are sentient:

According to the CA [capabilities approach], each sentient creature (capable of having a subjective point of view on the world and feeling pain and pleasure) should have the opportunity to flourish in the form of life characteristic for that creature.

Again, agency drives the view – choice, striving, seeking your good. It is not utilitarian. But again there is a restriction to sentient beings. Both books discuss plants as contrasting cases, and also some animals. The intention is not to have an *ad hoc* move to prevent the view from having an awkward amount of breadth, but to have a motivated restriction to sentience.

This gives us some questions. What *is* the relation between the agency side and this experiential side, both in these views, and in nature itself? Why the extra clause? Might a better case be made with the agency side only, or with a different restriction? How do these two elements – agency, sentience – relate to questions about gray-area cases and gradations?

#### ***4. Experience and Agency***

I'll discuss both how things look within my own preferred view of sentience, and also within other, more common views – the ones I called "narrow pathway" views in the first lecture. Narrow pathway views are worth discussing up front, to show some of the contours of the situation. The motivating thought in that project is that lots of coherent behavior can be produced wholly unconsciously. Felt experience involves a special path. This special path is relevant to agency – in us, it involves sophisticated attentive behaviors. But it's optional for the basics. Narrow pathway views tend to make sentience problematic in most nonhuman animals, certainly outside mammals.<sup>15</sup>

If narrow pathway views were right, agency-based ethical projects of the kind I am discussing here might associate moral consideration only with agency that is sophisticated enough to require consciousness, *or* treat conscious experience as a sort of

---

<sup>15</sup> For an agency-based ethical view: there is also the issue of sentience as a general feature of an animal, and sentience as relevant in a particular choice. I set that aside as I want to move quickly through this stage.



add-on: to be morally considerable, you need to be an agent pursuing your good and be sentient. Or perhaps they might leave the sentience principle behind.

I cover this topic quickly because I reject those views and see felt experience as more widely distributed. But this quick discussion gives us a shape that will appear in a different form in my view: *agency outstripping sentience*, to some extent.

The view presented in the first lecture has a lot of coordination between agency and sentience. The evolution of agency brings with it the evolution of a cluster of subjectivity-related features – the formation of a point of view or perspective on the world. I said that roughly speaking, the evolution of agency brings with it the evolution of subjectivity. That is one part of my account of the biology of felt experience. The other part involves the special features of nervous systems. Nervous systems are *how* animals generally achieve agency; non-neural animals (sponges, placozoa) show only simple forms of agency when compared to animals that do have nervous systems. But nervous systems I also see as important in explaining the *felt* side of subjectivity; this was, in the previous lecture, a two-part account.

I want to add at this stage another element to the philosophy of mind side. This has to do with the nature of *evaluative* experience, and this will push the category of sentience, even a little deeper, perhaps, that I have so far.

Pleasure and pain are the standard topics in this area, and I discussed octopus, insect, and crustacean pain. But there is more to aversive experience than this. As well as pain, there is stress, alarm, fear, and frustration. Perhaps surprisingly, the evidence for some of these in invertebrates has for a while been fairly strong. The main reason to hesitate in the face of this evidence is that some of these animals are *so* simple in their nervous systems. Here I have in mind what might be called mood-like states in insects and in gastropods (snails and slugs). The states are sometimes described as emotions, or emotion-like, but I think *mood-like* is a good description. An example of these medium-term phenomena, running well and truly into frontier territory, is *nociceptive sensitization* (here I draw especially on the work of Terry Walters).<sup>16</sup> Nociceptive sensitization is a heightened and general sensitivity after an aversive event. This is seen in gastropods, and also fly larvae (the grub-like early stages). Electric shock tends to be used as an aversive

---

<sup>16</sup> See Walters, "Nociceptive biology of molluscs and arthropods: Evolutionary clues about functions and mechanisms potentially related to pain," *Front. in Physio.*, 2018; Crook and Walters, "Nociceptive behavior and physiology of molluscs: Animal Welfare Implications," *ILAR Journal*, 2011.

stimulus in this work. As well as inducing immediate responses, and associative learning, it leads to a general sensitization. As Crook and Walters put it, "memory of a noxious event in snails can be linked to a fear-like motivational state that can dramatically alter the animal's response to other biologically significant stimuli." It's a kind of anxious, sensitized, "downer" state that is induced.

If we think back for a moment to more complicated cases, like pain in octopuses, it is in part the integrated nature of the animal's response to a bad event – the fact that the response is not reflex-like, but pervades several behavioral contexts – that supports an experiential interpretation. We discussed that in connection with Robyn Crook's 2021 octopus experiment. Here, in nociceptive sensitization, we have the *very faintest version* of something like that – something like that kind of integrated and pervasive effect of an aversive stimulus. It's not just a little local arc within the animal. And this, as I said, reaches deep into the puzzling territory, to organisms that have only tens of thousands of neurons.<sup>17</sup>

The question of how states like this relate to ethical views that make use of agency, but also include a role to sentience, is not completely straightforward. If one is concerned with experience *per se*, in an ethical context, then it's natural to include mood-like states as ethically relevant. And Walters and his colleagues do think that these features are adaptive – they do something useful in guiding action. But whether they condition the particular goals that are sought, the particular choices, in the way that pleasure and pain might is not so clear.

Anyway, if one accepts a sentience principle – in any context, utilitarian or otherwise – then one has to deal with the fact that sentience probably runs deep in the animal kingdom, deeper than has often been supposed. Korsgaard and Nussbaum both leave open the possibility that insects might be sentient. As I've said, I think they're probably "in," and they are not the only ones queuing up for the ark.

---

<sup>17</sup> There is less work on the positive-experience side. But it pops up, especially in this medium-term, mood-like setting. Bateson et al. ("Agitated honeybees exhibit pessimistic cognitive biases," 2011) found that an aversive shaking could induce a kind of pessimism in honeybees – an inclination to treat ambiguous stimuli pessimistically. They interpreted this state as emotion-like – it might again be called mood-like). Cwyn Solvi and collaborators followed this up and showed that a *positive* emotion-like state, a kind of optimism, can be induced in bumblebees by means of unexpected reward, one with analogous effects on the interpretation of ambiguous stimuli ("Unexpected rewards induce dopamine-dependent positive emotion-like state changes in bumblebees," 2016).

That's the first main point of this section. Next: although sentience and agency are closely linked, they're not completely inextricable. Agency runs deeper. If felt experience is a feature of nervous systems, then when we leave organisms with nervous systems, we are leaving felt experience behind. But agency – pursuing the organism's own good – goes further. In the case of organisms with very simple nervous systems, the same point might apply, but there we certainly encounter the problems of graded presence discussed in the first lecture. So I'll talk next about non-neural organisms, such as plants.

If it really is the active seeking of the organism's good that we are interested in, if that is driving the story, then forms of this feature probably outrun even the most liminal forms of sentience. I said in the first lecture that "minimal cognition" is all over cellular life – detecting what is going on, and reacting in a way that expresses a preference. Here is an example from plants. This is from a 2019 study of the evening primrose.<sup>18</sup> When exposed to the played-back sound of a flying bee, or a similar synthetic sound, the flowers produced sweeter nectar, and did so within three minutes.<sup>19</sup> This is thought to be aimed at increasing the chances of cross-pollination. The flowers themselves vibrated mechanically in response to the sounds, suggesting a mechanism where the flower serves as an auditory sensory organ.

There are many plant cognition reports somewhat like this, but they are often hard to interpret. This one looks clean, and also striking in how fast the adaptive response is (three minutes). The idea of a flower as an ear is, of course, also irresistible.

In some ways, this is not *action* because it's not movement. Venus flytraps are perhaps as good an example here, at least in some ways. Animal action involves, especially, movement of body parts of some kind. So does a lot of protist action – action in the complex single-celled organisms of the earlier and ongoing microscopic world from which animals came. The Venus flytrap mechanism is quite simple. This plant-sound case might also be simple, but it's new, I don't think we know much about how it works.

In some ways, then, this is different as a kind of agency. But the *good being sought* is not one that's derivative on our perspective as observers. It's a good for the

---

<sup>18</sup> Veits, et al., "Flowers respond to pollinator sound within minutes by increasing nectar sugar concentration," *Ecology Letters*, 2019.

<sup>19</sup> "Both the vibration and the nectar response were frequency-specific: the flowers responded and vibrated to pollinator sounds, but not to higher frequency sound. Our results document for the first time that plants can rapidly respond to pollinator sounds in an ecologically relevant way."

organism that is sought by the organism, as best it can, by tracking what's going on and responding appropriately.

We could have a category of goods sought by an organism *by way of feelings* – good that sought through acts guided by evaluative feelings of certain kinds. Does that make it a different kinds of good, with a different status? It's certainly a different kind of *seeking*. In the non-neural cases, including plants, the pursuit might always be relatively simple. More complex cases may always involve sentience, because of the unique role of nervous systems, the means by which agency is achieved and animals. The view I am resisting is the idea that the kind of good being sought is inherently different when sentience is absent.

Korsgaard at a few point wonders if plants might be at a low point on a continuum, for her purposes, with conscious animals at the other end. In relation to sentience, I doubt it, though I don't know. For me, plants are in the "no" category, at least so far, rather than the gray area. In relation to agency itself, though, I think they pass the kind of test that Korsgaard lays down as central to her view.

Korsgaard allows that if the plant/animal difference is one of degree, then "In that case, plants would be, in a very elementary sense, agents, and so might be said to have a final good" – a good of the kind that is ethically significant.<sup>20</sup> In other places, she stipulates that organisms who represent the world and seek their own good are *animals* – this is the special "animal" category she uses in the book – not a biologist's sense, but a sense fashioned for her purposes. Having plants in, or quasi-in, the animal category would make things look linguistically odd, but that's not a big deal; the category still might do the work intended. A more important issue would be the size of the category – that would be surprising, at least. And then I would add that the final-good category is not simply aligned with a category based on sentience.

---

<sup>20</sup> "An animal is an organism that functions, at least in part, by representing her environment to herself, through her senses, and then by acting in light of those representations."

"The difference between the plant's tropic responses and the animal's action might even, ultimately, be a matter of degree. In that case, plants would be, in a very elementary sense, agents, and so might be said to have a final good."

"The final good came into the world with animals, for an animal is, pretty much by definition, the kind of thing that has a final good--a good, in the sense that might matter morally." [Secs 2.1.7; 11.4.4]

I will note, again, the dependence of this point on a particular part of my account of the biology of sentience. In the first lecture, I used the evolution of agency as part of my attempt to make sense of subjectivity – roughly speaking, I said, the evolution of agency brings with it the evolution of subjectivity. With agency comes some subjectivity-related features, and this relationship runs all the way through to the minimal cases. As agency exists in non-neural organisms, agency there brings with it *some* of what is involved in felt experience, in a simple form. The other part of my view is the idea that nervous systems are special – they are part of the explanation of felt experience itself. Non-neural organisms, as well as having a simpler kind of agency (in at least many respects), lack something needed for felt experience. It would be possible to insist that the subjectivity-related features that are inherently linked with agency are all that is needed for felt experience. That is a hard debate to resolve, and I am working in this lecture within the view outlined in the first. We might at least say this: there are cases where the argument for genuine agency is much stronger than an argument for sentience.<sup>21</sup>

Nussbaum, in contrast to Korsgaard, is more firmly exclusive of plants. They lack the kind of individuality seen in animals, a kind that matters to action. This, she says, is true even before we consider sentience. Plants are indeed less definite as individuals. But some of what they do is like what we saw in the primrose. There's agency in the context of this tight collective, or a quasi-individual. It's natural then to wonder whether being a less definite individual – being a bit *they*-ish – may not be as generally antithetical to agency as it is, at least on my view, to sentience.

Again, it's possible to establish a category of agency in which sentience is just included, as an addition, for use in the ethical context. Nussbaum did something like this in an earlier discussion. In her Tanner lectures, she said at one point, as the ethically relevant form of agency seemed to be spreading uncontrollably, that we should "admit the wisdom in Utilitarianism," and just *require* sentience for the kind of concern that she was writing about, which is concern over justice. That is not what she does in *Justice for Animals*. There, she does want to argue that a sentience requirement is a natural condition on the sophistication of agency or genuineness of agency. I'll put the details in a footnote,

---

<sup>21</sup> That is how Nicolas Delon put the point in a talk on these topics, "Agential value," at the 2023 Pacific APA.

Most of this paragraph is an addition to the original lecture.

but we go through an analogous sequence of moves that we saw a moment ago with Korsgaard.<sup>22</sup>

So far in this section: Agency, seeking the organisms own good, doing so by tracking local circumstances and responding, is all over cellular life. Animals do this in a complicated, targeted way, a way that involves a *reach* into the environment that's unusual, made possible by the twin inventions of nerve and muscle. Sentience is probably spread widely. But the idea that sentience matters to the basic fact of agency, in a sense picked out by the kind of good being sought, is not so plausible.

"Sentience is spread widely," I just said. Once again, as in the first lecture, we need to look at the role of graded presence in that property. I said in that lecture that graded presence of sentience is likely. This is hard to describe using our present language. The case of *life*, as a category, is an earlier case that seems similar, and is now understood pretty well. That provides some encouragement. But what can we do now, while we wait for a better framework for addressing the gray-area cases of sentience or felt experience?

We can't do that much, but we can press forward a little, using the two elements that I use to make sense of the biology of felt experience in the other lecture: a more cognitive set of features, and another that draws on properties of nervous systems. One thing we can ask is: what if you have one side without the other? The idea is not that we

---

<sup>22</sup> Here is her main passage on this topic (from chapter 6):

The creatures discussed in the CA, the creatures whose significant striving the theory requires us to protect, must, it seems, be capable of perception and desire, and of moving in response to that combination. By perception I mean (however difficult it is to get at this in practical terms) the capacity to focus on objects in the world, in a way that's not just a causal collision, that has real directedness or what philosophers call intentionality. The world *looks like something* to these creatures. They have some sort of subjective experience. With desire, it is similar: the creatures we are looking for don't just mechanically jump away from harm or move toward food; they have a felt orientation toward what is seen as good and a felt aversion to what is seen as bad. That is what makes their striving significant. They are not just automata.

In other words, they possess that elusive property known as *sentience*.

A problem here has the rough character of a false dichotomy. Nussbaum contrasts perception that includes subjective experience to a mere "collision," and contrasts a "felt" orientation to what is seen as good with a purely "mechanical" response. Agency and minimal cognition in non-neural organisms show us that "directedness" on the world might outrun felt experience, and their responses need not be "mechanical" in a sense that does not apply also to ourselves.

can imagine full human-like complexity in one side and not the other, but a faint, minimal version.

Perhaps I went too fast there: why can't there be human-like cognition and behavior without a nervous system? Philosophers are very much used to such an idea. People say it would not be too hard to realize a mind like ours, including experience itself, inside a computer. Mental properties are "substrate neutral," because they are "functional" properties. They concern what's done by the parts of a system, not what the system is made of. So, philosophers say, we could slowly substitute artificial parts for the neurons in your brain, one at a time, preserving their functional properties. There would be no change in behavior, given that functional sameness, and it then makes no sense to think that your experience would "fade" as the substitution is done.<sup>23</sup>

These are not good arguments. I was convinced of this by Rosa Cao, here at Harvard when she was leaving neuroscience for philosophy, doing a postdoc with Sean Kelly. The next bit of the talk draws on her work.<sup>24</sup> "Functional properties," the properties that concern what a system or part of a part of it *does*, exist at all levels of grain. If we substitute a non-neural control system for a neural one, then coarse-grained functional properties may well be retained, but fine-grained ones will not. Philosophers, for some reason, got used to thinking of functional properties in a coarse-grained way, and talking about functional *identity* across systems that share coarse-grained functional properties. They also located the *mental* in those coarse-grained functional properties, ignoring the fact that if the mental is "functional," there is no reason to think that fine-grained functional properties drop out of the picture.

In reply, it might be argued that although this is generally true – there being coarse-grained and fine-grained functional properties, and so on – the important properties of neurons in our brains are *especially portable* across systems and hardwares. The important properties are the network properties, the local cell-to-cell influences, the firing of these cells causing firing of that one, and so on. If one thinks that that's all that matters in a brain, then it might be a special case. But in the first lecture, I argued that this is probably not all that matters, and a lot of what matters is less portable. I reject classical multiple realizability arguments for that reason. My view end up being mixed, on

---

<sup>23</sup> Pylyshyn might have given the first version of this argument. The arguments are discussed in detail in Chalmers, *The Conscious Mind*, 1996.

<sup>24</sup> See Cao's "Multiple realizability and the spirit of functionalism," *Synthese*, 2022.

questions of multiple realizability. There's lots of it within neural systems – evolution gave us that. But the multiple realizability doctrine seen within classical functionalism, especially in connection with artificial systems, I reject.

That's a general view of multiple realizability. In this particular context, talking about gray-area cases, the questions look a bit different. We're not asking about the replication of human minds with human experience; we're asking about minimal cases, systems that are analogous to animals with simple nervous systems, but now with a different control system inside. We wonder whether they would have something that we would crudely describe as another form of felt experience, or crudely describe as something else, something a bit like experience, but not it. When that's the question, the status of artificial systems and others that have part of the total package looks different. Then I think it is reasonable to think that there will, before too long, be artificial systems that are more brain-like in their construction, that do have a kind of minimal subjectivity, systems that look like they're getting into the gray area also seen in the animal case.

When we think about the other combination of presence and absence in relation to those two factors I talked about (cognitive, neural), there's also, both possibly and in early forms actually, systems that have some of the large-scale brain dynamics that I think are important, with none of the cognitive features associated with subjectivity that you find in an animal. These are seen, to some extent, in what are called *brainoids*, lab grown collections of nerve cells, that take on spontaneously some of the large-scale oscillatory patterns seen in our own brains. They do this with nobody being controlled, no senses, no action. Brainoids are produce some neural dynamics in a subjectivity-free context.

Future brainoid-like systems will raise questions akin to the questions that arise with robots and artificial systems with better hardware. Each is the flipside of the other, with respect to the two features that I used to explain the biology of experience in the previous lecture.<sup>25</sup>

---

<sup>25</sup> In the case of the artificial systems, another reason to question their role has to do with whether they really *have* the subjectivity-related properties associated with experience, or merely model them for us.



## 5. *A Positive View*

I'll now outline some parts of a positive view in the areas covered by this talk. I'll start with the meta-ethical side – what we might be *up to* when we make claims in this area.

I said earlier that especially in recent philosophy of mind, issues around sentientism tend to be discussed in a way that includes the surface appearance of moral realism and intuitionism. "Moral status" is treated as a real property, one that we learn about by considering possible cases and responding intuitively. When pressed on this, philosophers sometimes distance themselves from this meta-theoretic perspective, and offer a commentary that is different, endorsing a constructivist view or something like that. They may think that more realist modes of talk can be licensed by a subtle role they can have in the first-order discussion, as discussed by Simon Blackburn (*Ruling Passions*).

I think it is good for a person's view of the status of this debate to be on the table all the way through. I am a long way from moral realism and intuitionism, and this affects where things come out on the immediate, first-order issues.

Here is a general orientation to the ethical. I think that *valuation* is a basic thing that we do. Valuations are not reducible, in general, to something else – to description of how things are, expressions of emotional responses, or commands. Valuation is its own thing. It involves putting options in order (choices, outcomes, social arrangements...). A valuation has a variety of manifestations in discourse, other actions, and thought. Valuation in humans has various species, and ethical valuation is one of those. Ethical valuation is paradigmatically (though not always) concerned with behavior in social contexts, or at least that affects others. That place in social life gives it some distinctive features. Because this kind of valuation has a role in organizing our social life, disagreements very often can't just be let lie. They need to be resolved; we have to work out how to live, what to do.

Another element distinguishing the ethical from other kinds of valuations also involves this social role. I assume that our ethical habits of thought and talk derive from earlier contexts in which norms were established and enforced in societies. These norms and valuations are often entangled in theology and sheer coercion, but a stage that is reached, in many cases, is one where these valuations are brought into contact with our faculties for reasoning and public justification. When norms can be questioned and defended, considerations of parity become important: If you treated *this* case like *that*, why don't you do "the same" with this *new* one? Pressure towards non arbitrariness arises

very readily, and is a kind of ancestor to universalizability in the strong sense. In some cases, including our present cases involving nonhumans, pressure on valuations comes from parity *plus* new factual knowledge. We can learn that this case is, empirically, more like certain other cases than we had realized, and that can put pressure on what we've been doing up till now.

Ethical discussion of this kind has the goal of *reasonable valuation*. Ethics is about defensible, hence systematic, valuation that handles (especially) social questions. The goal of reasonable valuation is not the same as a goal of representation of what's "out there"; what it answers to is different. There can be a lot of constraint from factual matters, especially by way of parity, but there's also a good deal of inherent freedom of movement. A person can choose how well integrated with the rest of their thinking they want their ethical orientation to be. Claims about parity are also dependent on which similarities between cases are seen as important and which are seen as irrelevant. They depend on our rather flexible sense of what is similar to what.

A moment ago, briefly, I distanced myself from Simon Blackburn's absorption of realist meta-talk (where you thump the table and say things are *true*) into practical first-order ethical discussion. In the same book, *Ruling Passions*, Blackburn has a label for his view; he calls it *nondescriptive functionalism*. Blackburn's actual view, when filled out, is too close to traditional expressivism (valuation, again, is its own thing). But I think that label is a helpful one: nondescriptive functionalism. I suppose I am a kind of nondescriptive functionalist. The titles of these two lectures – "Limits of Sentience, Boundaries of Consideration" – are supposed to indicate a contrast between what we're doing in each case: roughly, discovering the limits of sentience and deciding on the boundaries of consideration. Boundaries are things we put there. We can determine them reasonably, or less so. With sentience, it's more a matter of discovery. The claims we make are answerable to nature in a different way.<sup>26</sup>

---

<sup>26</sup> Part of what happens, perhaps, when valuations and norms are made explicit and brought into a context of defense and justification is that they take on a propositional form that makes them *look* more descriptive – with generalizations, cases, conditionals, and so on.

A comment on how this view of the procedural side relates to Rawls's concept of reflective equilibrium: his view gives an important guiding role to particular cases, and to particular judgments about them that we are unwilling to give up. My view does not – although, again, there is plenty of inherent freedom of movement here.

In the case of nonhuman organisms and the problems that we are grappling with here, I see our goal along the lines just sketched. In general, we have the goal of reasonable or reasoned valuation. In the case of many of these problems with nonhumans, we are trying for the reasoned extension and modification of patterns of valuation that were initially shaped around human social relations. That's what we're trying, or might realistically try, to do.

Turning to the specific issues discussed, I endorse something like the sentience principle (with more to say some time about the "only" direction of the claim, in relation to species, ecologies, and the like). The sentience principle is historically allied with utilitarianism, and it also suggests a wholly experiential view of what is ethically important. This might take the form of what I called "welfarism" earlier, rather than utilitarianism, but that wholly experiential view of what is ethically important can in any case be rejected. In this lecture, I said that sentience runs deep but agency runs deeper. We can also say this: In the absence of sentience, agency does not matter. But in some cases within the class of sentient beings, agency matters a lot, and its role in reasonable valuation is not reducible to something hedonic.

That's a coherent combination of views. I'll say it again a little differently: Sentience runs deep, agency runs deeper; in the absence of sentience, agency does not matter; but in relation to very complex organisms like ourselves, agency matters a lot, and its role is not reducible to something hedonic or experiential. There's no clash there. There is some failure of systematicity, in agency looming very large in some contexts and disappearing from view in others. Failures of systematicity always generate some pressure in a context of challenge and justification. But that pressure is not the only thing.<sup>27</sup>

If our goal is reasonable valuation, why is this a reasonable valuation? I do not think we are *pushed* into this, or anything like this. It is one way to resolve some of the questions. I accept an agential moral framework, of some kind, in the ethical contexts involving human social relations. In this context, the classic arguments against utilitarian and experience-based views are good ones – arguments that involve fairness and distribution, neglect of the distinctness of persons, and also rejections of paternalism.

---

<sup>27</sup> Making one more reference back to earlier Harvard generations, this might sound reminiscent of a slogan that was discussed, but not endorsed, by Nozick in *Anarchy, State, and Utopia*. There he looked at the idea of "utilitarianism for animals, Kantianism for people." It's not the same, but I see the kinship.

Some of those arguments bear specifically on utilitarianism; welfarism, in my sense, does respect the "distinctness of persons." But others, such as those relating to paternalism, are more general. In organizing human social life, respect for autonomy and principles of liberal democracy are good organizers of social relations. I endorse those kinds of valuations (in some form – this is a big family of views). Human rights, for example, are very good political constructs.

Patterns of valuation that include concern for autonomy also get some purchase on the nonhuman case – when we leave the social-political human realm, those kinds of valuation need not just fall away. But a more welfarist pattern of valuation starts to get more traction in this context. The forms of agency seen in neurally simpler animals and non-neural organisms are so far from the form seen in the "home" context of agent-based views in ethics, human social affairs, that while an agential extension of concern does not become impossible or incoherent here, it does end up less motivated. Moral concern can stop with sentience.<sup>28</sup>

I just said it can "stop," but in the first lecture I argued that we can expect an outcome in which sentience is not a yes-or-no matter, but has a graded presence. *Partial* gradualism would retain a yes-or-no question about a minimal kind of sentience, and allow gradations from there; a *complete* gradualism does not retain a yes-or-no question about minimal cases. If sentience is strongly graded in this way, what consequences does that have for moral consideration?

The first point I want to make about this was not part of the lecture as given, but came up in the discussion afterwards. The point was made by Selim Berker. I expressed the sentience principle with the term "worthy" – all and only sentient beings have interests worthy of moral consideration. Worthiness, Berker noted, is itself a graded or gradable matter. One can be more or less worthy of something. This way of setting the principle up is friendly to a graded version of sentientism. I did not have this in mind when I wrote the lecture – I just wanted a formulation that was weaker than some others (avoiding notions of obligation). But this is also a good way to formulate the principle if one goal we have is a natural fit between this ethical idea and a graded view of sentience. We'd end up with something like this: beings that are more clearly sentient have interests more worthy of consideration. If the principle is expressed like that, it accommodates complete gradualism, not merely partial gradualism.

---

<sup>28</sup> This paragraph and the previous one are mostly new.

In the discussion of sentience, I also said that some of the puzzles and difficulties that arise around indefinite, gray-area cases can be expected to dissipate when our current, imperfect concepts in this area are replaced with a finer-grained treatment. What we call felt experience or phenomenal consciousness may resolve into a number of other properties (as happened in the case of *life*). This outcome may have a reflection on the ethical side; valuation and ethical choices can attach to the "strands" recognized in the finer-grained treatment. Different sorts of consideration might be put in place for different features within the older package we call "sentience."

Lastly, in these lectures I've been burrowing around the deep and murkier parts of the nonhuman terrain. This has been what scuba divers sometimes called a "muck dive," a dive in the strange crannies, often uncovering the most interesting animals. We've been looking at distant parts of the tree of life, and wondering about animals living long ago in the past. I also want to emphasize that in the sunnier and more familiar realms of the animal kingdom, there's more going on inside the animals than had been suspected, even recently. For example, self-control or delay of gratification, something that philosophers have quite often asserted is simply not present in nonhumans, has now been found in quite a few – in several birds, in primates, and even in cuttlefish.<sup>29</sup> We've seen the anticipation of the future in the preparation of tools. Dreaming has become another important topic; the evidence for dreams or dream-like states in a range of nonhuman animals is now quite good. This seems to be part of a toolkit of "offline cognition," with which animals form skills and explore possibilities for action. David Peña-Guzman gave a talk at the Pacific APA the week before this lecture, following up on a book he recently published about animal dreams.<sup>30</sup> He talked about the wide range of animals who seek out mind-altering drugs and seem to do so for hedonic reasons. As Kristin Andrews noted in discussion after his talk, this is a case of evidence for positive valence-seeking in animals, complementing all the work that's been done on pain. On the ethical side, modern industrialized farming, "factory farming," remains the most pressing ethical issue in this area. Its abuses are so appalling, and so visible from many ethical perspectives. That awareness can remain in place as we

---

<sup>29</sup> See Schnell et al., "Cuttlefish exert self-control in a delay of gratification task," *Proc. R. Soc. B*, 2021. The paper also has references to studies of other animals.

<sup>30</sup> The book is *When Animals Dream*, 2021.

also recognize, through a journey like the one we've done here, a world in which simpler forms of felt experience and agency are all around us.

---