Limits of Sentience

Peter Godfrey-Smith

University of Sydney

Whitehead Lecture 1, 2023, Harvard University

1. Introduction

Whitehead is famous for two contributions to philosophy: for *Principia Mathematica*, written with Bertrand Russell, and, in his later work, for his "process philosophy" (as it's now called) – an ambitious metaphysical picture featuring the primacy of processes and the importance of "continuity," to use a term that was also given a prominent role by James and Dewey.[1] In that respect, there's some kinship in outlook with these lectures. A more exact link to earlier Harvard generations can be made with this fairly well-known passage from William James:

> The demand for continuity has, over large tracts of science, proved itself to possess true prophetic power. We ought therefore ourselves sincerely to try every possible mode of conceiving the dawn of consciousness so that it may *not* appear equivalent to the irruption into the universe of a new nature, non-existent until then.[2]

[1] This is the text of the first of the two Whitehead Lectures given at Harvard University in April, 2023. The second lecture – "Boundaries of Consideration" – together with this one, can be found here: https://petergodfreysmith.com/philosophy/mind (at the top of the page). In the case of this first lecture, the text is close to the talk as it was given, with minor edits and some additions in the footnotes. In the case of the second lecture, I have rewritten and expanded some passages. I am very grateful to the Harvard Philosophy Department for inviting me to visit and give these talks.

[2] *The Principles of Psychology*, 1890, chapter 6.

That is my first topic in these lectures – the "dawn of consciousness," as James puts it. What would count as an "irruption" of a dubious kind, as opposed to an ordinary origin? How do those questions about faint and early cases relate to the animals we find around us now?

From there, I'll move to questions in ethics, moral philosophy, handling those topics with the same interest in borderlands and beginnings. The connection between the two topics comes from the idea, very often expressed, that there's a tight link between moral considerability and being sentient or conscious. I'll look at several versions of this idea, not just in utilitarian philosophies, where its role is pretty much immediate, but also within some non-utilitarian treatments of animals and ethics, in Christine Korsgaard and Martha Nussbaum's recent books. Those are approaches based on the primacy of *agency* rather than experience in ethical questions, but they include a role for sentience. All those views make it important to work out *who* is sentient, where it begins and ends. So the topics of the two lectures are "Limits of Sentience; Boundaries of Consideration," a combination of philosophy of mind and some moral philosophy, in that order.

I'll approach the topics in the philosophy of mind by looking first at the evolutionary path, and then looking explicitly at what kind of physical/biological basis there might be for felt experience. This will include quite a bit of speculation, but as well-anchored as I can make it in evolution and facts about the lives of animals we find around us today. The last part of this first lecture will be about origins, gradients, and the like.

One of the themes all through is the role of various kinds of *gradualism*. In evolution, the origin of subjective or felt experience is likely to have a gradual character. This suggests that looking backwards in time, the distribution of this feature is likely to have a gray area; there's no "lights on" moment in the history. Then, not necessarily but probably, we can expect a *graded presence* looking across animals alive now, too, with cases that are not a determinate *no*, or a determinate *yes*. Initially, this looks like quite a reasonable place we might end up in the case of animals like earthworms, tiny marine arthropods, maybe corals – perhaps even fungi and plants? A number of philosophers have argued explicitly, though, that "phenomenal consciousness" – basically, felt experience or sentience – cannot be a graded matter with respect to its presence; the historical transition has to be sharp, and the present borders have to be sharp also. There are views with a role for gradual change that avoid such objections; you might think that there's an initial yes/no question about a minimal kind of sentience, and then lots of gradations and differences of degree from there. I'll refer to those views as *weak gradualism*; there's a discreet hop onto

the escalator, and a smooth ascent after that. The harder questions arise from strong gradualist views, where there's no discrete step at the start.

I say "gradations," and a spatial metaphor of higher and lower has already crept in. But I assume, all through this, that we are not dealing with a simple scale. There's no new "chain of being" in the picture. I assume there are many dimensions of relevant differences between cases, and talk of "more versus less" will often make sense within those dimensions. The big contrast here is with yes-or-no treatments.

Lastly, in this introduction, a word about terminology: I'll move backwards and forwards between various related terms – sentience, consciousness, subjective experience – according to context. Differences between these will come into focus occasionally – the role of pleasure and pain in talk of sentience, especially. When it matters, I will sharpen things up. It's worth saying at the outset that there's one sense of the word "sentient," where it just means able to *sense*, where this can be understood without any implied element of feeling. That is not the sense of sentience I'm talking about here. In some ways, the phrase *felt experience* is the best term for the general topic. We're talking about the capacity for felt experience of any kind. And often I'll just talk about *experience*, meaning roughly the same thing that people talk about as phenomenal consciousness, subjective experience, and so on.

2. Animal Evolution and Agency

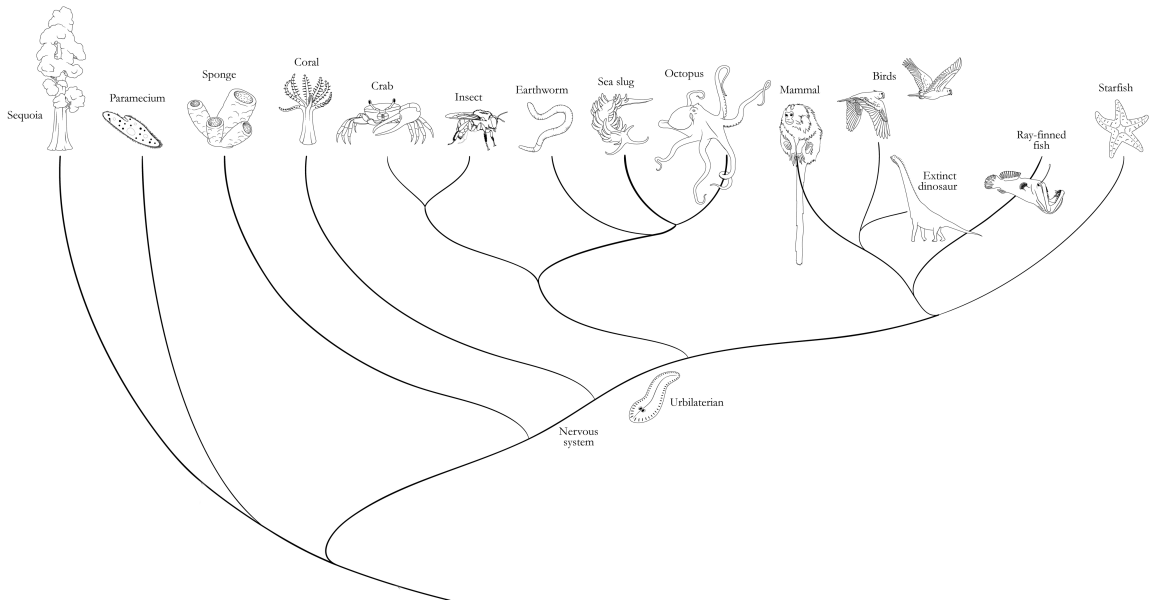
Most of this first lecture will be about animals. I'll look at plants and some other cases in the second lecture, but animals today. Before we get to animals, though, I want to put on the table some features seen in all cellular life as far as we know – in single-celled life, plants, animals, and so on.

The origin of life was in large part the origin of ways of containing and perpetuating some otherwise improbable forms of order, by taking in raw materials and using sources of energy to control chemical reactions. That already implies the evolution of *selves* of a certain kind – of units that are bounded, distinct from their environments, but also with traffic, a to-and-fro, across those boundaries. As things went, it also led, apparently very early, to what's now referred to as *minimal cognition.* This might be a universal, certainly a very general, feature of cellular life. Minimal cognition is a family of

relatively simple abilities to sense and respond to events in an adaptive way. [3]This seems to be very old, dating probably from several billion years ago. The idea of sensing, or a "sensitive soul," as coming later in the history seems possible in principle, but does not seem to be historically true. Perhaps this is not surprising; in life, boundaries are inevitable – self and other – and traffic is inevitable across those boundaries for thermodynamic reasons. Some of this traffic seems to have acquired an informational role quite quickly.

With that as background, on to animals. Animals evolved from a group of *protists*, single-celled organisms made of complicated, eukaryotic cells. Many protists live in loose collectives. In one evolutionary line, there was a move towards a more integrated style of living, featuring a particular combination that is distinctive to animals. This is the combination of action and a multicellular scale. The resources of those pre-existing protist cells included a contractable internal skeleton. This enabled an earlier, and ongoing, world of microscopic action and perception. In animal evolution, that invention, the cytoskeleton, was used on a new spatial scale, to coordinate whole body movement and action.

This chart shows a fragment of the total "tree of life" – it is not always a neat tree shape, but mostly so around animals.[4]



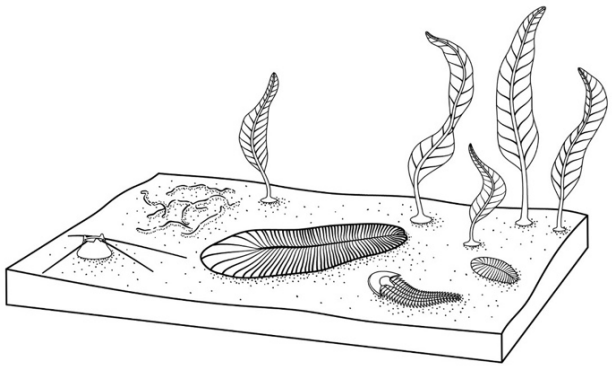
[3] See Lyon, "The cognitive cell: bacterial behavior reconsidered," *Front. Microbiol.* 2015, though Lyon does not like the term "minimal cognition" – why not just say "cognition"?

[4] This figure was drawn by Rebecca Gelernter.

Time goes up the page, with branching events at the common ancestors of various groups we can recognize today. At the top, we have the organisms around us now.

Nervous systems evolved early, and are seen now in nearly all animals (their likely origin is marked on the figure).[5] If we try to imagine the first roles of nervous systems, it's tempting to imagine something like a reflex arc, a simple sensorimotor response. But this might not be the best way to think about the roles of early nervous systems. Instead, we might think of them as having a role of enabling systemic coordination, *pulling the body together,* making it possible for this new invention, action on a multicellular scale, to occur.[6] I think of this is characteristic of the animal way of life, the animal way of being.

Very early animals, from the fossils we have available, don't look very active at all. They seem to have been slow movers or just anchored. (Here is a reconstruction of life about 560 million years ago.) [7]



What I think of as the distinctive animal combination – action on a large multicellular scale – became more visible in the Cambrian period, from about 540 million years ago, a period of coevolution of better perception in animals and real-time action. In the Cambrian, predation becomes conspicuous. So do animal bodies with parts designed for sensing and action. This seems to be have been a time of an opening to the world through the senses. Arthropods, the group that now includes insects, crustaceans, and spiders, apparently led the way in this process, and arthropods continue to have a particular *style* as animals – with hard external parts, toolkit bodies, much scope for action

[5] My tree diagram does not include Ctenophores, or comb jellies – the wild cards of early neural evolution. Some researchers believe that Ctenophores are further from us on the tree than sponges are, and given that ctenophores do have nervous systems, those must have evolved twice. For a recent discussion, see Burkhardt et al. "Syncytial nerve net in a ctenophore adds insights on the evolution of nervous systems," *Science*, 2023.

[6] See Jékely et al. 2015, "An option space for early neural evolution," *Phil. Trans. B*, 2015; Detlev Arendt, "Elementary nervous systems," *Phil. Trans. B,* 2021.

[7] This figure was also drawn by Rebecca Gelernter and appeared in *Metazoa*. As discussed in that book (chapter 3), it's possible that jellyfish were a more active but now-invisible part of this stage in animal life.

and manipulation. They seem to have been the first group of animals to achieve a kind of dominance. After the Cambrian, a group of molluscs became big and active; there was a first cephalopod rise, in giant shelled form. These animals then almost disappeared. Cephalopods later made a comeback, but never dominated ecologies again. A more minor player during these early stages was vertebrates, our group, in the form of fish. The vertebrate design features more centralization of nervous system than those other two. In fish themselves, there's less scope for action, other than swimming, but the vertebrate design went on and took new forms, especially when some animals made their way onto land.

That's a rapid historical sketch of animal evolution. Here are a couple of relevant features of all this for the project we're looking at today. First, there were multiple origins of behavioral complexity. Many of the deep branches in the animal tree occurred before the Cambrian. Because these branchings had already occurred when the Cambrian began, a more active lifestyle evolved separately in several lines, particularly in the three groups I mentioned: arthropods, like insects, vertebrates, like us, and cephalopods, a small group within the mollusks. Their common ancestor is back during the quieter time.

That multiple-origin picture though, is modulated by a recognition of gradualism or the likely importance of gradual change. Nervous systems evolved earlier; they are a shared inheritance down the three special lines, and others. So we have the three lines with something apparently special, and also the shared, inherited, earlier-evolving resources, especially nervous systems, and sensing and action itself.

3. The Biology of Subjectivity and Experience

Next, I will move explicitly to questions about the biological basis of felt experience, the mind-body problem in classic form. I do accept that there is a problem here; it's not just an artifact of bad philosophical pictures and habits of speech. I agree with some deflationists, that there are ways of *talking* about both minds and the physical that make the problem seem worse than it is – the reification of *qualia*, and so on. But there is something to grapple with here, and the solution will probably involve novel ideas on both the philosophical and the scientific sides.[8]

[8] My main discussion of the critical side is "Evolving Across the Explanatory Gap," *Philosophy, Theory, and Practice in Biology,* 2019.

At this stage, we know a lot more that is relevant than we used to know. But giving a definite answer to the main questions, especially the "where does it stop?" questions, is not something that can be done with a lot of confidence. What I want to do in this talk is push forward as much as I can, without emptily going beyond what it's reasonable to defend. So in the rest of the talk, I'll outline one view of the biology of felt experience, a view where it has a broad presence in animals. I'll contrast this with another approach that's often now preferred in neuroscience and psychology, also in parts of philosophy. Then, with the positive view in hand, I'll explicitly look at those questions about originations and transitions that bother James, in that passage I had at the start.

My view of the biology of experience has two parts, with uncertain relations between them. First, there's a set of features that have a broadly cognitive or functional character. One of the familiar ways of picking out the explanandum here (the thing to be explained) is "subjective experience." I see that phrase as a helpful one; it points us in a good direction. Subjective experience is experience of a subject. Subjects themselves are not distinct evolutionary products in the way that birds or ants are. But a system can become, through evolution, the kind of thing that occupies the role of subject. This involves the formation of a point of view or perspective on the world. That, in turn, involves becoming a kind of nexus of incoming and outgoing causal lines. It involves some integration of sensory paths and also, pretty uniformly, some kind of integration of present experience with traces of the past in memory. I think views of consciousness or felt experience that turn away entirely from this cognitive side (such as "integrated information theory" or IIT) are ignoring obviously relevant features – the subjects in subjective experience.

This is part one. And then we can ask, why did those point-of-view related features evolve? *Agency* is part of the animal way of being. That's what's most relevant to evolution in this setting – effective action. But coherent agency requires sensing requires fitting actions to the state of the world. It also involves evaluation of some kind, even if only tacit. It's imposing a little more separation on these features that I think is right, but, roughly speaking, the evolution of agency brings with it the evolution of subjectivity.

Those features – subjectivity reflected in agency – are seen conspicuously in the three "special" evolutionary lines or groups that I mentioned – arthropods, like insects, cephalopods, like the octopus, vertebrates like us. In these three groups, and nowhere else,

high resolution vision (what are called "class IV" eyes) evolved.[9] All this came to exist within different kinds of bodies, with different brain architectures serving different lifestyles, but in each case enabling targeted behavior and usually a good deal of movement.

Each of these groups has their own quirks. In octopuses, along with an extraordinary sensitivity to events, there's a particular orientation to the world, at least in some species – an attentiveness, an interest in novel objects – and with their extraordinary bodies, a tendency to forever perform new behaviors. Even in some of the more reticent, restricted views of animal consciousness I will mention later, that attentiveness and ability to produce novel behaviors is often seen as a mark of consciousness.

In the case of arthropods, we have a combination seen across different groups. Bees are sensorily acute, good learners and also problem-solvers. Due to the work of Robert Elwood and his collaborators, some of their crustacean relatives, marine crustaceans such as crabs and lobsters, have shown a subtle and complex handling of the evaluative side of subjectivity – pain-like states. This is seen in trade-offs between different kinds of aversive experiences, and their relation to risks. In an experiment that Elwood's group did, small electric shocks were found to induce a hermit crab to leave a shell. But this is conditioned by the value of the shell, and also by the level of perceived risk around at the time – you'll put up with more of this aversive event if the shell is more important to you.[10]

In cases like this, given the animals' behavioral facility, when we look closely we see additional apparent indicators, or at least features suggestive, of sentience. Hermit crabs don't only do this complex shell trading behavior, but also respond to the shocks with a thorough inspection of their shell: what on Earth could be doing this unusual thing to me?

---

[9] For eyes, see Nilsson, "Eye evolution and its functional basis," *Visual Neuroscience,* 2013.

[10] See Elwood, "Evidence for pain in decapod crustaceans," *Animal Welfare*, 2012, and "Hermit crabs, shells, and sentience," *Animal Cognition,* 2022. These findings are reassessed in Birch et al.'s report, "Review of the Evidence of Sentience in Cephalopod Molluscs and Decapod Crustaceans," https://www.lse.ac.uk/business/consulting/reports/review-of-the-evidence-of-sentiences-in-cephalopod-molluscs-and-decapod-crustaceans. That discussion is somewhat cautious about the trade-off results, less so about some other findings. All this work is discussed in more detail in my Jean Nicod lectures: http://www.institutnicod.org/seminaires-colloques/prix-jean-nicod/?lang=en.

We have octopuses, with that sensitivity and interest in novelty; also arthropods with their toolkit bodies and an initially hidden but real sensitivity; and the third line with these conspicuous marks of experience is our own. This is all in "element one" of my treatment of the biology of experience. The other element of my view involves some features of nervous systems themselves, and a picture of nervous system activity.

We often think of nervous systems as networks of relays and switch-like processes, with "spikes," or action potentials – discrete firings – occurring in neurons. all a matter of networked, cell-to-cell influences. There is also a collection of more diffuse, large-scale dynamic patterns in nervous systems, especially oscillations, rhythmic electrical activity spanning large parts of the brain.

The basis for these patterns is controversial, and everything I say in the next couple of minutes is probably more controversial than other parts of the talk. But these rhythms involve coordinated movements of ions across cell membranes, below the threshold that initiates a chain reaction and a "spike" or a "firing" of that neuron. These, or partly these, give rise to EEG patterns in animals like us, with distinctive brainwave patterns seen in sleep, in attentive states, in a relaxed but awake states, and various others.

Some decades ago, Francis Crick, Christof Koch, and some others conjectured that these large-scale dynamic patterns have an important role in conscious sensory experience, especially as an integrating mechanism.[11] They were thinking back then about the human case. Something that came as a surprise to me is the fact that these patterns are not specific to mammals, or vertebrates and so on, but are all over the animal kingdom. The neuroscientist Mac Passano in the 1960s studied and made conjectures about rhythmic patterns within *Hydra*, which are very small jellyfish-like animals.[12] He conjectured that the patterns he was observing in this very small nervous system (in a present day organism) might be telling us something about the ancestors of the brainwave patterns in more complex animals like us. The differences in brain architecture that I mentioned earlier, separating us from octopuses, and so on, don't prevent these dynamic patterns from being seen all over nervous systems in animals. We find some similarities in role, too. Bruno van Swinderen is an Australian neuroscientist who's influenced my thinking here.

---

[11] Crick and Koch, "Towards a neurobiological theory of consciousness," *Seminars in Neuro.,* 1990; for the oscillations themselves, see Singer, "Neuronal oscillations: Unavoidable and useful?" *Eur. J. Neuro,* 2018.

[12] Passano, "Primitive Nervous Systems." *PNAS*, 1963.

He and Ralph Greenspan found that a particular wave pattern is an indicator of "selective attention," or something like selective attention... in a fly.[13]

This, to me, was surprising. Building on that kind of thought, we can be led to a slightly unorthodox view of nervous system activity itself. A nervous system is an endogenously, or spontaneously active, system, prone to forming large scale oscillatory patterns. It also has those targeted, cell-to-cell influences – computation-like processes going on. And along with all this, its activity is modulated by sensory and other events. I would suggest that part of the peculiar glory of the nervous system as an evolutionary invention is this combination of targeted cell-to-cell influences, and the less local, inherently coordinated, especially rhythmic, side of what they do.

Something that's been noted in different forms in this area by several people – Passano himself, writing way back, Kaplan and Zimmer in a recent paper, and also Ginsberg and Jablonka – is a kind of *integration for free* that you get in his picture of neural activity.[14] When we have large-scale activity modulated by events in this way, even in quite a simple nervous system, the state of activity will be modulated simultaneously by several kinds of events – by sensing deriving from the external world (exterosensing), internal sensing, also by the overall state of the organism itself – what kind of condition it's in – and so on. I suggest that this kind of naturally integrative character in nervous system activity has a mapping to what might be called – I grope for the right word here – a kind of *textural* feature of experience, at least in us.

Here is what I have in mind. Suppose we sort of take a step back from the science for a moment, and ask: What is "in" experience? What is felt experience like? The tendency in recent philosophy and psychology has been to emphasize *selectivity* – conscious experience contains a small sliver of what's going on inside us – and also the idea of discrete items, or "contents." Things are either experienced or they're not. They're either conscious or unconscious, item by item. Stanislas Dehaene, an influential French neuroscientist, defends an extreme version of this picture. He argues on the basis of empirical work that felt experience – in humans, at least – contains one item at a time. A

---

[13] See van Swinderen, "The remote roots of consciousness in fruit-fly selective attention?" *BioEssays,* 2005; van Swinderen and Greenspan, "Salience modulates 20-30 Hz brain activity in Drosophila," *Nature Neuroscience*, 2003.

[14] Kaplan and Zimmer, "Brain-wide representations of ongoing behavior: A universal principle?" *Current Opinion in Neurobiology*, 2020; Ginsburg and Jablonka, *The Evolution of the Sensitive Soul,* 2019.

fast switching between items gives the appearance of a multiplicity. Other people, less extreme than Dehaene, say there can be a handful of things going on in there at once.

That picture is coming from the experimentalists. Recent philosophical responses to this, I think, have been a bit like this. People have thought: well, if that's what the experimental work shows, then I guess that's where we are. And also, perhaps there are some everyday phenomena that support that highly selective picture. Some decades ago, David Armstrong introduced the case of an inattentive long distance driver – steering, making small adjustments, doing basic navigation, without being aware of what's going on – and then suddenly "coming to" and realizing that he's driven quite a long way in this inattentive state.[15] This case, in recent discussions, is taken to be friendly to view a bit like to Dehaene's; the driver's experience at each moment is occupied by *either* the road *or* the radio, or a daydream or a plan. The road is outside of experience entirely, until it's suddenly attended to.

When I said that last bit, I hope that, if you hadn't already been feeling some resistance, then at least some of you began to resist. Here is an alternative, and, I suggest, very natural way of describing cases like this – describing them as they first appear. A driver has, at each moment, what we can call an *experiential profile*. That profile typically includes something that's in attentional focus, and a lot more that contributes to experience while being to various degrees in the background. This can include sensory elements that are not being attended to – the road, the car seat, the temperature – also a mood, and an energy level. These all contribute to the driver's experiential profile, even if the driver is attending entirely to tomorrow's plans. If they were not driving along *and* dealing with that road *and* doing so a certain number of hours from their last meal, on a seat with that shape and hardness, then it would *feel different to be* them at that moment.

Energy level seems an especially clear case. It's an aspect of experience that is rarely in attentional focus. But it does make a difference to how things feel.

If we think of this profile-like feature as a general feature of experience, not just a feature in humans, then it maps naturally to the picture of nervous system activity that I mentioned a moment ago.[16] We might then envision the evolutionary history like this. Nervous system evolution in early animals was driven by a coordinative role – trying to get the body to hang together. to do stuff together, to produce coherent action. But the

[15] See Armstrong, "What is Consciousness?" In *The Nature of Mind and Other Essays*, 1981.
[16] This is discussed in more detail in my "Gradualism and the Evolution of Experience," *Phil. Topics,* 2020.

kind of activity established was always being modulated by events, was always sensitive to influences, even if just *de facto*. In the Cambrian, there was a new premium on this openness – on the detection of events and making responses to them. And if you ask, at least from this stage, in our three special evolutionary lines, why there should be "something it's like" to be one of these organisms to have this going on – why there should be something it feels like to have an endogenous large-scale pattern of nervous system activity, modulated by events and also put into service into the service of action from a point of view – then to me, there's not much of a gap here anymore. Once such systems exist, it should feel like something to be them.

This picture, where felt experience is a natural concomitant of some core features of the animal way of being, is not where a lot of the recent scientific and philosophical work has ended up. According to this research tradition, one thing we've learned since the late 20th century, is that while views like that might be initially appealing, they are not sustainable. This is because much of the basic business of sensing, cognition, and action in humans can occur without consciousness being involved. Here is a quote from the psychologist Chris Frith -

> The key insight driving research on this topic is that much precise and well-adapted behaviour in a fully awake person can occur without consciousness.[17]

If you're guided by this view, then what we need to understand and study is a *contrast* between mere competent sensing and behaving, and so on, and a special kind that's conscious – a special kind of processing or routing of information, through some particular kind of brain. That routing may involve working memory, or an "attentional spotlight" that can be put on particular bits of information – features that go well beyond what might be seen as a basic toolkit for a behaving animal.

The term I use for this cluster of views in neuroscience, philosophy, and psychology is *narrow pathway* views. They tend to be narrow in a couple of linked senses. Dehaene is a perfect example of a narrow pathway view. In the brain, there's a particular path or a configuration that gives rise to experience. This can only exist, as far as we can tell, in brains of a certain kind, like ours. It's natural then to think that consciousness itself will be an evolutionary late-comer. It's on a narrow evolutionary path as well, rather than being a broad feature of animal life.

---

[17] "The neural basis of consciousness," *Psychological Medicine,* 2019.

There's a lot of experimental work that can be used to support such a picture. This is the experimental paradigm based on blindsight, dorsal stream vision, and the subliminal processing of stimuli that come in too quickly for us to be aware of them, but that do have an effect on our brains and behavior. In the case of blindsight, due to a brain injury, a person can manage visually-guided behavior, while denying any visual experience. "Dorsal stream" vision refers to one of two streams of visual processing inside us. A woman lost one of those streams (the other one) due to an accident, and retained the ability to do many kinds of visually guided behavior, while professing to be nearly blind, and being unable to describe objects around her. That led to the positing of a conscious and an unconscious stream of visual processing, where the unconscious one is still *visual,* you're still using vision to navigate around the world.[18]

The empirical side of this work has to be accommodated by any viable view – you can't pretend it's not there. But the interpretation of the work is in play. I'll spend a few minutes on how these empirical discoveries relate to my picture.

Here's how I think people tend to think about the situation with these findings, especially in relation to questions about nonhuman animals. They think something like this: There are lots of things that can be done unconsciously in humans – certain kinds of perception, the guidance of action of certain kinds.... The various capacities that we've learned can be done unconsciously in humans could be put together, it seems, to yield a package, a total cognitive and behavioral profile, that might be roughly like what we see in various kinds of nonhuman animals. We know that each of these things can be done unconsciously, so it seems that the package could also be unconscious. And, you might then infer, it probably *is* unconscious in lots of nonhuman animals.

That's a of train of thought it's easy to fall into, when reading about this kind of experimental work. But in the experiments that motivate this kind of view, the human subjects are all conscious. It's not that they're doing all these things while unconscious, it's just that they aren't conscious of everything they're doing. There's no reason given by this work to think that because at any time, *some* of what we do is done unconsciously, a basic total combination of activities could be *all that way at once.* It might instead be that once you're an awake human subject, doing the sorts of things that people do, *some* of it has to be conscious – or more exactly, you'll have some kind of experiential profile. For any

[18] See Milner and Goodale, *Sight Unseen*, 2005.

normal and wakeful human being, there's something it's like to be that person, even if some processing is being done deep in the background.

What I think we learned from these experiments over the last few decades is that experiential profiles can be very different from what might have guessed they should be like, when people carry out various tasks. That is something we have to grapple with. There are surprises in the literature. But the idea that we're learning about the possibility of a complex and totally unconscious behavioral profile – that's a different thing. We have not yet learned about the possibility of that.[19]

The last thing I want to do in this section is to push back more specifically against the picture that one ends up with, within narrow pathway views, and the relation between where you end up, with these views, and the evolutionary story as I've told it so far.

My account is guided by the idea that experience-related features, as far as we can initially tell evolved in animals with different histories, different bodily resources, and different nervous system architectures. Octopuses have no visual cortex, but they can clearly see. A visual cortex can't be required for vision. In that case, a person might say: we know that octopuses can see, but maybe not consciously, if they lack a visual cortex. They could say this drawing on what we've learned about unconscious dorsal stream vision. Particularly significant in response to that kind of thinking is a paper that came out in 2021.[20] This is the best study so far on pain in octopuses. It was done by Robyn Crook at UCSF. I am switching from vision-like capacities to pain now, but I think the general message will be clear. In octopuses, we get an initial impression that injury produces acute pain, in some cases. A bite from a fish leads to what you'd expect it to lead to, if it was painful. It's then natural to ask: Is this misleading? Is it just our over-rich interpretations? The 2021 Crook paper studies this phenomenon from several directions. What's important here is the way that the lines of evidence interlock and work together.

Here is what Crook did. Acetic action (vinegar) injection into an arm was the stimulus – just once, a little injection of vinegar. This showed up in a number of linked

---

[19] Here is a more exact way of putting it. The data might show that you can do X_1 unconsciously, while being conscious of something else, and can do X_2 unconsciously, while being conscious of something else.... The conclusion that would not follow is that you can do X_1, X_2,... (etc.) all together, without being conscious of *anything*. This is covered in more detail in my Jean Nicod Lectures (online for now, a book with MIT Press to follow).

[20] "Behavioral and neurophysiological evidence suggests affective pain experience in octopus," *iScience*, 2021.

responses by the octopus. Firstly, it showed up in its behavior in relation to choices about places – what's called "conditioned place preference." The octopuses have initially preferred a particular location, and avoided that location when the bad thing had occurred in that location. You might say: conditioned place preference – big deal; that's been seen in flatworms. But here are further parts of the study. The octopuses also wound tended, groomed the injured part. Wound-tending is fairly rare in animals far from us, invertebrate animals. Partly that's because it's probably hard to *do*, with many of the bodies of animals far from us; it's not something all sorts of animals can even attempt. But some animals might do it, it seems, and don't, while octopuses do wound-tend. Thirdly, there was another behavioral choice. An area or a chamber in the tank where analgesic drugs were given after the injury was preferred, though this location was not preferred otherwise. So a second place preference was established by giving the animal what appears to be a pain-killing chemical. (It's surprising how many drugs that work as work as painkillers in us seem to work as painkillers in various other animals.) There was also some physiological evidence; they looked at what was going on in the nervous systems of the upper arms in these various conditions – with the acetic acid, with the painkiller, and so on.

In this study, we have a behavioral first impression and a vague sense of the role that pain might have in octopus life. We recognize that this could be defeated; there might be something more we could learn that would undermine our impression, when we look more closely. We might see something that's very peripheral and reflexive and not suggestive of experience. In fact, we do not see that. We see a kind of integrated, multifaceted handling of the aversive stimulus, one that's quite similar to profiles of behavior seen in fish and in chickens. It's the coming together of various responses to the aversive stimulus that suggests that there's pain, or something like it, present.

It also suggests that what philosophers call *multiple realizability* of the basis for experience in different neural architectures is present. It suggests that you can have different architectures in nervous systems that serve the same function, that mediate the perception and handling of aversive stimuli, and are the basis for aversive experiences. It suggests that different brains have the basis for pain. Going back fewer Harvard generations than William James, Hilary Putnam guessed this for the case of octopus pain, In a paper published in 1967. He said: if octopuses feel pain (and he was confident they do), then it's going to be a different architecture that makes this possible.[21] Feeling pain

[21] Putnam, "Psychological predicates," 1967.

won't be always a consequence of having the same kind of neural structure that we find in us. That was one of the first papers on what is now called multiple realizability. The example, only empirically solid from 2021, is about as old as the idea of multiple realizability itself.

4. Limits, Gradations, Irruptions

Suppose the picture I have outlined is okay, despite all the many uncertainties, and we move ahead within it. What might be the distribution of felt experience of sentience? Where is it found? We've already come through a series of previous expansions of the limits of sentience. These would be accepted, taken on board, in this picture. Fish made it into the club especially through Victoria Braithwaite's work.[22] Cephalopods, I've just talked about. I've talked about crustaceans in the context of Elwood's work. The more recent frontier has been insects, and I think they are probably no longer the frontier; I think there's reason to see them as *in*.[23] I'll look at this in more detail in the second lecture, again in the particular context of pain.[24]

The frontier, the difficult cases, are now elsewhere – other invertebrate animals with smaller nervous systems: gastropods like snails, earthworms, flatworms... and then animals that don't have our bilaterian or left-right symmetrical design, like anemones and corals. It's hard to imagine it in the *Hydra* that I mentioned, but I'll come back to that. And then we have to think about other kinds of life – protists, plants, fungi, bacteria, and also artificial systems. All that now has to be thought about, but I'll stay with animals for now, and next look at a general point introduced briefly at the start of this paper.

If we look back at the phylogenetic tree, the tree of animal life, there are two dimensions there. We have the history, going back through time, and we have the present range of animals, looking along the top. In a Darwinian context, traits that are complex arise gradually, for the most part, and in a broad sense of that term. "Gradual" here does

---

[22] *Do Fish Feel Pain?* 2010.

[23] Andy Barron and Colin Klein were well ahead here: " What insects can tell us about the origins of consciousness," *PNAS* 2016. See also van Swinderen, "The remote roots of consciousness in fruit-fly selective attention?"

[24] Part of the shift we are seeing is a result of moving back towards something sensible, away from excesses of restriction and anthropocentrism. But it's not just that. We are not just going back to a broader view, but one that is surprisingly broad.

not mean at a uniform speed. It means that the feature is built up piece by piece, rather than in a sudden jump that introduces a lot of new complex structure in one move. Let's assume that evolution in this case is gradual. That suggests that we have a kind of *graded presence* looking back into the past. There will be cases in the past where there were animals that had *some* of what it takes, but it's hard to tell if they had *enough*. We can then look across the top of the tree. A gradualist view of the history suggests a graded presence of sentience now. This is not automatic, because no present-day simple animal is an ancestor, and need not be especially similar to any ancestors. But looking sideways along the top, we do see a lot of diversity in nervous systems and behavioral complexity. This suggests that graded presence is likely on both dimensions.[25]

Some people in both philosophy and science have suggested, partly in response to this kind of initial thought, that graded presence, either in the past or now, is just not possible - that phenomenal consciousness is a yes or no matter, in principle. A few recent papers have argued this.[26] Startlingly, it has also led Michael Tye from a fairly standard form of materialism, developed in lots of earlier books, to a form of panpsychism. In his most recent book, called *Through the Looking Glass,* Tye says there's an aspect of consciousness that can't arise from its absence; it can't arise from something else. So given that it's around in us now, it has to have been around forever, as a basic feature of reality.

Here is a quote from a paper by Bayne, Hohwy and Owen:[27]

> [T]he notion of degrees of consciousness is of dubious coherence. According to the standard conception of consciousness, a creature is conscious if and only if it possesses a subjective point of view.... Arguably, the property of having a subjective point of view is not gradable – it cannot come in degrees.
>
> In this way it resembles being a member of the United Nations rather than being healthy, which clearly can come in degrees.

---

[25] In the lecture I said: there are no animals today with "half a nervous system." They all either have a nervous system, or they don't, and almost all of them do. But this might be questioned in the case of sponges. See Leys, "Elements of a 'nervous system' in sponges," *J. Exp. Bio.*, 2015.

[26] For example, Simon, "Vagueness and zombies: Why 'phenomenally conscious' has no borderline cases," *Philosophical Studies,* 2017.

[27] "Are there levels of consciousness?" *Trends in Cognitive Sciences,* 2016.

These are two philosophers and a scientist, Adrian Owen. Owen was responsible for the extraordinary work on how some patients thought to be in a persistent vegetative state are in fact fully aware and able to communicate with the outside world by using their voluntary imagination in combination with brain scanning (fMRI). You tell the person to think of playing tennis if the answer to a question is yes. By looking at their brain, you can work out that they were imagining something of that general kind, and you have a contrasting imaginative act for them to engage in if the answer is no. You might think that Owen would be sympathetic to graded and gradualist views, but not so; that quote above is quite strong.

Their view is a kind of "weak gradualism," in the sense that I had earlier. There's a discreet hop onto the escalator, and gradual change of various kinds from there. But the yes/no matter is sharp.

It's not that I claim we can rule out weak gradualism at this time. It's empirically possible. Maybe some of the crucial large-scale dynamic patterns in brains can appear in a "phase transition" – in a sudden event, with only a minor change in the physical resources.[28] Sharp divides might even apply on *all* of the three scales where transitions from unconscious to conscious occur: evolution, individual development from a fertilized egg, and waking from sleep or a coma. In all three, I suppose that it could strictly speaking be sharp. But I certainly resist the idea that we can now know it has to be that way, by inspecting our concept of consciousness, or thought experiments and intuitions. Strong gradualism is certainly possible.[29]

Some of the ways we *talk* in this area suggests a sharp divide. This is part of the problem. The famous "something it's like" formulation from Nagel – there's something it's like to be you – suggests a dichotomy and in some ways the most basic dichotomy, something versus nothing. But although Nagel's wording is quite a good gesture towards the problem, it does not constrain solutions. The *something*, in "something it's like," is not

[28] A member of an online audience in a talk I gave "at" Tübingen in 2020 in raised this point. I don't seem to have kept track of her name, unfortunately – I will try to find out.

[29] The context of this passage is a discussion of anesthesia and brain damage, rather than evolution and other animals. And part of what Bayne and his coauthors wanted to emphasize is the absence of a single ordering from more conscious to less, a point I agree with. But the claims made about degrees are broader, and I disagree with those. A single ordering is not the only alternative to a dichotomous (member of the U.N., or not) situation. As Fazekas and Overgaard say in a commentary on the Bayne et al. paper, there might be many dimensions of consciousness, with gradations on each.

a *thing* that needs to be determinately called into existence. In this case, and others, quirks of language may be responsible for a range of apparent difficulties with a strongly gradualist view.

Once the origin of these difficulties in ways that we talk is seen, the right response will be to discount them, and to expect to find ourselves, when more is known, talking a bit differently, or talking differently in contexts where the details matter. I gave a talk on this topic a few months ago at Sydney, and David Braddon-Mitchell asked me: on my view, is there a *no* category? There's a yes category and there's a gray area; is there a no category – within living things, for example? I said yes. He said: okay, is there an organism who is the first one where the answer is not a definite no? I seem to be committed to a new hop onto the escalator, not a hop from no to yes, but a hop from no to *not definitely no.* Or else the view becomes panpsychist, (or something like that), because the *no* category has collapsed.

My response to this was that this also is putting too much weight on our present concepts. When more is known, they will look rough and ready in their application to phenomena. We can expect to get a more fine-grained account of the transitions, probably with several different strands underlying the differences between cases. The simple *yes/no*, or *yes/no/indefinite*, distinction will then be replaced. This is not just something I say to deflect the case, and that might be true in principle. This is exactly what happened in the case of the concept of *life*.

One can imagine, at an earlier stage in knowledge, an argument being presented that life is a definite thing in principle, which cannot exist in degrees. As more came to be known about life, including special cases like viruses, we came to see that our categories require some reorganization. Life is better understood when broken into a cluster of different capacities. Actually, I think it can be broken into just two capacities, or families of capacities. One of them involves metabolism, the use of energy and materials to maintain otherwise improbable forms of order, and the other involves reproduction and evolution. Viruses can be recognized as well and truly in the vicinity of life, but lacking some of the cluster. You might say that viruses are "not fully alive, "but you can also say it's better not to talk as if life is a definite property any more. "Alive" and "life" can remain useful as loose terms, but the varieties and the tricky cases are best handled using other language, the language of metabolism and the language of reproduction and heredity.

Just as we see in the case of life, a gradualist account of experience can take the form of an explanation of the history and distribution of a cluster of things, none of which

are named *experience* when they are being explained as strands within that cluster. In us, that cluster comes together and we call the result experience. But the elements of the cluster, as in the case of life, can each have their own histories.

I want to add one more point in opposition to the "gradualism is impossible" view. This one I got from Geoffrey Lee, in discussion at the APA conference in San Francisco last week. This is from a book he's writing, with the provisional title *Searching for the Inner Light*. Lee has quite a deflationary view of consciousness. When we talked about gradualism, he suggested that one basis for the intuition of a sharp divide regarding is that what we first do with the tricky cases – flatworms, crabs, plants, computers – is try for an imaginative projection. We try to project ourselves inside their shoes or claws. We try to guess what it all feels like for them. If we can do the projection, if we feel we've done it, then we say *yes* – they have a minimal kind of consciousness. If we can't, we say that the case is a likely no. But this test is done from the basis of our own full subjectivity. We modify, dim, and blur from that basis. Such a test is obviously not diagnostic, either about the case in question, or about the kind of question that we're asking. It makes graded presence look problematic, without real evidence.

I am getting towards the end, and have just a couple more points to make. Much of this last part of the talk has been presented in expectation of a finer-grained account. I'm mostly just gesturing towards it. I don't think we have the resources to go too far on this yet – to achieve the kind of resolution that we achieved in the case of the concept of life. We can probe a little, we have a few things on the table already. In my account of the biology of experience, earlier in this talk, I had two elements: some schematic or functional properties related to subjectivity and point of view, and some physical features of nervous systems. In animals, they're very much tied together. The neural features are how we achieve the subjectivity-related side of things. But we can ask some questions about each without the other, it seems. What if you had the schematic, subjectivity-related, properties, and no nervous system? What if you had a nervous system with these peculiar dynamic features without the subjectivity related properties? That's one of the philosophy of mind topics that I'm going to push to the second lecture. I'll finish today with just a bit more reflection on the main questions that have guided this lecture, and that I keep coming back to.

We start off looking at those three evolutionary lines coming out of the Cambrian – lines long separated, stretching through to the present. We find today that there are marks of felt experience within some animals in each of them: the octopuses with their

sensitivity and novelty seeking; the arthropods with their mixture – the intelligence of bees, the complex handling of aversive events in crustaceans – and then our own group. We try to get an account that makes sense of these cases. The case for recognizable subjective experience in each of those groups is good; I think we have good reason to believe that animals in each of those groups experience their lives. Experience, if it is present, is not due to some brain architecture shared across all these cases. There are different ways of doing things at that level. But there's also those similarities in some physically distinctive, dynamic features of neural activities. These are lower level features, in a sense – those dynamic features that figure in wakefulness and sleep across many animals, that have a link to attention (even in flies), and so on. They also link, I suggested, to the "experiential profile" phenomenon, that feature of the texture of experience in us and perhaps in others. From here, two views initially look like live options. One is where I found myself earlier, when thinking about the problem. It has something like three, perhaps a few more than three, independent origins of subjective experience in different evolutionary lines. In three lines, a transition was made from absent to present. Those evolutionary lines are special in how the animals live, and how they're set up.
The handling of puzzles of gradual appearance, the "irruptions" that bothered William James, will be handled by changes that made those animals special on each of their three different lines.

That's one option. Another option is a broader view, where we say the features that in special forms made subjective experience *conspicuous* to us, in those three lines – both the subjectivity related features, and the neural features – are found in lots of other animals, in simpler form. They shade off into simpler forms, all over the tree. There are fainter versions of what matters, lacking those sharp behavioral marks, all over the place in animals. That line of thought pushes us towards a view where the three special groups just have *more* of what the others have, including experientially, and where we conclude that the core phenomenon is more widespread.

The simple way of expressing that choice between views is "one origin" versus something like a "three origins" choice. Either the big innovation was early and then shared, or later and arising several times. I can't get away from thinking about the question that way, and I do think part of that question, as asked just then, is substantive rather than terminological. But it is wrapped up in the problem of describing liminal and partial cases. Did the common ancestor of us and an octopus have a simpler form of experience, or something that's *not* experience, but *like* it? Was it one of the ones that was *not clearly no*,

rather than being a simple case of *yes*? Once we get to that point, I think we need our finer-grain vocabulary, yet to come, to answer the question. For what it's worth, though, and with all these uncertainties and difficulties acknowledged, and conditioning all this by gradualism, I find myself thinking more along the lines of the second option than the first, and thinking more along those lines that I did before. The thought I find powerful here is the thought that (this is almost a repetition of what I said a minute ago) the features that, in special forms, made subjective experience *conspicuous* to us on those three lines, both the subjectivity-related ones and the neural ones, *are found* in lots of other animals. When we look closely, we'll find a simpler form, but we'll see a version of the same kind of stuff. I suspect that future revisions of our thinking, even with new language on board. will probably tend in that direction. I'll say a little bit more about that in the second lecture, but the goal then will be to start to link all this to questions about moral consideration.
