Animal Evolution and Subjective Experience

Peter Godfrey-Smith
University of Sydney

For the NYU "Animal Consciousness" conference, November 2017.

* The main text here is my NYU talk, minus the discussion of videos, with a few passages filled out and other changes indicated in the footnotes. The notes also add a lot of extra material.

I will first give a sketch of animal evolution and some relevant phylogenetic relationships, then locate in that framework some features that seem *likely* to matter to subjective experience, and talk about their history – did they arise once or often? Do they cluster or dissociate? I'll next outline, more cautiously, how I think everything might fit together – the rough shape of an eventual view. Lastly I'll say something about the special case of the octopus.[1]

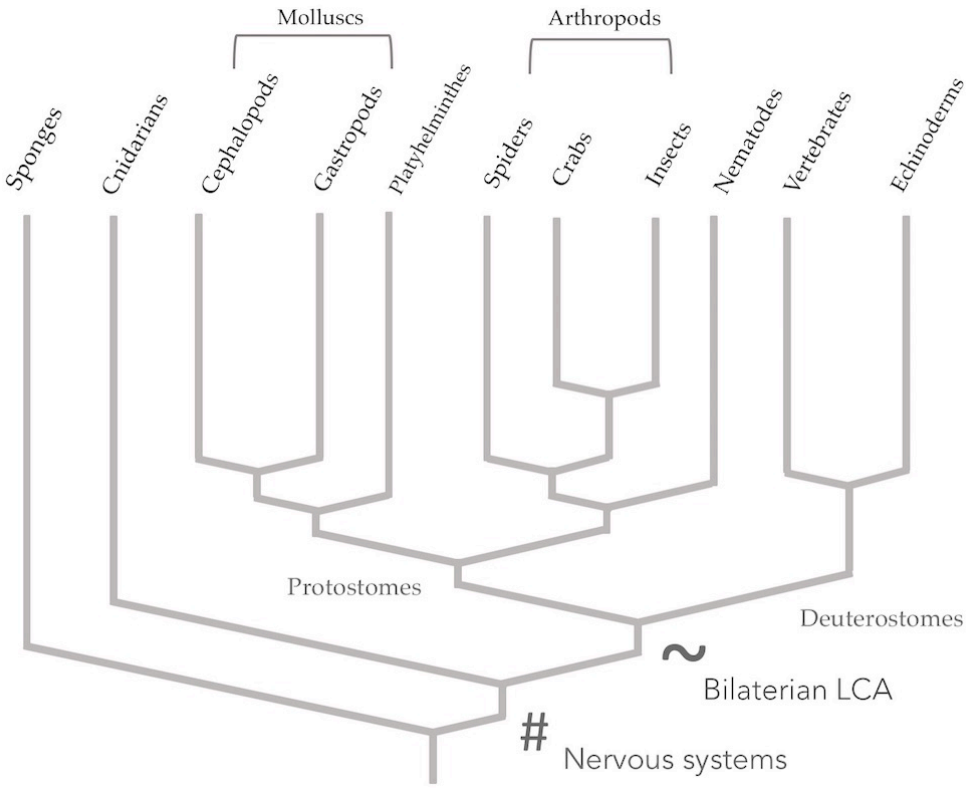
My title has "subjective experience." The conference title has "consciousness." These are often seen as same, especially round NYU, at least for one sense of "consciousness." I think this is a bit of a problem. "Consciousness" inevitably suggests a rich human form of subjective experience. Many discussions try to set that aside: consciousness is just there being *something it's like to be* an organism or other system, in Nagel's phrase. Any sort of rudimentary feeling – any vague wash of felt reward or pain – is said to be conscious. OK, but richer associations return, I think, moments after the

[1] Several other papers add detail in different areas. See "Mind, Matter, and Metabolism" and "Materialism, Subjectivity, and Evolution" on the philosophical side; "Individuality, Subjectivity, and Minimal Cognition" and "The Evolution of Consciousness in Phylogenetic Context" for more biology.

weak sense has been endorsed, and make the problem seem harder. I think "subjective experience" is better, for the basic target, but I'll speak the local language much of the time.

The history of animals has the shape of a tree; animals are a branch or sub-tree within a larger tree of life (which outside animals is not always so tree-like). The tree-based framework is used here not with any philosophical claim about the nature of species; it just represents what happened. All present-day animals are linked by common ancestry to various degrees of temporal depth. The large-scale shape of animal evolution is the formation of species by branching events, with subsequent change within species, accompanied by extinction.

Here below is a first picture of the tree. Read it with time running up the page. It is very incomplete – only the most relevant ones are included, plus extras that are useful landmarks. (Taxonomic ranks are mixed, which is not a problem in a tree-based framework).



People still sometimes talk of a "phylogenetic scale." They might ask: where in the phylogenetic scale did consciousness arise? But there is no phylogenetic scale. A scale is a matter of higher and lower. In the tree, there is higher versus lower in the sense of *earlier versus later*, but no one alive now is lower in that sense. There is also a distinction between *simple and complex*, but there are many different kinds of complexity, and no single scale.[2] A tree is just a different sort of thing from a scale, and animals form, or are part of, a tree.

Animals arose something like 700-900 million years ago. I will describe the sequence of branchings in a way organized by relations between present-day animals. One early branching led on one side to sponges and on the other to everyone, or nearly everyone, else.[3] Later, a branching occurred that led on one side to *cnidarians* – jellyfish, corals, anemones – and *bilaterians* on the other. Bilaterians or bilaterally symmetrical animals have a left and right as well as an up and down, and include most of the familiar animals.[4] Nervous systems may have evolved once or perhaps twice. Everyone (or nearly everyone) on the big branch that includes cnidarians and bilaterians has a nervous system.[5] Sponges do not have nervous systems.

The bilaterian line split into two sides, the *protostomes* and *deuterostomes*. We are deuterostomes, as are starfish. Protostomes include most of the familiar invertebrates – insects, crustaceans, molluscs, worms. The "bilaterian LCA" on the figure is the *last common ancestor* of living bilaterian animals.

Many of the branchings mentioned so far probably took place in the *Ediacaran*, the first period with any fossil record of animals, 635-540 million years ago (mya). This was a time of ambiguous forms, many of which lived fixed in place or drifted, while a

---

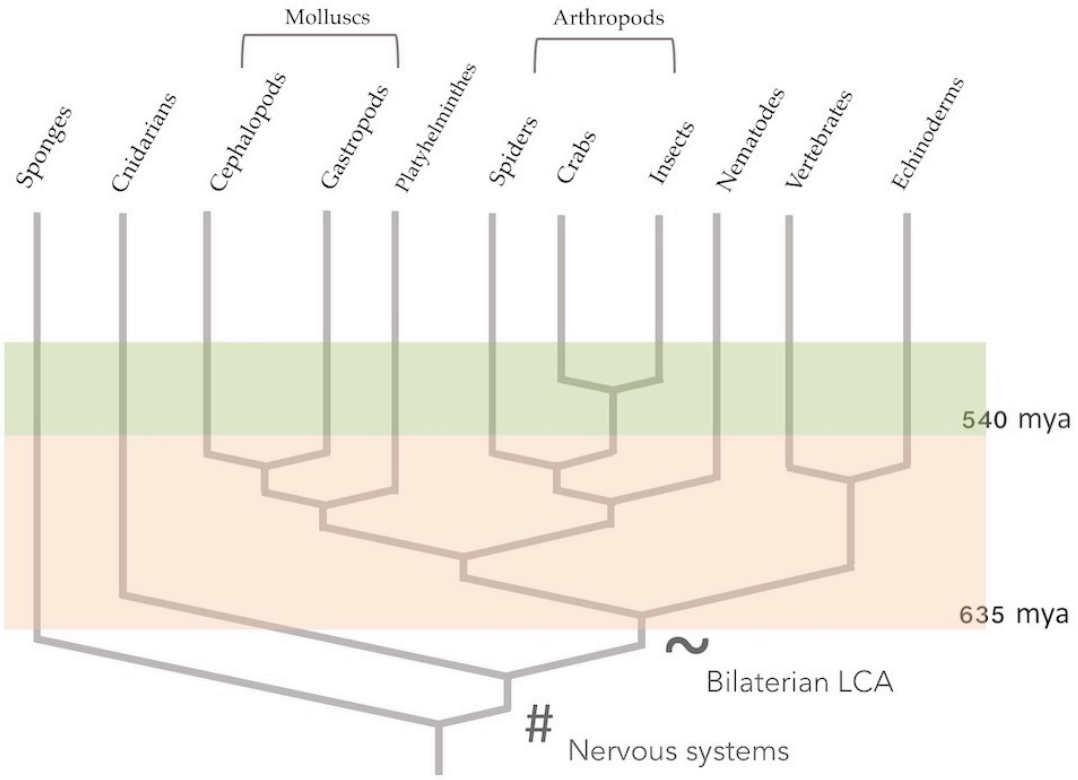
[2] Perhaps the 'scale' is across the top? But apart from anything else, left-right flips over a branching are meaningless in the figure, as my placement of starfish outside vertebrates reminds us. There is also no combination of horizontal flips that brings both nematodes and platyhelminths, two of the neurally simpler worm-like animals on the chart, close together over to the left hand side; they are just not very closely related. "Lower" sometimes seems to mean m*ore similar to an ancestral form*: not *old*, but *resembling* something old. There are old animals that look like sponges and millipedes, and no old animals that look like people. But there are also probably no old animals that look like various present-day parasites. It's best to discard talk of a phylogenetic scale. There is just a tree, with various tree-related features.

[3] Problem cases, whose location is uncertain, include comb jellies (ctenophores) and placozoans. Comb jellies may have branched off from other animals even before sponges.

[4] *Cnideria* is a phylum; *Bilateria* includes many phyla and has no official rank.

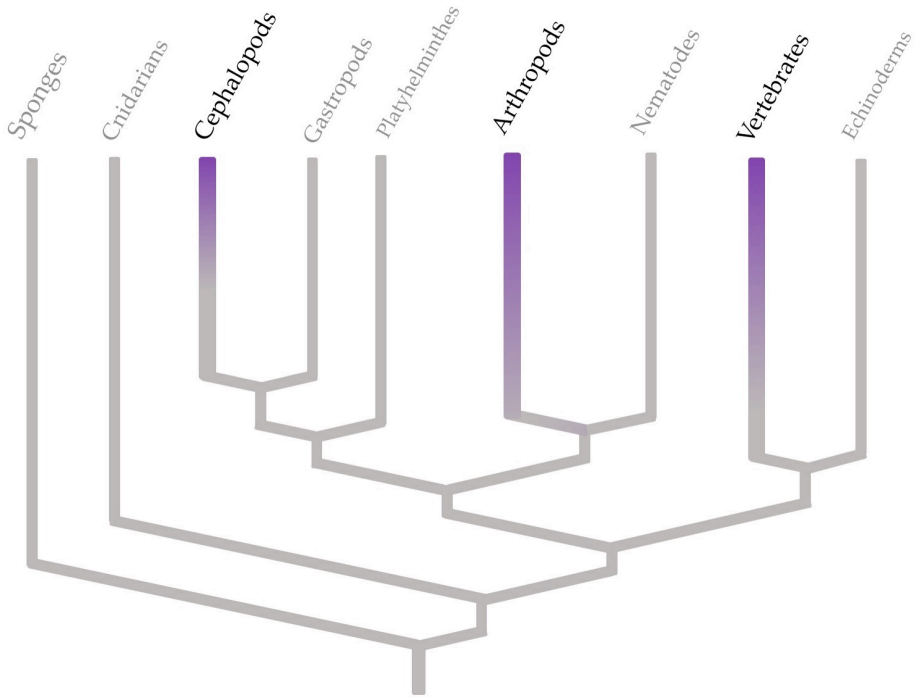
[5] "Nearly" because of uncertainty over placozoans. There are also a few other oddities.

few probably crawled, grazing microbial mats. The *Cambrian*, from 540 mya, saw the sudden appearance of bodies with hard parts and obvious means for complex behavior – legs, claws, sophisticated eyes.



Three groups, which had diverged in the Ediacaran, eventually produced *some* species showing conspicuous behavioral complexity: arthropods (especially some insects, spiders, and crabs), vertebrates like us, and cephalopod molluscs (especially the octopus).[6]

[6] The dates are all controversial. I am plotting what I take to be common rough estimates. The website *Time Tree of Life* (http://www.timetree.org) is useful. See Peterson et al., "The Ediacaran emergence of bilaterians," 2008 (though it has some dates older than the ones marked here) and Peterson et al. "Estimating metazoan divergence times with a molecular clock," 2004. There is a lot of disageement about dates, less about the order of branchings.



As Michael Trestman puts it, these three groups contain species with *complex active bodies*: animals with the capacity for rapid and directed movement, manipulation of objects, and sensory tracking of objects in space. They also have complex nervous systems that enable all this. Arthropods evolved this combination first, with the others following. The last common ancestor of these three groups, living in the Ediacaran, was probably a simple worm-like creature with much more limited capacities.

So we could start by recognizing three origins for complex behavior and large nervous systems (though it's possible to split more finely).[7] These animals all seem initial candidates for subjective experience, but we've not said enough yet to make much of a link to that question – there's not been much "bridging of the gap." So we can next look for some more specific features.

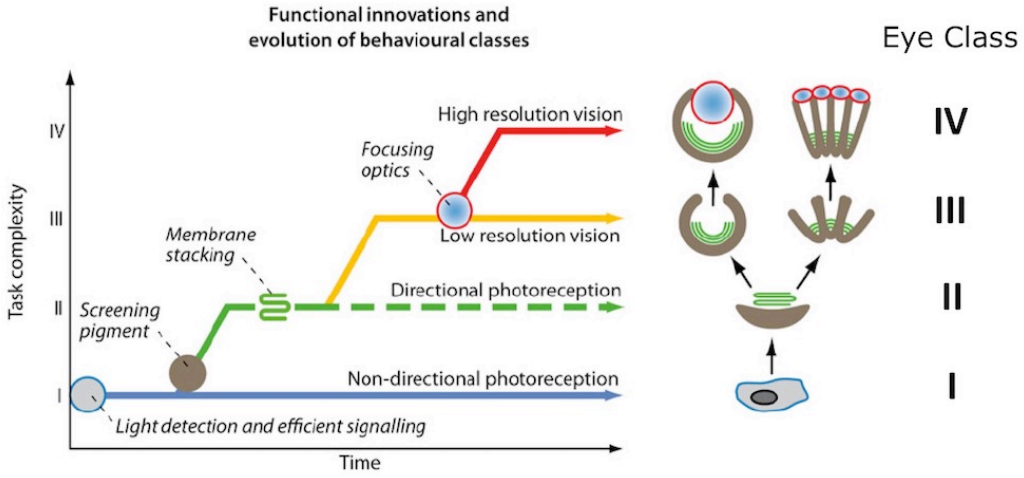
The pointer we'd get from a huge amount of recent work is: look at the sensory side. For many, subjective experience is either *inherently* sensory, or this is the paradigm

[7] Splitting more finely: behaviorally complex insects are a fair distance from behaviorally complex spiders, and mammals and birds are more neurally complex than other vertebrates, including their common ancestor, so you can count two origins for complex behavior rather than one in each of these cases.

case. For some this claim goes via a representationalist theory of qualia, but other approaches have the same sensory focus (Prinz, for example).[8]

Sensing of *some* sort is ubiquitous, not just in animals but in all known cellular life. But there are certainly differences in complexity that might be relevant. Here I'll look just at vision, a sense with unique importance in animal evolution. Which kinds of animals have eyes that can form a useable image and, with the aid of downstream processing, present objects in space?

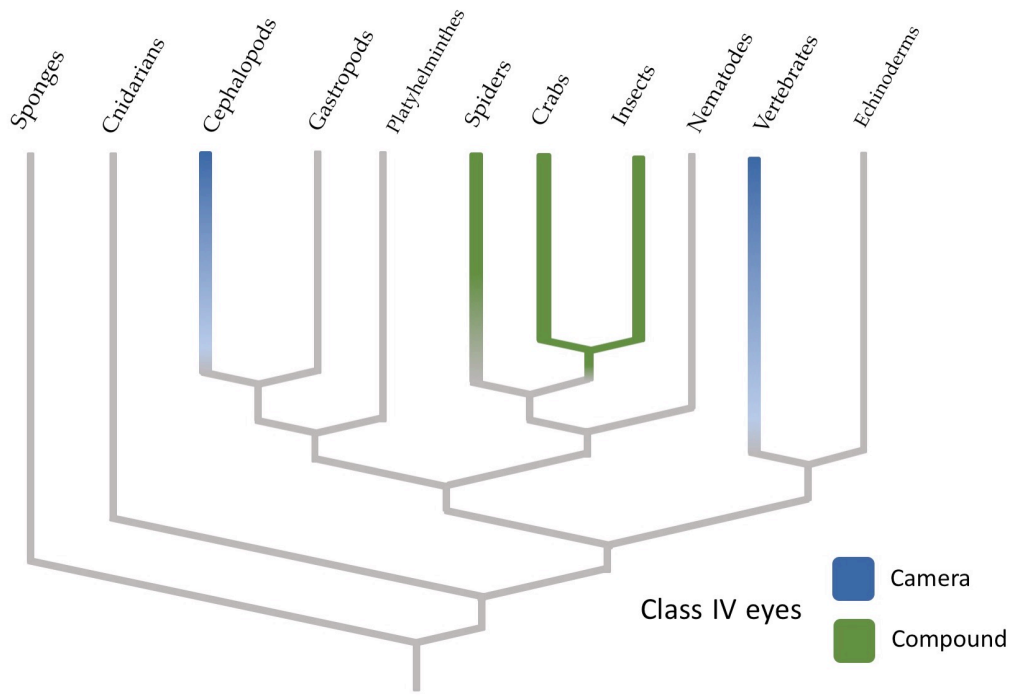
Dan-Eric Nilsson discusses four "classes" of eyes, I-IV, with IV being those that enable "high resolution vision." These eyes evolved perhaps three or four times, with two designs, the camera eye of vertebrates and cephalopods and the compound eye of arthropods.



Modified from Nilsson, 'Eye evolution and its functional basis,' *Visual Neuroscience* 2013

The groups in which these eyes arose are the same three picked out in more general terms earlier.

[8] Dretske: *"If* one chooses to talk about state consciousness (in addition to creature consciousness) at all, the clearest and most compelling instance of it is in the domain of sensory experience and belief."



I don't think perception is *the* key to subjective experience. In arguing that there is non-sensory conscious experience, people use cases like moods, and also "cognitive phenomenology." I agree with both. In my Jack Smart lecture (2017) I add the example of *energy level* – fatigue and its relatives – to the list.[9] Not all subjective experience is *presentation* of facts or things; there is just a way some internal states and processes feel.

A family of approaches to the evolution of consciousness that are not based on sensing set out from the idea that a plausible early form of consciousness is evaluative or affective. *Feelings*, especially washes of positive and negative affect, might be basic and widespread forms of subjective experience.

Valuation itself, like sensing, is ubiquitous in life.[10] But there might be forms with special relevance to the problem. An idea that has been picked up by several is a possible link between consciousness and instrumental learning – learning by tracking the good and bad consequences of your actions.[11] Other relevant kinds of evaluative sophistication

---

[9] "Materialism, Subjectivity, and Evolution," on my website.

[10] More than in the sensory case, and even bracketing questions about consciousness, there are uncertainties about stages and natural categories here. Some people think the usual distinction between classical and instrumental learning, with the latter being rarer, is partly erroneous.

[11] See, for example, Ginsburg and Jablonka, "The transition to experiencing," 2007, and Damasio, Panksepp, and Denton for the vertebrate case.

include trade-offs, as seen in Elwood's work on hermit crabs.[12] But I'll make my next point with the aid of learning itself.

Instrumental learning appears to be scattered through the tree in a way that suggests it evolved several times, and it may have an interesting pattern of presence and absence. It is present in vertebrates and cephalopods, and *some* arthropods – very prominent in bees. But it has not yet been seen in all arthropods, including some behaviorally complex ones, such as spiders. (Spiders have occasionally been said to have instrumental learning, but the standard citation is sometimes misdescribed and does not support this, or might be seen as an interesting borderline case.[13]) The review I make use of here – Perry et al. 2013 – also does not report instrumental learning in wasps, though I've found one report of its presence.[14]

That raises the possibility of a dissociation between features relevant to subjective experience. We can imagine, in principle, an animal with a very sophisticated sensory side and much simpler evaluation – an animal more "robotic" on that side. And there is the flipside – *perhaps* – an animal with less of an idea of what is going on, but with a stronger sense of whether what's happening is good or bad.

Is that second combination harder to make sense of? The possible asymmetry is reflected, or may be, in facts about distribution. For the first combination – rich sensing and simpler evaluation – we have various terrestrial arthropods, and the combination makes sense there. These animals often have short lives dominated by routine, by a definite list of actions – complex actions, but regimented ones. They don't live very "open" lives, but they face considerable sensorimotor demands, especially those that can fly. On the other side – rich evaluation, simpler sensing – there are no clear cases I know

---

[12] See especially his "Evidence for pain in decapod crustaceans," 2012.

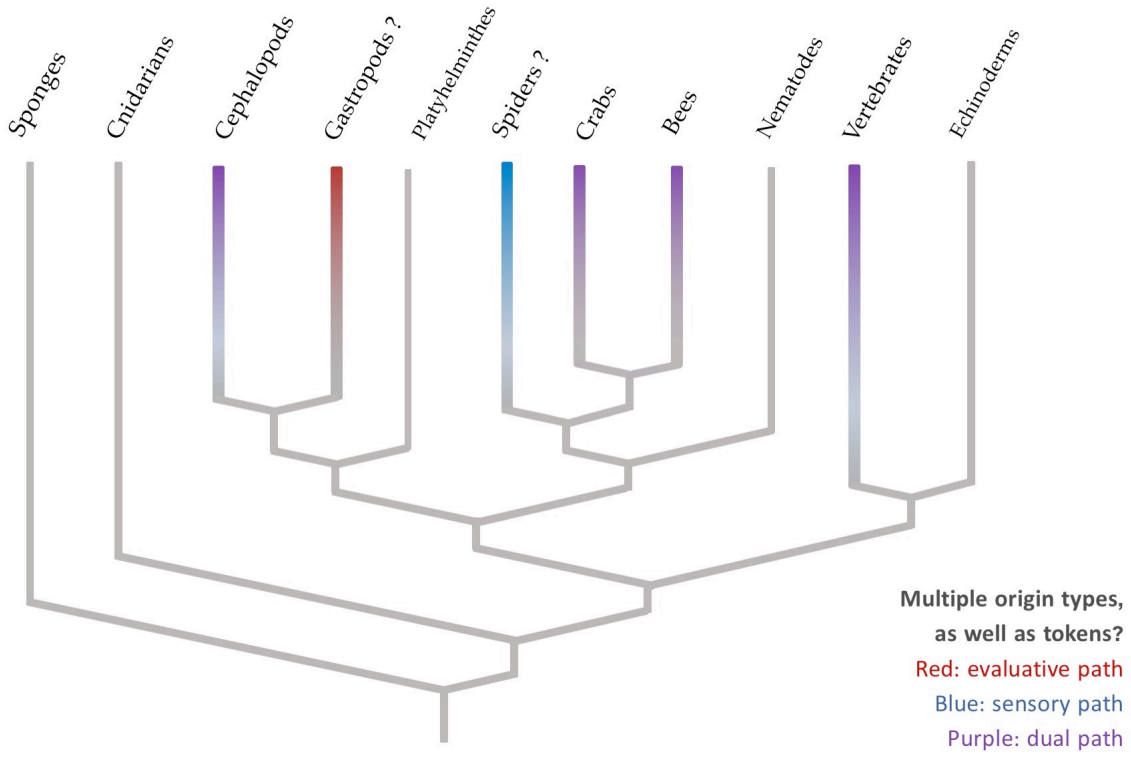
[13] Spiders can be classically conditioned. It's not that they can't learn at all.

[14] The review I use is Perry, Barron, and Cheng, "Invertebrate learning and cognition: Relating phenomena to neural substrate," *WIREs Cognitive Science*, 2013. This review lists what has shown to be *present* in various groups; it's much harder to show that a trait of this kind is *absent*. The negative claims I make use of here should be taken with caution. The wasp study is "Hitch-hiking parasitic wasp learns to exploit butterfly antiaphrodisiac" by Huigens et al., 2009.

Adamo (in a commentary on Barron and Klein, *PNAS,* 2016), notes that insects are under intense selection to keep their brains small, and she suggests they'll only have as much of a reward system as they really need. This point might also be taken to cast doubt on some arguments from homology – the idea that even if wasps don't obviously behave in the same sort of way as their close bee relatives, they probably have much of the same machinery. Fliers are unlikely to carry excess baggage.

of, but *perhaps* gastropods. They can do some instrumental learning, and have simpler eyes than other such learners (Class III eyes in Nilsson's sense).

Here is a summary slide, very tentative, for this part of the treatment. The slide has purple for complexity on both evaluative and sensory sides, blue for the sensory side only, red for the evaluative side only.[15] (In the talk I had wasps, also, in blue, as I was following the list in Perry at al. on this point. See footnotes 14 and 15 on this.)



If there is a problem with this picture it is probably one of making gulfs wider than they are. Something it illustrates in principle, though, is the possibility of different evolutionary paths to things that are different from each other, but amount in each case to a form of subjective experience. Less starkly, there might at least be paths that differ with respect to a reddish or blueish tinge.[16] Human experience features a definite combination of the sensory and the affective or evaluative, and the familiarity of this combination

---

[15] Bees, like other insects, do not display pain behavior in the relatively recognisable form seen in crabs and the like, but they are very sensitive to reward and aversive events in other ways, as Barron emphasized at the NYU conference.

[16] Feinberg and Mallatt (*Ancient Origins*...) make a three-way distinction; they recognize *sensory*, *affective*, and *interoceptive* consciousness.

guides various intuitions we might have about the problem. But these features may be separable.

From here, further traits can be put on the table, along with a finer-grained treatment of some of the ones I've discussed.[17] Integration of the senses (as opposed to sophistication in one particular sense) yields a multi-modal registration of the world. Another is integration of sensory input with the present state of the body. Through the evolution of traits of this kind, an animal becomes *a center of action and perception* in a way that is not true of other living things. One consequence of becoming a center of agency and perception of this kind is the animal's gaining of a sense of itself *as* such a thing. As emphasized by Merker, a mobile animal is continually generating by its own actions sensory inputs that must be disambiguated, must be registered as self-caused rather than informative about external changes.[18] As Björn Brembs has also noted, instrumental learning requires something of the same kind – the animal must register what it is doing or has just done, as opposed to what has merely happened.[19] Competent agency, in a mobile animal that can actively manipulate objects, requires, or at least encourages, some registration *of* one's own agency.[20]

I've put a number of traits on the table – complex sensing, evaluation, integration, registration of self and agency. I am not sure how they fit together biologically, and which are most important to the problems at hand. I'm also unsure how deep the differences in "style," discussed just above, might run. A question that would repay some work is: do subjectivity-related features form a cluster of associated properties, such that animals with one tend to also have others? To what extent is there a package of traits reliably found together, as opposed to more variety? But suppose we had a filled-out theory of this general kind. What contribution would it make?

I think a theory like this, when filled out, would be a biological theory of subjectivity. Once we have an animal with features like this, appropriately combined, we

---

[17] What about attention? Yes, especially if seen not as a gate (as in Prinz) or a spotlight, but as a matter of amplification (Desimone and Duncan, "Neural mechanisms of selective visual attention," 1995).

[18] Merker, "Liabilities of mobility," 2002.

[19] Brembs, "Operant Behavior in Model Systems," 2016.

[20] Perhaps not for the simplest instrumental learning that only affects the time and place of production of a species-typical behavior? See both Jablonka and Colin Allen on the importance of a category of "open-ended" learning in this context.

have a perceiving and acting subject, in a strong sense. Subjective *experience* is then some of what goes on in such animals, it's part of what they undergo.

Some descriptions of the problem at hand encourage this sort of bridging. Nagel has said that the main problem is explaining in naturalistic terms how a "subjective point of view" could be a feature of the world.[21] That is exactly what a theory of this kind would give us. Subjective experience is the biology of subjectivity from the subject's point of view. And subjects, in turn, are comprehensible evolutionary products.

To fill this out a little, I will make some comments about how a theory with this general shape relates to some lines of research that might seem to push *away* from the story told here, and might also seem to push away from recognizing subjective experience in animals far from us. This is work that suggests that it's *not* true that subjective experience is just the first person point of view on what complex animals find themselves doing.

First, a message often taken from current research on human consciousness is that there is a great deal of *selectivity* here (see, for example, Dehaene's work, and Milner and Goodale). A lot of what humans do is not conscious, and the unconscious seems to include a lot of the basic activity of getting by – perception, basic learning, and guidance of action. We seem to do a lot of this "in the dark," and this package of capacities might seem to include a lot of what animals get up to.

This could be expressed as an argument – though I don't think I've seen it made explicitly: humans can engage in a range of basic cognitive activities unconsciously, and what they do in this way seems similar to what many animals do, so it's probably possible to have an entire cognitive profile, of the sort typically seen in animals, that is unconscious.

This would not follow, though – it might be true, but it's not yet shown. In the experiments that motivate this kind of view, the human subjects *are* conscious, though they may not be conscious of everything they do. 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 combination of activities could all be that way at once. It might instead be that once you are an awake human subject, doing the usual sorts of things people do, *some* of it has to

---

[21] See *The View from Nowhere*.

be conscious. For any normal and wakeful human being there is *something* it's like to be that person, even if some processing is being done deep in the background.[2223]

Here is a second line of work that seems to push away from the view I'm defending.[24] Once we get to what *is* conscious in us, there are experimental phenomena that seem to show that what is experienced is not just ordinary perception and cognition (etc.) from the inside, but something more "constructed" or "synthesized." I'll use an example emphasized often by Hakwan Lau.[25] We experience "normal" color saturation of a visual scene even at the periphery of vision. That must mean we infer it and add it in, as color processing in the periphery is different and not very good.

---

[22] Here is a bit more detail about this, in the light of discussion at the NYU conference.

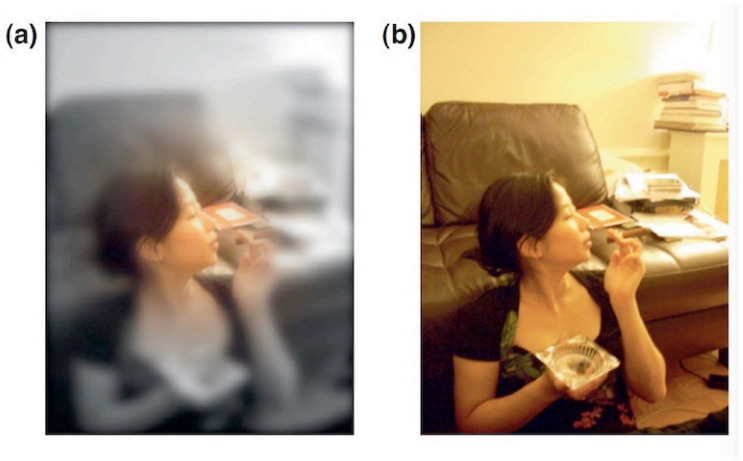
The data might show that you can do X_1 unconsciously, while being conscious of something else, and can do X_2 unconsciously, while being conscious of something else.... The conclusion that would not follow is that you can do X_1, X_2,... (etc.) all together, without being conscious of *anything*.

Block compares actual blindsight to what he calls *super-duper-blindsight*, where there is a normal level of visual discrimination, freely exercised, but the subject does not experience this as vision (as opposed to real-world blindsight where subjects have reduced visual discrimination and do not freely exercise their abilities but have to be prompted and cajoled). We can take Block's thought experiment further and define *global super-duper-blindsight*, in which all senses and also cognitive processes have this character. This is not "zombie" status, as the person does not insist they have feelings, and so on. One way of expressing the argument that I think some people are tempted towards is to say that we can infer from facts about ordinary blindsight, and its relatives, to the possibility of global super-duper-blindsight, as this would be just a combination of things we have reason to believe in. I am saying that this argument would be fallacious if offered.

[23] Relatedly, in *The Ancient Origins of Consciousness*, Feinberg and Mallatt say that the capacities used as evidence for "complex" sensing and processing without consciousness are so "weak and incomplete," when you take the ecological context of behavior into account, that "any fish relying on them could not sense dangers well enough to survive in nature" (p. 210).

[24] This one was cut from the presentation on the day.

[25] See, eg., Lau and Rosenthal, "Empirical support for higher-order theories of conscious awareness," 2011 and Solovey, Graney & Lau "A decisional account of subjective inflation of visual perception at the periphery," 2014.



Lau and Rosenthal, 2011.

Above is a figure from Lau and Rosenthal (*TRiCS*, 2011); the idea is that we see the world as in (b), but a "raw" version of experience would be more like (a).

This sort of thing suggests that even basic sensory experience in us is "synthesized," with the aid of sophisticated capacities.[26] One might also suspect that the capacities that do this "synthesis" are not things that lots of animals, especially invertebrates, have. (Again, I don't think I've seen this argument made explicitly in print.) Subjective experience, again, is not just the inevitable result of a biological encounter between a complex animal and its environment.

One might reply that animals *can* do this sort of synthetic work, in their own way. But that is not necessary to offer a response. The more basic reply is that the traits responsible for these effects give rise to *our* kind of subjective experience. That does not mean that without these traits, you have none. Without this you have, or may have, some other kind of experience.[27]

---

[26] This is the sort of work that induces some people to say "we live inside a model" or "live in a simulated reality." That is not the conclusion to draw, but it's an exaggerated description of something real.

[27] Relecting on all these ideas and some discussion at the NYU conference: I think some standard ways of describing the problem at the outset may be misleading. People say that phenomenal consciousness is "a property of mental states." (Carruthers said this at the start of his talk.) What might be misleading is the idea of a collection of states present in an agent at a time, some of which are conscious and some of which are not. An example used in discussion was the sensation of pressure on one's backside while sitting normally in a chair. This, someone said, is usually unconscious – it's an unconscious sensory state. But perhaps instead we should start from the idea of an entire psychological profile present at a time. Individual "states" are often somewhat arbitrary abstractions from this profile. As a person sits and writes, there will be something it

The last challenge I'll discuss is more a challenge to *us*. I discussed a range of subjectivity-relevant properties, properties that look like partial gap-closers in different ways. As I noted, these all have simple forms, running all over the tree of animal life and outside animals to some degree. If these *are* subjectivity-relevant, does this relevance only appear in the complex forms? Or in all forms? I don't know. But to me it seems that the closer we look on the biological side, the more graded the relevant properties are. Those sensory and evaluative transitions I discussed earlier - Class IV eyes, instrumental learning – are all surrounded by various sorts of close approaches and partial cases.[28] That leaves open the possibility that there is some reason why only some values of these variables, only some forms, are relevant to the existence of genuine subjectivity. I don't see why this should be, but don't think we know either way. A true gradualism about subjective experience is both the more difficult option to think about, but also given our empirical picture, the more natural option, the one that looks suggested. If so, we will have to move beyond an account of which features *suffice to make you* an experiencing

---

feels like to be that person – the whole profile is conscious (in the thin sense of the term in use here). Then we can ask: what difference does the pressure on the person's backside make to the totality of that person's experience? Does it feel different, overall, that there is the pressure? Is the "conscious" processing of the act of writing affected by the "unconscious" sensing of pressure from the seat? Will a less comfortable seat slow the writing a little, or affect overall mood? If so, then the seat pressure is making a difference to what it feels like to be that person. It is making a contribution to a psychological profile that is subjectively experienced.

[28] Eyes: nudging the category of class IV eyes Nilsson has several borderline cases - *almost* class IV (in free-swimming molluscs charmingly called "sea elephants," and some swimming annelid worms). Between III (low-resolution vision) and II (directional photoreception without vision) he has a couple of "IIA" eyes (scallops, sabelids). That is right on his border between vision and its absence. Nilsson has the simplest *possible* case of an image-forming eye in a cyanobacterium. See his "Eye evolution and its functional basis" and other papers. If you think that reafference compensation in handling sensory input is significant (as I do), then you must note that nematodes with 302 neurons have a circuit dedicated to this.

On the evaluative side: Instrumental learning may be rare, but momentary reward-guided behavior is ubiquitous, and has much of the same neurochemical profile (especially dopamine systems). (See Barron et al., "The roles of dopamine and related compounds in reward-seeking behavior across animal phyla," 2010.) Combinations of momentary reward-guided behavior and classical conditioning can yield "conditioned place preference," which looks a lot like instrumental learning though it may be entirely classical. This is seen in flatworms (platyhelminths). See my "The evolution of consciousness in phylogenetic context," 2016.

subject, to an account of how having richer versions of these features implies internal goings-on that are *more experiential.*[29]

I do think that there is probably *something it feels like to be* various invertebrates: an octopus, and probably a bee and a crab, all in different ways. The bee has a miniature brain but one exquisitely tuned to seeking rewards, and able to control flight, with all flight's complex reafferent relationships between action and perception. The crab has a more laconic pace of life, but complex sensitivity to aversive experiences. Octopuses I'll discuss in a moment. From there on, I am more wary, and start to work within a gradualist framing – one that posits something *more approximating experience* in various other animals, with their different mixes of the properties I've discussed.

I'll finish with the special case of the octopus. This is the most neurally complex invertebrate, with something like half a billion neurons. Octopuses first have all the basics I've discussed here, on both sensory and evaluative sides. They have class IV eyes, instrumental learning, and wound-tending. They also have some "extras," a *style* of cognition that is suggestive of the phenomena discussed in workspace theories and the like – not one suggestive of there being a special place inside the animal, a workspace,

[29] The question of gradualism interacts with the questions about multiple origins discussed earlier. The deeper subjective experience lies in the tree, the fewer distinct origins we can expect (though more sophisticated forms of the trait can still have many origins).

but suggestive of the sort of attention-driven, novelty-directed way of being that workspace theorists like Dehaene associate with consciousness.

Octopuses are hence the best cases for invertebrate consciousness, and, given the history of animals, for multiple independent origins of consciousness.

I think the "smartness" of octopuses is sometimes exaggerated; they are probably not *ruminative* animals, in the way some birds, for example, seem to be, but they have an exploratory style of behavior, continually manipulating objects in novel ways.[30] These complex manipulative behaviors, along with their array of receptors of different kinds, must generate a very rich sensory world.

In making this inference, and in trying to imaginatively "get inside" them, we have to contend with the fact that octopuses' nervous systems are surprisingly decentralized, with most neurons in the arms themselves. The sensing in the arms, which is very extensive, seems to have a combination of more local and more global consequences. This is relevant to questions of integration, and the generation of a self. But octopuses can "pull themselves together" and act as attentive, evaluating subjects, featuring sometimes chaotic mixes of curiosity and caution.

---

[30] See Tayor et al. "Spontaneous metatool use by New Caledonian crows," *Current Biology* 2007.

Thanks to Andrew Barron, Gáspár Jékely, Fred Keijzer, John Sibbick.