Animal Evolution and the Origins of Experience

Peter Godfrey-Smith

CUNY Graduate Center, &

HPS, University of Sydney

1. Introduction

How can we find a way to understand the simplest and most basic forms of subjective experience. What is the set of living organisms for which it *feels like something to be* one of those organisms? When did this phenomenon begin, and what was its earliest form?

The intrinsic interest of these questions is obvious, I take it, and they are important in at least two other ways. Progress here should help in other areas of philosophy of mind, including the most basic debates about how mental and physical are related. To this you might say: attempts to answer my questions won't *help* with the mind-body problem itself, but rather, would *be* helped by resolving (if we can) the more fundamental questions. However, it may be that if we better understand the relations between simpler and more complex forms, this will help us see how subjective experience can have a basis in the material. I think that the shape of an eventual theory will be one that relates the material to the living, the living to the cognitive, and subjective experience to the kind of cognitive operations that living systems engage in. In a companion paper to this one ("Mind, Matter, and Metabolism") I address the first couple of these relationships in detail. In this paper I discuss the later ones, looking especially at the evolution of animal life, and how stages in animal evolution might be related to subjective experience.

These issues are also pressing in a more practical way. Here I have in mind ethical questions about the treatment of animals in farming, experimentation, and elsewhere.

Questions about how we should treat animals of various kinds are closely connected to questions about subjective experience, especially questions about suffering. A plausible view is when we are dealing with an organism with no subjective experience at all, there are few ethical concerns about ways we might treat it. Or perhaps there are, but they are different questions from those that apply to an animal that has the capacity to suffer (questions in environmental ethics may still apply, for example). For animals that do have subjective experience, especially of a negative kind (pain and suffering), there is a strong initial case that this should be factored into decisions about how we treat them. In those discussions, the important questions are not evolutionary *per se*, but about the distribution of subjective experience among present animals. Does a fish or crab have any subjective experience at all? But evolutionary questions are connected to these ones.

The next section sets up the topics of the paper in more detail, and sketches some features of evolution *before* animal life. I then look at the early history of animals, focusing on stages that seem likely to have some relation to the evolution of subjective experience. The last section looks at the relation between the evolutionary history and recent work in neurobiology and philosophy of mind.

2. Subjective Experience and Early Evolution

I said that the aim is to understand subjective experience. How does this relate to questions about consciousness? It is common now to use the word "consciousness" in a broad way, to cover all kinds of subjective experience. Distinctions might then be made between different kinds of consciousness, some more complicated than others. I don't think there is any error in setting things up this way, but it's not the best. An earlier way of framing the issues, seen more often in the 1980s, distinguished three main problems for philosophy of mind: *qualia*, *consciousness*, and *intentionality*. The problem of qualia was seen as the problem of explaining the first-person feel of the mental, intentionality involves semantic content or "aboutness," and consciousness was seen as a sophisticated kind of mentality, with special features on both the cognitive and qualitative (subjectively felt) side.

Now "qualia" and "consciousness" are often seen as amounting to the same thing, not because of an argument for reduction of one to the other, but because there is only

one phenomenon to consider.[1] If there is something it feels like to be a system, then the system is said to be conscious, or have some kind or degree of consciousness (Nagel 1974). I prefer the earlier set-up, and think the difference is not merely verbal. "Qualia" was an awkward term, but it captured the possibility that there might be some sort of very diffuse *feeling* present in the activity of a system which is distinct from anything we would usually call consciousness. I wonder whether squid feel pain, but I don't think of this as wondering whether squid are conscious. "Sentient" is a better adjective for the more general property, and some people use that term, though others would probably say that to wonder whether a squid feels pain is to wonder whether it is *phenomenally conscious*.

In this paper, the phrase "subjective experience" will be used for the broadest category of phenomena in this area, which we can also pick out by saying that some states of some systems feel like something to the system and others do not. "Qualitative" (in a sense derived from *qualia*) will sometimes be used as an adjective for the felt features of mental states that are subjectively experienced. Consciousness is something beyond mere subjective experience, something richer or more sophisticated, though it is hard to say how this is best understood. I'll use the term "cognitive" in a very broad way for the processes in organisms that manage sensory input, establish and access memories, control behavior, and so on. I don't assume that an information-processing or computational view of all these processes is the right view (though it might be). I want a general term for the side of the mental that involves behavioral control and intelligence.

The next section will work through some of the history of animal life. Before that, I will describe some of the evolutionary setting *before* animals, especially as it's important to appreciate that a substantial amount of cognitive or proto-cognitive capacity was in place before animals evolved.

Suppose we approach the history of life from the point of view of functionalism in philosophy of mind, looking for the initial evolution of perception, memory, and behavioral control – the things functionalism tells us are important in giving a physical system psychological properties. All these capacities evolved well before animals, and

---

[1] Chalmers in his influential 1996 book also frames things this way.

some are seen in quite sophisticated forms even in single-celled organisms, including prokaryotes (bacteria and archaea). A good example is chemotaxis (seeking or avoiding chemicals) in the bacterium *E. coli*. This system controls bacterial swimming, and it makes use of a form of memory. The swimming choice of a bacterium at each time-step is controlled by a comparison made between the levels of good and bad chemicals that are presently sensed, and the levels sensed a few seconds before. If conditions are improving, the cell continues along the line it has been following. If conditions are getting worse, it takes a random "tumble." This is quite a smart system – much smarter than the usual philosopher's example of magnetotaxis – especially as it involves something beyond the simplest relations between input and output.[2] The present stimulus has a significance dependent on the previous time-step.

Bacteria are prokaryotes, cells with no nucleus that also lack further internal structures that other single-celled organisms have. An important event before the evolution of animals was the evolution of *eukaryotic* cells, which are larger and more complicated, and whose initial evolution features the engulfing of one prokaryote (a bacterium) by another (an archaean) something like 1.5 billion years ago. One feature of eukaryotic cells that is especially important to the evolution of behavior is the *cytoskeleton*. This is a skeleton-like collection of fibers within a cell whose movements can be chemically controlled. In particular, they can contract. This makes possible changes of shape in the cell as a whole, and is the beginning of non-trivial manipulation of objects and new kinds of locomotion. Single-celled eukaryotes also evolved richer forms of sensing, such as detecting the direction of light.[3]

---

[2] For the *E. coli* system see Baker et al. (2006). For magnetotaxis see Dretske (1986) but also O'Malley (2014).

[3] See Spang et al. (2015) for an important bridge between prokaryotes and eukaryotes with respect to quasi-behavioral capacities in unicellular organisms. For the evolution of light sensing and vision in unicellular organisms, see Jékely (2009).

3. Transitions in Animal Life

From a world of the more complex single-celled organisms described just above, the evolution of multicellarity occurred perhaps a dozen times, independently, with different results. One of these gave rise to animals.

3.1. Multicellularity

Animals are a branch of multicellular organisms originating perhaps 800-900 million years ago. There is much uncertainty about the dates and the pattern of the first branchings in this part of evolutionary tree of life. I'll work provisionally in this paper within a fairly traditional view of the history of animals. This view has been challenged, but debates about the first events do not have too many consequences for the principles central to this paper.

Though sensing and the control of behavior were not animal inventions, multicellularity made possible great shifts in the evolution of these capacities, as it enabled a specialization of sensing and acting parts that were formerly crammed into one cell. This division of labor requires interaction between parts – some sort of effect of one cell on another in real time. There are various ways to do this, and some of it can be achieved without a nervous system, but only a small range of extant animals do not have nervous systems – sponges, placozoa, and a few reduced oddities whose ancestors had and lost them.[4] So that is the next step to consider.

3.2. Nervous systems

Nervous systems arose perhaps 700 million years ago. There is ongoing debate about whether they arose once or several times, but they certainly evolved early and are present in nearly all animals. For someone interested in the evolution of subjective experience, this might look like *the* transition, the landmark. And so it may well be, but I said all that without addressing the question of what a nervous system *is*. This is not a question with a straightforward answer.

---

[4]   See Jékely et al. (2015).

Nervous systems enable interactions between cells with respect to their electrical properties. Cells can "depolarize" – the usual charge difference across a cell membrane can be lost and quickly restored. Nervous systems induce patterns of these changes in collections of cells. Those features, however, are also seen outside animals – depolarization of cells, and effects of one cell on another's electrical properties. Suppose the category *neuron* were to be understood purely in terms of excitability and cell-cell interaction; perhaps any cell is a neuron if it is electrically excitable and can influence another cell's electrical excitation, by means of chemical intermediaries or more direct effect. If that is what a neuron is, then various organisms that are usually called "non-neural" do have neurons, including some plants.[5] That broad functional definition of a neuron is quite a reasonable one, though it is at odds with many habits of description in biology. What might a reasonable narrower definition look like? In a paper about early nervous systems co-written with Gáspár Jékely and Fred Keijzer (forthcoming), we opt for a definition that is partly functional and partly morphological: a neuron is an electrically excitable cell that influences another cell by means of electrical or secretory mechanisms*,* and whose morphology includes specialized projections. Neurons in this sense are not seen outside animals. A nervous system, then, is an interacting collection of cells that are (or include) neurons in this sense.

There might appear to be an element of arbitrariness in adding this morphological criterion about projections, to narrow down what counts as a neuron. And as I said above, I think there's nothing wrong with the broader definition. But the combination of excitability, chemical signaling, and a morphology with projections is an important one; it enables nervous systems to achieve specific *kinds* of cell-cell interaction, including interactions that are tightly targeted. The projections on a neuron enable that cell to affect the electrical properties of another cell some distance from it, without affecting all those along the way. The contrast between neural and non-neural organisms is not one between organisms with networks of cells that affect each others' electrical properties and organisms that do not, but a contrast between different modes of cell-to-cell interaction. An important and almost exceptionless generalization can also be stated: all and only

---

[5] For a review, see Volkov and Markin (2014).

organisms with neurons (in the narrower sense defined above) also have muscle cells. Muscle and neurons seem to have coevolved.[6]

Did neurons evolve once or several times? This has been debated at length over recent years, due to changes in our understanding of the early evolutionary branchings in animal history.[7] A traditional view has it that sponges are the "sister group" to all other animals: there is an evolutionary branching deep in the past that goes on one side to sponges, and on the other to all other animals. Another possibility, based on genetic evidence, is that comb jellies, or *ctenophores*, are the sister group to all other animals, including sponges. In other words, there is an early evolutionary branching that goes on one side to ctenophores, and on the other to other animals including sponges. Ctenophores used to be grouped with jellyfish – and some still think this is right – but a number of recent papers have made the case that ctenophores are more distant from us than any other living animal. This debate is important because sponges do not have nervous systems, while ctenopores do. If ctenophores are the sister group to all other animals, then either nervous systems either evolved twice (once for ctenophores and once for everyone else) or the ancestors of sponges had nervous systems and lost them.

Figure 1 represents some animal groups and their evolutionary relationships. To keep things simple, in this figure I leave out ctenophores altogether. The term "neuralian" was introduced by Claus Nielsen (2008) for all animals with nervous systems. When he introduced the term, Nielsen assumed that the neuralia are a single branch of the tree – a "clade." (Any portion of an evolutionary tree like that in Figure 1 is a clade if it can be generated by picking a point on the tree and including all organisms downstream of it.) Nielsen assumed that ctenophores were located somewhere internal to the neuralian clade. If ctenophores belong outside sponges on Figure 1, then animals with nervous systems do not form a clade. It would be a bit unusual to keep the term "neuralian" if that is how things turn out, but I am going to set this issue aside and allow that there might

---

[6] For the exceptions, see Jékely et al. (2015).
[7] For this debate, see Moroz (2015), Jékely et al. (2015).

also be neuralia outside Figure 1.[8] This does not affect the points made later in this paper, as my focus will be on evolution in the organisms that are represented in the figure.
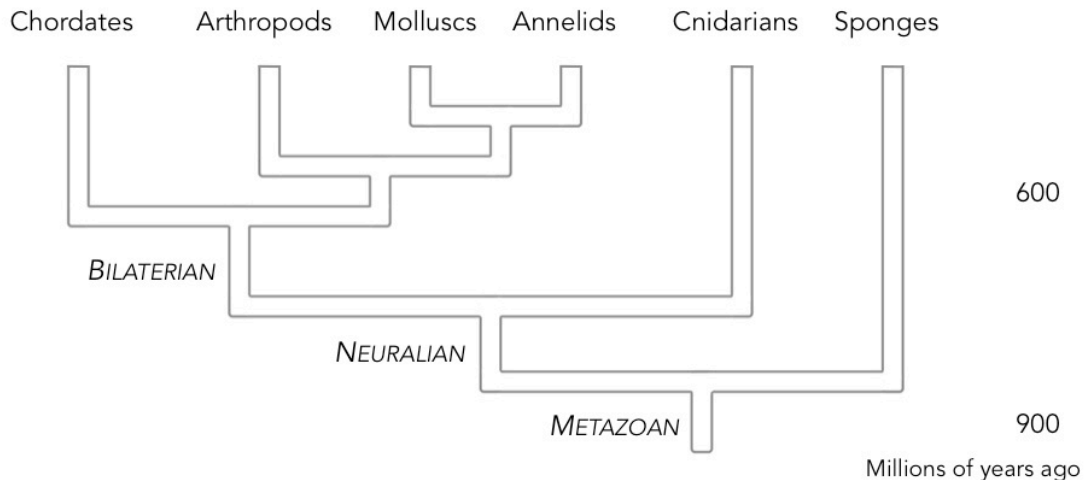

Figure 1: A representation of some early branchings in the animal part of the tree of life, with dates (in millions of years) tentatively associated with some events. Names along the top designate some of the main taxonomic groups within animals. The italicized and capitalized labels show the initial appearance of broad kinds of organisms important to the paper. First is the evolution of *metazoa*, or animals; then *neuralia*, animals with nervous systems (or perhaps a subset of these animals – see the main text); then *bilaterians*, animals with bilaterally symmetrical bodies, including ourselves.

Suppose for a moment that nervous systems were a one-time animal invention. Again, this might look like *the* landmark for the early evolution of the mind. But what were the first nervous systems doing for their owners? One natural assumption is that these early nervous systems had a simpler version of the same role of coordinating perception with action that is seen now in us. In bacteria, in early animals, and in ourselves, a crucial task is coordinating what is perceived with what is done, and nervous systems evolved in animals to do this in a more complex way than other organisms.

---

[8] As well as sponges, *Placozoa* are animals without a nervous system. I've not marked them on the tree. They are thought to have branched off later than sponges but earlier than cnidarians.

Perhaps that is right, but we should not simply assume it. First, many things present-day nervous systems do aren't a matter of controlling behavior, and these may have been very important in the early stages; nervous systems often control aspects of development and physiology.[9] And even within behavior, there is a possible discontinuity between then and now. The ideas I'll sketch next draw on a paper by Fred Keijzer, Marc van Duijn, and Pam Lyon (2013), building on earlier work by Carl Pantin (1956).

When people imagine the role of early nervous systems, they often picture a flow chart starting with perception and terminating in behavior. Behavior itself is taken for granted: something is done. But *how* is it done? In the case of a multicellular animal, it is a substantial task to perform a coherent behavior at all, coordinating the micro-acts of cells into a useful macro-act by the whole organism. There's an important *internal coordination* role that nervous systems play, which is distinct from their role in coordinating perception with action.

I noted earlier the coevolution of nervous systems and muscle. Without muscle, an animal can't do much. Motion must then be achieved with cilia (little hairs) whose powers are limited. Keijzer and his co-authors argue that it was the demands of coordinating muscle action into useful behavior that first gave rise to the patterns of interaction between cells associated with nervous systems. Guidance from the senses in simple animals can be done non-neurally, at least in large part. They suspect that the first nervous system evolved as a way to control a complex new effector system – muscle – in something that might have looked like an early cnidarian (a relative of jellyfish and anemones). In some passages above I assumed what Keijzer and his co-authors call an *input-output role* for nervous systems – the emphasis was on a division of labor between some cells specializing in sensing, others in acting. Keijzer and his colleagues want to challenge this assumption. Early nervous systems might have had a lot to do with just *pulling the animal together*.

So far these are points of principle, regarding possibilities for early nervous systems and their function. Is there any way to make claims about how things actually

---

[9] These roles are discussed in more detail in Jékely, Keijzer, and Godfrey-Smith (forthcoming).

went? I will raise some possibilities (which depart now from Keijzer, van Duijn, and Lyon).

Before 600 million years ago or so, we have no idea what the lives of animals were like. The only evidence that animals existed at all, and had nervous systems, is genetic evidence. The branching point that connects humans and cnidarians, for example, may be found around 650 million years ago.[10] We have no idea what animals at those times looked like, beyond inferences that can be made by extrapolating backwards from animals living more recently. Then we reach a period now called the *Ediacaran (*635-540 million years ago), from which some soft-bodied animals are preserved as fossils. Once we find animals whose lives we can say something about, we see something of philosophical interest. Many Ediacaran animals seem to have lived on the sea floor, grazing on microbes or filter-feeding. Some appear to have been mobile, and some probably had nervous systems. What were they doing with them? We can make some defeasible inferences from their bodies. Ediacaran animals have no legs, no antennae, no sign of complicated eyes, no shells, no spines, no claws. They had none of the bodily tools of complex interaction between animals, and none of the obvious tools of complex real-time behavior at all. There appears to have been little or no predation – there are no fossils of half-eaten individuals.[11] In Mark McMeramin's apt term (1998), it seems to have been "The Garden of Ediacara."

If we employ the distinction introduced above between *internal coordination* and *input-output* roles for early nervous systems, then this feature of the fossil record suggests that in the Ediacaran, a lot of what nervous systems did was internal coordination. What was going on that needed *reacting* to? Not much; lives appear to have been quite self-contained. Nervous systems in the Ediacaran may have functioned mostly in "pulling the animal" together, as I put it above, enabling simple locomotion and feeding, and controlling physiology and development, without complex real-time sensorimotor arcs being present at all.

Another piece of evidence may push in a different direction, however. This comes from the evolution of associative learning. Standard frameworks in learning theory

---

[10]   Here I draw on Petersen et al. (2008).

[11]   There is just one possible known exception, some *Cloudina* fossils from the late Ediacaran.

distinguish between *classical* and *instrumental* conditioning. Classical conditioning, exemplified by Pavlov's dog, is a means by which correlations between stimuli can be tracked – a behavior apt as a response to *A* comes to be produced in response also to *B*, when *B* is a predictor of *A*. Instrumental conditioning is learning to produce (or avoid) behaviors that have been previously followed by good (or bad) consequences (or such consequences in specific situations). The origins of associative learning are unclear, but classical conditioning is very widespread across bilaterian animals – animals with left-right symmetry, such as ourselves.[12] Within this group, classical conditioning is seen in animals as simple as nematodes, which have only 302 neurons and lack action potentials. Instrumental conditioning, in contrast, has been seen only in invertebrates with larger nervous systems, such as crabs, various insects and some molluscs.[13] Classical conditioning may still have evolved independently in bilaterians a number of times, but suppose, for the sake of argument, that it evolved once and was passed down many lines from there. If so, it probably evolved something like 600 million years ago, either in the Ediacaran period or before it. Classical conditioning appears to be activity whose function is tied very much to an *input-output* role for nervous systems, a role that involves adapting to environmental events and patterns; it is a tool for dealing with external patterns, not a tool for internal coordination. (Instrumental conditioning, in contrast, has both possible roles.) If classical conditioning evolved in or around the Ediacaran, it makes it less likely that nervous systems at this time were concerned mostly with internal coordination.

Neither of these historical arguments are strong ones, though they point towards directions from which further evidence might come.

3.3. Sensorimotor complexity and CABs

At the end of the Ediacaran period we reach the Cambrian "explosion," when many new kinds of animal appear in the fossil record. From these bodies we can again make

---

[12] See Perry et al. (2013) for a review. They also accept a single finding of classical conditioning in an anemone, an animal outside the bilaterians. This would push a single origin for classical conditioning further back in time, or else be an independent origination of the ability.
[13] See Perry et al. (2013) again.

inferences – stronger ones, this time – about lifestyles. From the early Cambrian we see *do* see legs, antennae, complicated eyes, shells, spines, and claws. There is much controversy and rampant speculation about the Cambrian, but a range of mainstream views form a family that has particular importance here.[14] These views hold that at least one important thing that happened in the Cambrian was a process of *feedback* that linked the evolution of behavior and bodies in many groups. This shift may have first taken place in arthropods (which now include insects, and back then included trilobites). Whether they were first or not, the evolution of more complex behavior in some animals seems to have made life more complicated for others. In the early Cambrian, predation arose – seen clearly in the fossils – and with predation, a series of "arms races" appear to have followed, improving the senses and the means for bodily action. The evolution of rapid and fine-grained behavior in one animal makes the choices of others more acute. The evolution of image-forming eyes might have been especially important in this process (Parker 2003), or perhaps it was a suite of features. Whether eyes were pivotal or not, the *details* of what was going on around an animal came to matter to its life and prospects. This rather obvious feature of present-day animal life may not have been in place at all before the Cambrian.

The coevolutionary processes of this time linked senses, behaviors, and bodies. Michael Trestman, in a useful categorization (2013), marks out what he calls *complex active bodies*:

> This is a cluster of related properties including: (1) articulated and differentiated appendages; (2) many degrees of freedom of controlled motion; (3) distal senses (e.g., ''true'' eyes); (4) anatomical capability for active, distal-sense-guided mobility (fins, legs, jet propulsion, etc.); and (5) anatomical capability for active object manipulation (e.g., chelipeds, hands, tentacles, mouth-parts with fine-motor control).

These are bodies that can manipulate objects, sense things at a distance and react to them. CABs originated in the Cambrian, and as Trestman has it, only three groups of animals have given rise to bodies of this kind: vertebrates, arthropods, and a small group of molluscs, the cephalopods. With these bodies, the role for nervous systems that we are

---

[14] See Marshall (2006) and Budd and Jensen (forthcoming).

familiar with – the fine-grained linking of perception and action – becomes prominent. There's an opening up of senses to the world, and through new capacities for behavior, tighter *loops* between perception and action. Not only does what you do come to depend in a finer-grained way on what you see, but what you *do now* affects what you *see next*.

Figure 2 summarizes the steps described in the last few pages. As in Figure 1, we start from the evolution of animals (metazoa) and nervous systems (neuralia), at least on the non-ctenophore line. A shaded band marks the Ediacaran. Around this time we see the evolution of bilaterian animals. As the figure shows, genetic evidence suggests that many of the major animal groups had already diverged at this early stage, without much morphological fanfare. Then we reach the Cambrian, a band shaded differently, a time of rapid evolution of bodies and behavior.
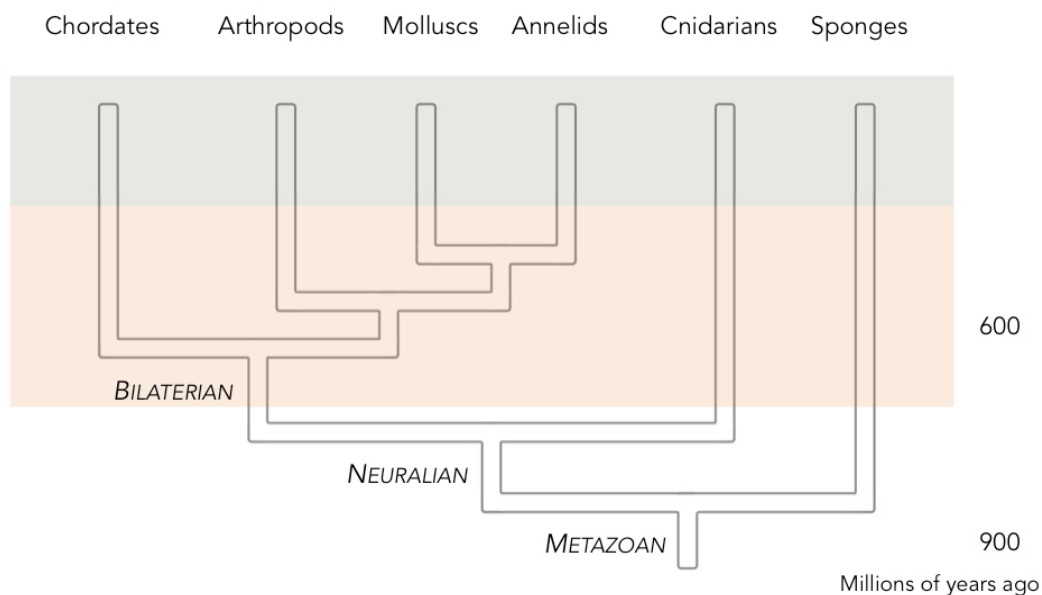
Figure 2: Further events in animal evolution. The lower shaded band marks the Ediacaran, the upper shading marks the Cambrian. Genetic evidence suggests that various familiar animal groups branched off from each other in the Ediacaran, though we have little fossil record of them.

If this is right in broad outline, we reach the following picture. The first nervous systems may have done rather little of what we now see nervous systems as enabling –

behavior in real time, the fine-grained processing of what the senses tell us. Eventually, these did become central to animal life, in a process that began perhaps in the Cambrian. From that point on, the mind evolved in response to *other* minds – in response to demands that the speeding-up of behavior, more complex senses, and an ecology of individual-on-individual interaction placed on each organism. Further, new *bodies* evolved in response to other minds. Bodies that would not have been advantageous before these new behavioral regimes now became essential. The ecology in which new bodies evolved was an ecology of behavior.

The stage we've reached is well before we get to any of the animals that people usually think of as having subjective experience. We are in a world in which the behaviorally significant animals are arthropods, simple fish, and (more so a little after the Cambrian) some molluscs. With respect to the senses, behavior, and the nervous connections between them, though, some plausible basics are now in place. From this point onwards, the evolving differences between animals have a more quantitative character: some of them gain more neurons, more sophisticated learning and categorization, more complex behavior. Others remain, or become, simpler.

4. Latecomer and Transformation Theories
I'll now start to bring this historical material into closer contact with the philosophy of mind. I'll organize this discussion with another diagram, Figure 3.
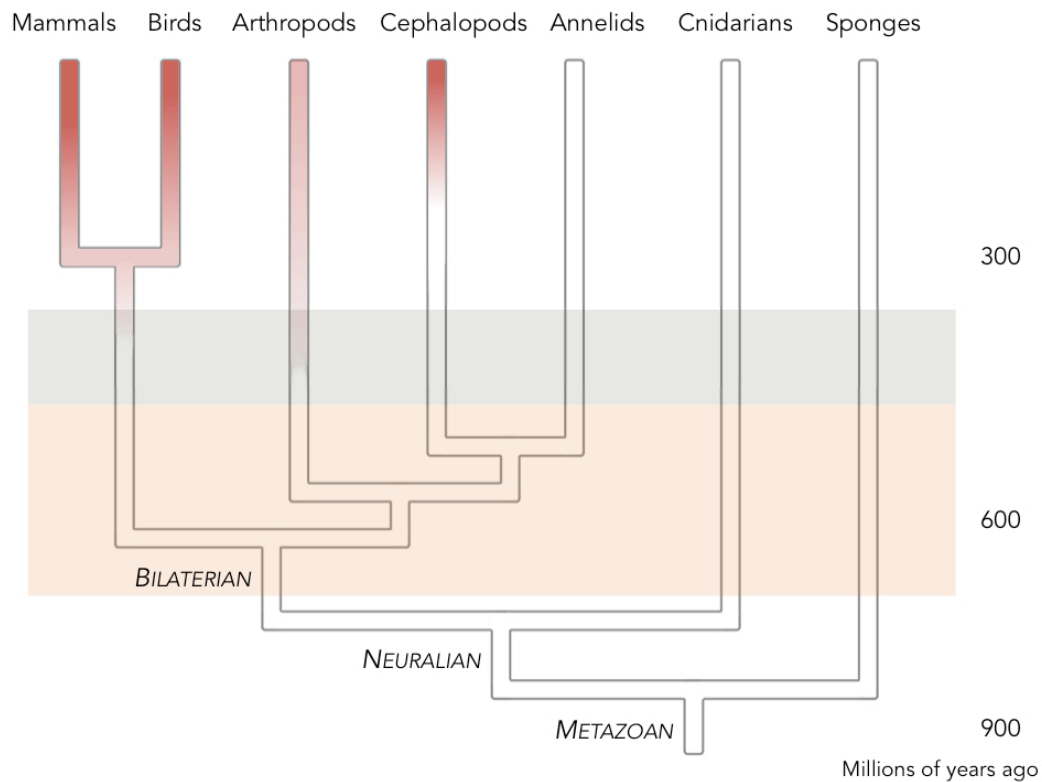
Figure 3. Part of the animal branch of the tree of life, with the graded shading within branches on the left showing the location of high levels of sensorimotor and cognitive complexity within particular groups.

Figure 3 has more than one interpretation. First, it is a picture of part of the animal tree in which the shading within branches represents the sensorimotor and cognitive complexity seen in some species within particular groups. It is also an attempt to represent complexity that is indicative of subjective experience, but I will reach that second interpretation in a moment. For now, set subjective experience aside and think of sensorimotor and cognitive complexity only. The figure doesn't represent overall values for each group, but *high* values within each group. The figure mixes taxonomic levels and leaves a great deal out; I'm very aware of the limitations of this kind of representation, but at a coarse grain, I think it's informative. The drawing shows a partially *parallel* development in these features, beginning especially in the Cambrian. I've replaced "molluscs" in Figure 2 with cephalopods, a small group within them. Within arthropods, in contrast, complex behavior is seen across several different groups – bees, spiders,

hermit crabs, mantis shrimp. Cephalopods of the relevant kind appear later than the other animal groups with complex behavior, but once they arise, they evolve very large nervous systems, especially in octopuses.[15]

Stepping back for a moment, the overall picture looks like this. Animals are a branch of the tree of life – one way of being a multicellular organism – unified even across non-neural groups by shared patterns of cell signaling. Nervous systems arose quite early in this branch, though, making use of pre-existing electrical capacities of cells, and create a new kind of control system. Bilaterian bodies, like ours, also seem to have evolved well before any regime of complex behavior. Such behavior evolved in the Cambrian, and it evolved in parallel in several groups, not as a radiation from a single source. This process was genealogically parallel but also coevolutionary, with one animal's changes making life more complicated for others. The graded shading within some branches is supposed to suggest that a lot of what happened along these lines was quantitative change, in a rough sense of that term. The behavioral advance is not one that can be measured on a single scale of complexity, but shows many forms – different kinds of sensory sophistication, different means for locomotion, different ways of acting on the world.

Now let's look at another interpretation of the figure. It can be seen, with its shading, as a hypothesized map of the evolution of subjective experience. To offer such any such interpretation we need to make further assumptions. I'll set up these next steps with a distinction between two general views. The first asserts a kind of *proportionality* – though that is not quite the right word – between the cognitive side of the mind and subjective experience. This can be seen as an application of a simple kind of functionalism. What we call the "qualitative" side of the mind is just the first-person point of view, the *subject's* point of view, on cognitive processes at work in the system. The qualitative is in no sense an *extra* feature of the mind, something that might in principle be absent; it's an insider's point of view on the cognitive.

This view gives us a way of thinking about the origins of subjective experience, a way that emphasizes gradients and differences of degree. It might be hard for us to

---

[15] See Darmaillacq et al. (2014).

imagine simple and minimal kinds of subjective experience, because we can't give ourselves the point of view of an animal very different from ourselves. But on the cognitive side, we can probably understand the grey areas quite well, and get a reasonable grip on the differences between a minimal scrap of mind and none at all. The qualitative will then exhibit a gradient of the same general shape. A failure of imagination is encountered on the qualitative side, but that's just a limitation in us. There can still *be* a gradient on the qualitative side that maps to a gradient on the cognitive side.

This view does lead to surprises. On the cognitive side, as I've emphasized, there is a gradient in complexity that stretches well past animals, all the way to unicellular life. In Figure 3 I started my shading in the Cambrian, but why not much before, in earlier animals or single-celled life, very faint? Isn't that the message of the analysis? A view like that initially looks absurdly generous, but it need not be; after all, gradients reach very low values.

Clearly there is much uncertainty about how the details of this first view would go, but the overall picture is one that makes sense. The second view I'll discuss has developed in more recent work – the literature has seen a shift, I think, from the first to the second. This view rejects any sort of "proportionality" assumption about the richness of the cognitive and qualitative. *Divergences* between these are now emphasized, and a large body of work charts the apparently quirky manner in which some of the cognitive activity going on inside humans has a subjective feel, along with much that does not. This I take to be the theme of much recent neurobiology (Dehaene 2014). There's a distinctive *style* of cognitive processing that brings with it subjective experience, embedded in much that does not. Work of this kind can motivate a view in which subjective experience is an evolutionary *latecomer*. The small fraction of what's going on in humans that has a subjective feel seems to be indicative of a particular way of organizing perception and cognition, a late-evolving way that features the achievement of forms of cognitive unification that many non-human animals probably not have.

I'll take a closer look at these arguments. First, it is uncontroversial that there is a lot of sophisticated processing going on in our brains that we do not subjectively experience. The initial stages of visual processing, and the processing of the syntax of sentences we hear, are standard examples. These, however, might have limited force as

arguments against the first view. Perhaps those kinds of processing (the second, in particular) are just like doing sums or accounts in the background, very different from the sensorimotor capacities we might associate with simple forms of subjectivity in animals. Other work, though, shows that there really is a problem here. An example is the work of Dale Milner and Melvyn Goodale (2005) on vision. They argue that there are two "streams" of visual processing in our brains. Only one, the "ventral stream," leads to experiences *felt* as vision. This stream is concerned with tasks like the categorization of objects. The "dorsal" stream handles tasks related to basic navigation, and dorsal stream vision feels, perhaps, like nothing – or perhaps like *something* but very different to vision. Here, allegedly, we have a sensorimotor arc guiding biologically important behavior in a way that does not give rise to subjective experience. Or if some faint subjective experience is present in this sort of perception, there is still a surprising divide between the cognitive and qualitative sides. Dehaene (2014) surveys a wide range of work that shows further divergences between complex cognitive processing and subjective experience; we do a huge amount of sensing and thinking in a way that feels like nothing at all.

It is sometimes unclear how work of this kind relates to the ideas developed earlier in this paper. Much of the neuroscientific work is presented as an investigation of *consciousness*. Some of the scientific writers may be using, tacitly, a framework similar to mine, in which a theory of consciousness is *not* a theory of subjective experience in the broadest sense. But other parts of this work do seem committed to the idea that recently-evolving sophistications are necessary for an animal to have any subjective experience at all.

What marks the difference between processes we experience and those we don't? A range of views are being defended. According to "workspace" theories, what we are conscious of is information made available in a "global workspace" that integrates information from various sources (Baars 1988, Dehaene 2014). This machinery of integration is something that many animals probably do not have, as it is machinery linked in specific ways to memory, attention and executive control. Views of consciousness that give a special role to "working memory," such as Jesse Prinz's AIR theory (2000) and Peter Carruthers's view (2015) have a similar character. All this work

shares the following picture: a lot of cognitive activity goes on in us that has no felt side, and we need to work out which are the special pieces that do have this feature. Once we find those special cognitive activities, and – better still – their neural correlates, we know what other animals need to have. It remains imaginable that animals without brains of the right kind do have subjective experience; those inner structures might only be necessary in our case, not in everything. But why should we believe this? To entertain that possibility threatens to remove the study consciousness from science once again. It is better to conclude that when these features first evolved, so did subjective experience, and not before; vague talk of "gradients" in this area does not take seriously what we have been learning.

If so, subjective experience is not something that arises in all animals with complex sensorimotor capacities, but only those with a particular kind of organization. There is a difficult question of what an animal has to be like to have machinery that is *close enough* to what enables subjective experience in us. Prinz (2000) thinks this question is probably unanswerable. But according to this second family of views, there is no reason to regard Figure 3 as telling us much about subjective experience. Instead we should probably shade a narrow band for subjective experience at the top of the mammal branch (or perhaps the mammal and bird branches) and leave the rest blank. The evolution of significant amount of sensorimotor complexity in other parts of the tree is beside the point, as an animal can have a lot of sensorimotor complexity and no subjective experience associated with it – we know that from our own case.

I'll now offer a reply to these ideas. I agree that some earlier work assumed too simple a mapping between cognitive and qualitative. A latecomer view is one response to this realization. There's an alternative, though, which I will call the *transformation* view. According to this view, some late-evolving features of our brains do greatly *affect* the nature of subjective experience, but they don't bring it into being. They modify more basic kinds of experience that were already there. They make it richer, perhaps, and bring it into contact with memory and verbal report. This may include pushing some kinds of subjective experience into the background, so far back as to make it hard to report on or remember. Basic forms of subjective experience were present earlier and require less, and in us these have been transformed.

What argument can be given for this view? Is it a vague plea for retention of a more generous attitude, and no more? The best argument I can offer at the moment is based on the ongoing role of what seem like old forms of subjective experience that appear as *intrusions* into more organized and unified kinds of processing. Consider the intrusion of sudden pain, or of what Derek Denton calls the "primordial emotions" – bodily feelings which register important metabolic states and deficiencies, such as thirst, and the feeling of not having enough air. As Denton says, these bodily feelings have an "imperious" role, when they are present: they press themselves into experience and can't easily be ignored. Do you think that those things (pain, shortness of breath, etc.) *only feel like something* because of sophisticated cognitive processing in mammals that has arisen late in evolution? I doubt it.

I will focus on the case of pain, and evidence for pain in animals that are unlikely to pass the tests for consciousness that people like Dehaene and Prinz would impose. Making this argument is not straightforward. One might initially say that it's obvious that even simple animals respond to pain in a way that indicates they feel it. But many responses to bodily damage that *seem* to involve pain and distress probably do not. For example, rats with a severed spinal cord, and hence no channel from body damage to the brain, can exhibit some of what looks like "pain behavior," and can also respond in quite sophisticated ways to the damage.[16] Given this, it is significant that other experimental work has shown that more complex pain-related behaviors are present in animals far from us on the evolutionary tree, including some invertebrates. What I see as important in this work is its indication that these animals respond to damage with more than reflexes, with modifications to their behavior that are flexible, sensitive to novelty, and balanced by other cost-benefit considerations.

The clearest results are in vertebrates, though some invertebrates have also shown this pattern. In one study, prior testing was used to work out which of two environments (empty or enriched) were preferred by some zebrafish. After injection with a chemical

---

[16] "The spinal cord distinguishes noxious stimuli from other stimuli, and adaptive changes in behavior result.... [L]earning about noxious stimuli can occur in the absence of conscious awareness of pain" (Allen 2004). For recent work on animal pain see also Key (2105), who defends a latecomer view, and Jones (2013).

believed to cause pain, the fish then preferred the normally less-favored environment when it had painkiller dissolved in it, and not otherwise: "the fish were willing to pay the cost of being in an unpreferred environment to obtain analgesia, and thus it can be inferred that these fish must have obtained some reward possibly in terms of pain relief such that the pain was reduced" (Sneddon 2011). Similarly, in a study in chickens, birds with damaged bodies chose a food that would usually be less preferred, provided that it contained analgesic: "lame birds selected significantly more drugged feed than sound birds, and... as the severity of the lameness increased, lame birds consumed a significantly higher proportion of the drugged feed" (Danbury et al. 2000).Finally, Robert Elwood reports that hermit crabs could be induced to leave their shell by a shock, but they were more reluctant to leave a higher-quality shell, or to leave when the odor of a predator was around: "hermit crabs trade-off competing demands in their responses to electric shock in a way that cannot be explained by a nociceptive reflex response" (Elwood 2012, p. 26).

It is important, also, that other animals appear to fail these tests. Crabs may be very different from their fellow arthropods, the insects. An older review, but not one that has been superseded as far as I know, says: "No example is known to us of an insect showing protective behavior towards injured body parts, such as by limping after leg injury or declining to feed or mate because of general abdominal injuries. On the contrary, our experience has been that insects will continue with normal activities even after severe injury or removal of body parts." (Eisemann et al. 1984).

These results do provide some support for a view of pain as a basic and fairly widespread form of subjective experience, one unlikely to be dependent on late-arriving mechanisms of working memory, integration of information, and so on.

One response to this argument is to say that it suggests that many more animals than we realized *have* the complex features that enable subjective experience in us, including fish and hermit crabs. It would require further empirical work to assess this view. Another possibility, one which surely becomes vivid once these results are on the table, is that there are forms of subjective experience that are simpler and older than the form of consciousness that recent neurobiological work on humans has been investigating. If so, there is something it's like to be a fish or hermit crab, even if (as I

would put it) they are not conscious.

If the arguments offered over the last few pages are right, the transformation view may well be correct, and the latecomer view is not as well supported as it might have appeared. A case can then be made for some sort of the separation of categories that have recently been conflated. There's the evolutionary origin of a subjective *feel* to life, in a very broad sense, and this was later shaped eventually into something with the familiar features of *consciousness*. In these pages I've not attempted to say much about the relation between those two things. My aim has been to say something about the evolution of subjective experience in a broad sense. The uncertainties in this area are enormous. The shape of the tree of life around the time that nervous systems first evolved is not yet clear, and there are puzzles about the relation between genetic and fossil evidence. But the idea of parallel evolution of sensorimotor and cognitive complexity from the Cambrian onwards is better supported. As I emphasized, this process was genealogically parallel but also co-evolutionary, with one animal responding to behavioral evolution in another. Any mapping between behavioral complexity of these kinds and subjective experience will also be controversial, at least for now, but I think Figure 3 might be a reasonable rough map of the history of subjective experience.

—————————————

References

Allen, C. (2004). Animal Pain. *Noûs* 38: 617-43.

Baars, B. (1988). *A Cognitive Theory of Consciousness*. Cambridge: Cambridge University Press.

Baker, M., P. Wolanin, and J. Stock (2006). Signal Transduction in Bacterial Chemotaxis. *BioEssays* 28: 9–22.

Budd, G and Jensen, S. (forthcoming). The Origin of the Animals and a "Savannah" Hypothesis for Early Bilaterian Evolution. *Biological Reviews.*

Carruthers, P. (2015). *The Centered Mind: What the Science of Working Memory Shows Us About the Nature of Human Thought*. Oxford: Oxford University Press.

Chalmers, D. (1996). *The Conscious Mind: In Search of a Fundamental Theory*. Oxford: Oxford University Press.

Danbury, T., C. Weeks, A. Waterman-Pearson, S. C. Kestin, and J. P. Chambers (2000). Self-selection of the Analgesic Drug Carprofen by Lame Broiler Chickens. *Veterinary Record* 146:307-311.

Darmaillacq, A-S., L. Dickel, and J. Mather (2014). *Cephalopod Cognition.* Cambridge: Cambridge University Press.

Dehaene, D. (2014). *Consciousness and the Brain: Deciphering How the Brain Codes Our Thoughts*. New York: Farar, Straus and Giroux.

Denton, D., M.J. McKinley, M. Farrell, and G.F. Egan (2009). The Role of Primordial Emotions in the Evolutionary Origin of Consciousness. *Consciousness and Cognition* 18: 500–514.

Dretske, F. (1986). Misrepresentation. In Radu Bogdan (ed.) *Belief: Form, Content and Function*, New York: Oxford University Press: 17–36.

Eisemann, C. W. K. Jorgensen, D. J. Merritt, M. J. Rice, B. W. Cribb, P. D. Webb and M. P. Zalucki (1984). Do Insects Feel pain? - A Biological view. *Experientia* 40: 164-167.

Elwood, R. (2012). Evidence for Pain in Decapod Crustaceans. A*nimal Welfare* 21: 23-27

Jékely, G. (2009). Evolution of Phototaxis. *Phil. Trans. R. Soc. B* 364: 2795–2808.

Jékely, G., J. Paps, C. Nielsen (2015). The Phylogenetic Position of Ctenophores and the Origin(s) of Nervous Systems. *EvoDevo* 6:1

Jékely, G., F. Keijzer, and P. Godfrey-Smith (forthcoming). An Option Space for Early Neural Evolution. *Philosophical Transactions of the Royal Society of London B*.

Jones R. (2013). Science, Sentience, and Animal Welfare. *Biology and Philosophy* 28: 1-30.

Keijzer, F., M. van Duijn and P. Lyon, What Nervous Systems Do: Early Evolution, Input–Output, and the Skin-Brain Thesis. *Adaptive Behavior* 21: 67–85.

Key, B. (2015). Fish do not Feel Pain and its Implications for Understanding Phenomenal Consciousness." *Biology and Philosophy* 30: 149-165.

Milner, D. and M. Goodale (2005). *Sight Unseen: An Exploration of Conscious and Unconscious Vision*. Oxford: Oxford University Press.

Marshall, C. (2006). Explaining the Cambrian "Explosion" of Animals. *Annual Review of Earth and Planetary Sciences* 34: 355–84.

McMenamin, M. (1998). *The Garden of Ediacara.* New York: Columbia University Press.

Moroz, L. (2015). Convergent Evolution of Neural Systems in Ctenophores. *Journal of Experimental Biology* 218: 598-611

Nagel, T. (1974). What is it Like to be a Bat? *Philosophical Review* 83: 435-450.

Nielsen, C. (2008). Six Major Steps in Animal Evolution: Are We Derived Sponge Larvae? *Evolution and Development* 10: 241–257.

O'Malley, M. (2014). *Philosophy of Microbiology*. Cambridge: Cambridge University Press.

Pantin, C. (1956). The Origin of the Nervous System. *Pubblicazioni della Stazione Zoologica di Napoli* 28: 171–181.

Parker A. (2003). *In the Blink of an Eye: How Vision Sparked the Big Bang of Evolution*. New York: Basic Books.

Perry, C., Barron, A., & Cheng, K. (2013). Invertebrate Learning and Cognition: Relating Phenomena to Neural Substrate. *WIREs Cognitive Science 2013*. doi: 10.1002/wcs.1248

Peterson, K., J. Cotton, J. Gehling, and D. Pisani (2008). The Ediacaran Emergence of
Bilaterians: Congruence Between the Genetic and the Geological Fossil Records
Philosophical Transactions of the Royal Society of London B 363: 1435–1443.

Prinz, J. (2000). A Neurofunctional Theory of Consciousness. In A. Brook and K. Akins
(eds.) *Cognition and the Brain: The Philosophy and Neuroscience Movement*.
Cambridge: Cambridge University Press.

Sneddon, L. (2011). Pain Perception in Fish Evidence and Implications for the Use of
Fish. *Journal of Consciousness Studies* 18: 209-29.

Spang, A. J. Saw, S. Jørgensen, K. Zaremba-Niedzwiedzka, J. Martijn, A. Lind, R. van
Eijk, C. Schleper, L. Guy & T. Ettema (2015). Complex Archaea that Bridge the Gap
between Prokaryotes and Eukaryotes. *Nature* 521: 173–179.

Trestman, M. (2013). The Cambrian Explosion and the Origins of Embodied Cognition.
Biological Theory 8: 80–92

Volkov, A. and V. Markin (2014). Active and Passive Electrical Signaling in Plants. In
U. Lüttge, W. Beyschlag (eds.), *Progress in Botany* 76. DOI 10.1007/978-3-319-
08807-5_6