Mind, Matter, and Metabolism

Peter Godfrey-Smith
City University of New York

For the NYU Modern Philosophy Conference, 2014. Work in progress.

1. Introduction

This paper is about the relevance to philosophy of mind of some biological topics – the nature of life in general, the evolution of animal life in particular. I'll look especially at a cluster of questions about qualia, consciousness, and the "explanatory gap." What I'd eventually like to do is put together a picture in which the basis of living activity in physical processes makes sense, the basis of proto-cognitive and then cognitive processes in living systems makes sense, and the basis of subjective experience in cognitive processes also makes sense.[1] I think that working through all the links here will make a difference; things won't look the way they do when we just ask: how can consciousness exist in a physical system?

This paper is a first sketch and some stages are more filled out than others. In the early sections I'll discuss, drawing on recent biology and biophysics, how living activity relates to the physical. The physical basis of life is quite definite and constrained, and different enough from other aspects of the physical to have consequences for the

[1] This sort of project goes back a long way. Recent works in the tradition include Thompson's *Mind in Life* and Deacon's *Incomplete Nature*. Thompson's book has influenced this paper.

mind/body problem. I'll also discuss the *proto-cognitive* nature of living activity. Thinking about life itself only goes so far, though, and the second part of the paper will be about the kind of life seen in animals. I'll look at a series of transitions in animal evolution which seem to have *some* role in the evolution of subjective experience, and connect these to the questions in philosophy of mind.

2. Changing views of life and mind

It used to be common to think of life as a sort of bridge between mental and physical. Aristotle's view of the different kinds of soul is a position of this kind. Descartes, in contrast, asserted a mechanistic view of life and isolated the mental/physical relation as the fundamental problem. Later views within materialist and "emergentist" programs revived claims of continuity – Spencer, Lewes, Dewey, Broad. I am curious what identity theorists like Place and Smart thought about the issue. The situation changed with the rise of computers and AI. This work seemed to show that some aspects of cognition are mechanizable in principle, and in a non-living system. There's no question of life being present in a classical AI system, or a familiar sort of robot, and given that there seems a real possibility that such a system might capture of all of mentality, there can't apparently be too close a link between life and mind. *Computation*, rather than life, became the crucial bridging concept between mental and physical.

Another development pushing the same way was a change in the understanding of life itself. I think there has been a partial deflation of the concept of life, especially when we compare it to its previous role. As I see the present situation, we have theories of the different things that living systems *do* – they maintain their organization, using energy and other raw materials, they develop and reproduce, and many of them perceive and behave. Our understanding of those activities is a "theory of life" of a sort that removes any appearence of a large-scale problem that might motivate vitalism. But there is no need to say much about which of these activities, or which combination of them, comprises *life*. As a result, biology textbooks feel able to say a few general things at the start of the book about what living systems do, without taking much of a stand on the nature of life. Living systems have distinctive features, and these tend to cluster. The

book will go on to say definite things about each of these activities; the books are not relaxed about all biological concepts, but they are about life.

This development makes it seem even less appealing to use life as a load-bearing concept in treatments of the mind/body problem. As life has become less mysterious, it has become less important as a tool. All these developments (around computers, and life) are reasonable, but I think that some of what has resulted is a wrong turn.

3. Matter at the scale of metabolism

If we look more closely at the "cluster-concept" treatment of life that has developed, the cluster has two main parts. Life has a *metabolic* side, and a side that has to do with *reproduction and evolution*. Living systems maintain their organization in the face of thermodynamic tendencies towards disorder and decay, by taking in raw materials and using sources of energy to control chemical reactions. They also reproduce and evolve. There's no point, I think, in asking whether a system has to do *both* these things to be alive, whether either is sufficient, or one is primary. That's a non-question. Instead the theoretical connections are something like this. Metabolisms are shaped through evolutionary processes. This involves a role for reproduction, as opposed to mere persistence. That's not because reproduction is essential in some way to metabolism – it is not. But a metabolic system that can multiply instances can evolve in ways that a non-reproducing system cannot, as the proliferation of each improvement creates many independent platforms on which further innovation can occur. Among metabolizing systems, then, those that can reproduce will become more complex and orderly as well as more common. Reproduction also requires control of energy somewhere in the system, though maybe not direct control by the reproducer itself. There are tight evolutionary connections between metabolism and reproduction, but no impediment to seeing them as different things.

The metabolic side of life, in a broad sense of that term, is the side that is important in this paper. Let's now look at what metabolism is like, especially at its physical basis. In this part of the paper I draw on some recent commentaries on cell

biology and biophysics, especially Peter Hoffman's book *Life's Ratchet* (2012), and Peter Moore's paper "How should we think about the ribosome?" (2012).[2]

Metabolic processes in actual cells occur at a particular spatial scale, the scale measured in nanometers – millionths of a millimeter. They also take place in a particular context, surrounded by water. In that context and at that scale, matter behaves differently from how it behaves elsewhere. In a phrase due to Hoffman, what we find is a *molecular storm*. There is unending spontaneous motion, which does not need to be powered by anything external. Larger molecules rearrange themselves spontaneously and vibrate, and everything is bombarded by water molecules, with the larger molecules being hit by a water molecule trillions of times per second. Electical charge also plays a ubiquitous role, through ions dissolved in the water and charged regions of larger molecules. The parts of a cell that *do* things in the usual sense – making proteins, for example – are subject to forces that are much stronger than the forces they can exert. The way things get done is by biasing tendencies in the storm, nudging random walks in useful directions, thereby getting a consistent upshot out of vast numbers of mostly meaningless changes. Moore, though not Hoffman, thinks we should conclude from all this that "Macromolecular Devices Are Not Machines." Moore thinks that a machine is a quite definite sort of thing, where low-level interactions are predictable and parts are tightly coupled. A storm-like collection of random walks influenced by friction, charge, and thermal effects, in contrast, is non-mechanistic.

The metabolisms of even the simplest known cells are also very complex, with many hundreds of chemicals involved. Some are more complex than others, but there are no known simple metabolisms.

Metabolism *happens* to operate at this special spatial scale, but does it *need* to be that way? Metabolisms are now very complex, but they, too, surely don't need to be? Surely they once were *not* complex, as they evolved from simple beginnings?

Here I go out on some limbs and my biochemical knowledge is not strong enough to defend things in detail. But my impression of the way work is going in this area is as follows: it's probably not true, first, that present-day metabolisms evolved from very

---

[2] I've also benefitted here from discussion with Derek Skillings. See his "Mechanistic Explanation of Biological Processes" (PSA 2014).

simple ones. There probably never were any simple metabolisms – simple in the way that older models of the origins of life are simple. The transition that occurred went not from simple to complex, but from disorderly to orderly. Disorderly complex chemical systems gave rise to more orderly complex ones. When I say that, I am stating explicitly something that I take to be a *semi*-explicit message of a lot of more detailed work.[3] A reason to believe this comes from the inevitability of *side-reactions* in chemical systems. Simple metabolic models use imaginary chemistries in which each part has only one or two effects. In real chemistry, the parts have many effects of different sizes. The evolution of life was a matter of channeling and taming this sea of interactions, not taking a few simple interactions and stringing them together. Once you have basic metabolisms, they can become more complicated. But, again, the simplest ones are very complicated, in comparison to earlier models and present artificial systems, and there's a good chance things have always been that way.

"Metabolism *happens* to operate at the nanoscale in a molecular storm, but does it *need* to?" How contingent are the special features of material interaction in living systems? I am not going to ask a series of "nomically possible? logically possible?" questions, but let's ask in a more informal way how hard it would be for things to be different, given the nature of matter and how matter comes to be laid out on planets. The physical features of metabolism discussed above may be very far from accidental. Hoffman argues that the scale and chemical context seen in actual present-day metabolisms is the only place where we will find devices that run themselves: "The nanoscale is the only scale at which machines can work completely autonomously." At this level there is spontaneous motion, but there is enough structure and the relations between forces are such that a lot can happen, by biasing tendencies in random walks. It is at least very difficult, then, for life to *arise* outside this scale and context. Life could not have arisen in a dry-land macroscopic realm, on the scale of familiar machines. Perhaps once life exists in a "chemically easy" form, artifactual systems can be made that have different relations to energy and self-maintenance. The message I am trying to

---

[3] Orgel, "The implausibility of metabolic cycles on the prebiotic earth," Száthmary, "The hard problems of the origin of life," Koonin and Martin, "On the origin of genomes and cells within inorganic compartments."

emphasize, though, is that things are more constrained in this area than quick acts of imagining would suggest.

Some issues in the "mind/body problem" are about how any sort of mind could be physically realized. Others are more concerned with our minds, and their actual physical basis. I am aiming for claims of the general kind, but here is a point that is more local to our case. Let's consider again arguments against materialism based on conceivability, and the apparent separability of the mental and physical (in Kripke, Chalmers, and others). One kind of argument begins with the fact that it seems that we can conceive of an exact physical duplicate of an ordinary human, where the duplicate does not have any subjective experience. It's said that this exercise shows the separability of mental and physical, and hence the failure of materialism.

I don't think we need the ideas in this paper to reject those arguments. The general reply is that although there is indeed an imaginative act we can engage in that shows this apparent separability, it can be diagnosed as arising from quirks of the imagination, especially from the separability of what Nagel once called "sympathetic" and "perceptual" imagining (sympathetic: imagining *being* something; perceptual: imagining *seeing* it). If materialism was true it would still seem false, because of how our imaginations work.[4] But the ideas above do add something. In us, the material basis for mental activity is tied to cells and metabolism. When we look at what's actually going on in our bodies and brains, we find that many of the imaginatively familiar features of the physical are not present. Many of the features of the physical that *seem un-mental* are not present. And it is difficult to imagine the crucial processes at all, to get any sort of intuitive handle on what they are capable of.

In the context of this conference it is natural to link this point to Leibniz and his "mill" argument against materialism (*Monadology,* section 17).

> [W]e must confess that perception, and what depends upon it, is inexplicable in terms of mechanical reasons, that is through shapes, size, and motions. If we imagine a machine whose structure makes it think, sense, and have perceptions, we could conceive it enlarged, keeping the same proportions, so that we could enter into it, as one enters a mill. Assuming that, when

[4] Think of it in Bayesian terms: the evidence (in imagination) is equally likely given the truth and the falsity of materialism, so the prior probabilities, whatever they were, remain unchanged.

> inspecting its interior, we will find only parts that push one another, and we will never find anything to explain a perception. And so, one should seek perception in the simple substance and not in the composite or in the machine.

If we were observers at the intra-cellular scale, things would not look to us as Leibniz describes. Leibniz's imagined mill was a macro-scale dry-land object. An aqueous nano-mill would be a very different place. Inside a cell we would see some "parts that push one another," but not in the manner of macroscopic machines, and we would not only see pushes. We would see a storm of activity biased by charge and shape, generating an enormous number of random walks that, on average, tend in orderly directions. The processes are more causally holistic, noisier – more a matter of "herding molecular cats" – than a push-pull model allows.

Explaining how the whole process amounts to human "perception," as Leibniz asked, still requires working at a different level of description from the intracellular (as in a standard reply to Leibniz's argument), but what seemed to be an obvious antipathy between mental and physical is much reduced. Our immediate imaginative response to the scene would surely tend towards panpsychism, if anything. That would be another mistake. But macroscopic machines provide a poor model for the material basis of living activity, and for the material basis of mental activity in living things like us.

Suppose our brains did contain processes of the sort Leibniz envisaged – smaller versions of mechanistic processes of the sort familiar from the macro level. Might we then conclude that subjective experience *is* impossible in such a system, without some other contribution? Might Leibniz have been right about that much? I suspect that the answer is moot (or a faint yes for Leibniz): our brains could not have turned out to be that way.

4. Life and cognition

Next I'll look at the other side of this "bridging" role that biology may play. This involves the link between living activity (in the metabolic sense) and the mind.

Starting with some obvious facts: all the systems we encounter that are clear and uncontested cases of systems with minds are also living systems. The same is true of

nearly all the usual contested candidates – here I'm thinking of simple animals and very sophisticated AI systems. A converse principle is also true: all known (metabolically) living systems engage in some cognitive or proto-cognitive processes. I'll say more about the "proto-cognitive" in a moment, but we can think of it initially as involving sensing and responding to events, perhaps in minimal ways, and doing so in a way that helps keep the system alive. I'll discuss the status of both generalizations further in a moment, but first I'll say more about the second one by looking at some examples at the low end of the scale with respect to complexity: bacteria.

Bacteria (and archaea, which are superficially similar but distant in evolutionary terms) are the simplest known systems with a metabolism. They do a fair bit of sensing and responding to events around them. I'll divide the phenomena into two main categories, one that involves gene regulation and another category that contains everything else.

First, a lot of the "control" processes seen in bacteria work through the genome, by the regulation of gene expression. The output of these systems is chemical, rather than "behavioral" in the usual sense. Genetic systems in all cells work via processes with a quite strongly computational character, with cascades of interactions that can be described in terms of *and*s, *or*s and *not*s. This looks like an immediate help to my case in this paper, but that is not really so straightforward. A person might say that *computation* is the crucial concept here, and computation is seen both inside and outside living systems. Computation is important for genes and important for thinking, but those are separate matters and computation also does not have any essential connection to life. That would be a reasonable point, and I think it shows that describing the biological role of computation, in an ordinary sense of that term, is not enough. But what we see prefigured in basic kinds of (metabolic) life is something more specific. It is the use of sensing and responding, often coordinated with boolean or boole-approximating operations, to maintain the integrity of a system and its activity, seeking and maintaining some states while avoiding others. A complicated collection of *and*s and *if-then*s with no metabolic point to them would not be the same sort of thing. When the genome is used to control the synthesis of metabolically important chemicals by means of feedback, or by tracking

conditions in the external environment, *that* is proto-cognitive in the sense I have in mind.

The second category consists of proto-cognitive control devices that are not immediately dependent on the genome for their operation (though they do depend on it for their construction). A good example is chemotaxis (seeking or avoiding chemicals) in the bacterium *E. coli*. This system controls bacterial swimming. It is quite a smart system – much smarter than the common example of magnetotaxis. *E. coli* chemotaxis makes use of memory: swimming choices at each time-step are 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*, then the cell swims straight. If they are getting worse, the cell takes a random "tumble."

Is this sort of thing *always* present in cells? Even if it is always present now, does it need to be? Perhaps proto-cognitive activity is a good idea for any (metabolically) living thing, and hence it readily comes about, even though it has no necessary connection to the metabolic side of life. If a metabolically active system had an easy enough environment, might it get away with *none* of this? How lacking in proto-cognition could a viable living system be? Locomotion, for example, is optional for bacteria, not essential, though the majority of bacteria apparently do it.[5] I said earlier that control of metabolism by sensing and feedback, also, is proto-cognitive. But if the environment was easy enough to deal with, might a cell just keep its reactions running in a "dumb" way with no proto-cognitive control?

If we look at the simplest organisms that are known, the bacteria called *Mycoplasma*, we find that they do engage in adjustment of metabolism to external events. One of the findings taken to be quite striking when these organisms were recently studied was the dynamic nature of their gene regulation, and their adjustment of metabolism to conditions. Mycoplasma are not ideal examples, though, because they went backwards with respect to complexity, from ancestors with larger genomes and more complicated metabolisms. They are not remnants of old forms, but cases of reduction due to a parasitic lifestyle.

---

[5] "The majority of bacterial species can swim," Armitage and Scott, "Bacterial Behavior," 2013. (*The Prokaryotes*, 4th edition, Springer).

Perhaps there are well-studied cases I don't know of which do approximate being metabolisms with no proto-cognitive adjustment of activity to conditions, and perhaps there are cases that are presently unknown. It would take a lot to show that there *never* were any organisms like this; it would take a lot to show that the only ways to maintain a viable metabolism include processes that can, on independent grounds, be considered proto-cognitive. (I say "on independent grounds" because it might be claimed that *self-maintenance* is itself proto-cognitive. If so, all metabolism has a proto-cognitive nature, but only by a questionable extension of the latter category.) What we can say, at least, is that proto-cognition comes along very readily. It is widespread in bacteria and present also in archaea. The bacteria/archaea split is the oldest evolutionary split between kinds of life on earth, dating from something like 3.5 billion years ago. Archaea have been studied less than bacteria, but some can swim much faster than a cheetah can run, if speed is measured in body lengths per second.[6]

So at present I think there is both theoretical and empirical uncertainty about how closely proto-cognitive activities and metabolism are connected. When I say "theoretical" uncertainty, I mean that is unclear where the boundaries of the proto-cognitive lie. This boundary will not be sharp, and non-competing broader and narrower concepts will probably be defensible, but a much better specification than the one I've been using here ought to be possible. It would be a mistake to trivialize the connection by saying that self-maintenance itself is proto-cognitive, but it might also be a mistake to restrict the concept to capacities that have a sensorimotor character. And then empirically, we would like to know what the lower limits are on control-related activity in a real metabolism.

One view that might be defended is that proto-cognitive abilities are entirely distinct from metabolic life, but are a natural and expected addition, something living

---

[6] "Diurnally Entrained Anticipatory Behavior in Archaea," Kenia Whitehead et al. *PLOS* 1 2008. Swimming behavior of selected species of Archaea. Herzog and Wirth. *Appl. Environ Microbiol*. 2012. "The two Euryarchaeota M. jannaschii and M. villosus were found to be, by far, the fastest organisms reported up to now, if speed is measured in bodies per second (bps). Their swimming speeds, at close to 400 and 500 bps, are much higher than the speed of the bacterium E. coli or of a very fast animal, like the cheetah, each with a speed of ca. 20 bps. In addition, we observed that two different swimming modes are used by some Archaea. They either swim very rapidly, in a more or less straight line, or they exhibit a slower kind of zigzag swimming behavior if cells are in close proximity to the surface of the glass capillary used for observation. We argue that such a "relocate-and-seek" behavior enables the organisms to stay in their natural habitat."

systems will quickly gain. Another possible view is that they are more inextricably tied together, that proto-cognition is an inevitable aspect of a functioning metabolism. Closer ties might also be present in some particular kinds of life. Consider multicellular life, for example. In multicellular organisms, a great deal of signaling between cells goes into the making of the body itself, and keeping that body running. "Cell-to-cell signaling" is only signaling in a rather minimal sense, but this is not just any old causal network; there are specialized producers and receptors of evolved signal molecules.

Even if proto-cognition is not inextricable from metabolism in principle, metabolic processes – and their special material basis, discussed in the previous section – are what cognition grows out of. In organisms like us, the line between the "information processing" side of brain activity and the metabolic side is very porous, not really a line at all.[7] In biological systems, the active structures change just from being used – they reflect their immediate history, in a way that computers do not (or don't unless programmed to) – and are weakly affected by what many other parts are doing. If you watch how the same neuron responds to the same stimulus, it does not simply reproduce the same pattern of firing, but behaves slightly different each time (see Wu et al. 1994, Gal et al. 2010).[8] This sensitivity to history and context is not merely "noise," but raw material on which adaptive plasticity can be built.[9] The character of cognition in living organisms is affected by its embedding in metabolic processes. The idea of a non-living "functional duplicate" of a living system like a person, routinely appealed to by philosophers, is quite problematic. There might one day be functionally *similar* non-living systems, to some appreciable degree, but metabolic activity is part of the "functional" profile of a human agent.

I'll make one more point – or reiterate and extend an earlier one – about the nature of proto-cognitive processes in simple organisms. What's important is not merely something like logic, or something like communication. The term "cognitive" is a flexible

---

[7] See, for example, Moore and Cao, "The Hemo-Neural Hypothesis: On The Role of Blood Flow in Information Processing," *J Neurophysiol* 99 (2008).This section of the paper has been greatly affected by discussions with Cao.

[8] The Wu et al. paper is "Consistency in Nervous Systems: Trial-to-Trial and Animal-to-Animal Variations in the Responses to Repeated Applications of a Sensory Stimulus in *Aplysia."* J. of Neuroscience, 1994.

[9] See Ralph Greenspan's 2007 *Introduction to Nervous Systems*: "If the intrinsic variability in the nervous system cannot be reined in, then perhaps it is exploited." (p. 70)

term of art which certainly can cover things like that. What's important is something more specific. Simple organisms sense and respond to events, both internal and external, in a way that helps keep them alive, implementing a distinction between states and outcomes that are sought and maintained and other states and outcomes that are avoided. Boundaries between the system and its environment are controlled. At the risk of proliferating protos further, there is *proto-subjectivity* present. I don't mean this in a sense that implies first-person felt experience, but in a weaker sense. Simple organisms, in their sensing and acting, are subjects and have a point of view in a richer sense than, say, a digital camera, which also senses and responds and computes, but does not control its boundaries and maintain a metabolism. Further, it might be argued that metabolism itself, with or without a proto-cognitive side, brings proto-subjectivity into being.

Making another brief historical excusion, we can summarise some of this section by saying that if Aristotle's nutritive soul corresponds to metabolism and his sensitive soul corresponds to proto-cognition, then there is less of a gap between those two than he (understandably) supposed. Certainly the proto-cognitive is not the particular domain of animals; it is found in at least nearly all of life, perhaps in all.

5. Transitions in animal life

So far I've been discussing (metabolic) life in general, and how it helps us think about matter on one side and mind on the other. I think that some progress can be made through this general reshaping of the terrain, but it can only take us so far. The main points about life made above apply as much to bacteria and plants as to animals like us. *How* far those points take us depends on questions about continuities and discontinuities, and the current literature contains an extremely broad spectrum of views on that issue. The literature includes a revival of panpsychism in a full-fledged form (eg., Galen Strawson, Philip Goff, sympathy from David Chalmers), along with a near-panpsychist view from Guilio Tononi, based on a measure of "informational integration." This view (endorsed now also by Christof Koch) holds that all systems containing interactions that can be described in terms of information flow have some sliver of consciousness – even a simple non-living switching device. Given the previous sections, it's natural to put another radical view on the table. I'll use the term *biopsychism* for the view that all and only systems that are

alive, in the metabolic sense, have subjective experience. (The term was introduced by Ernst Haeckel, 1892, with a meaning close to this. Haeckel himself endorsed panpsychism of a very robust form.) According to biopsychism, the low end of the mental scale is inhabited not by simple matter (panpsychism), or simple machines (Tononi and Koch), but the simplest forms of life.

Biopsychism can be motivated by combining the ideas in previous sections of this paper with a simple form of functionalism. Suppose you think that the "mental" has two sides, a qualitative side and a cognitive side. Both are found in different degrees of sophistication. Proto-cognitive activity is found in the simplest metabolizing systems, including bacteria. But, on this view, the qualitative and cognitive sides of the mental are tied closely together – the qualitative is just what the cognitive feels like from the inside. Simple functionalism about the mind then tells us to embrace a gradient, from a lot to a little, for both the cognitive and qualitative, and minimal forms of subjective experience should then be found in very simple living things. It is very hard for us to think about the low end of the scale with respect to the qualitative side, because "thinking about" it involves the use of our sympathetic imagination – imagining what it would feel like to be such a system – and here our sympathetic imagination founders. But that's our fault, not the fault of biopsychism.

I don't think biopsychism is insane, though I don't believe it. It certainly has advantages over the other radically generous views. A living system maintains itself, controls its boundaries, and thus has a kind of proto-subjectivity. There are no very simple living systems (though there once, presumably, were systems with marginal amounts of the relevant kind of order). Biopsychism in something like this sense has been endorsed explicitly by some people – by Herbert Jennings round 1900, Lynn Margulis, and Pamela Lyon.[10] We don't know how to think about the difference between a complete absence of subjective experience and a minimal but nonzero scrap of it, and this inability makes all these "generous" views hard to assess. If functionalist arguments, along with further information about proto-cognition in unicellular life, motivate a biopsychist view, then what is to tell against it other than sheer weirdness?

---

[10] As I read Evan Thompson's *Mind in Life*, he does not endorse it, though he seems tempted.

Though I don't write off biopsychism, I'll spend the rest of my time working within the more familiar range of views that hold that some distinctive features of *animal* life are essential to the presence of subjective experience.[11] This is the family of views that many of us now spend time navigating hesitantly around, trying to find criteria that divide the cases sensibly – striking a balance between views that are too brutally human-centric, on one side, and views that are so generous they seem flaky, on the other. Before I start, I'll make a point about the set-up. This topic is often now discussed as the problem of *consciousness*. Thirty years or so ago, people usually said there are three main problems in the philosophy of mind: qualia, consciousness, intentionality. The problem of qualia was seen as the problem of explaining the feel of the mental, and consciousness was seen as a more sophisticated kind of cognition with special properties, including a special qualitative side. Now "qualia" and "consciousness" are often seen as amounting to the same thing, a change in set-up largely due to writers who happen now to be at NYU (Nagel, Block, Chalmers). In particular, if there is something it *feels like to be* a system (in Nagel's phrase), then the system is said to have a kind of consciousness (perhaps "phenomenal consciousness"). 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 consciousness. Pain is a good example; I wonder whether squid feel pain, whether damage feels like something to them, but don't think of this as wondering whether squid are conscious. As well as pain, there are the states 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.[12] Like Denton, I think of these as good candidates for being the most basic states that can feel like something for an organism. You can say they are kinds of consciousness if you want, but I'll use the terms *qualia* and *subjective experience* for the broadest category of

---

[11] Haeckel in the same 1892 paper coined the term "zoopsychism," but *not* for the view that all and only animals have minds (as one might expect). He set this up as a view about the more complex forms of soul seen in higher animals and in man.

[12] See Denton et al, "The role of primordial emotions in the evolutionary origin of consciousness," 2009.

phenomena here, also describable by saying that some states of some systems feel like something to the system and others do not.[13]

Next I'll survey some transitions in the history of animal life which seem to have some relevance to the origin of subjective experience.

5.1. Cytoskeleton: The first transition I'll mention comes *before* animals, when all life was single-celled. One part of the evolution of the eukaryotic cell (the larger and more complex cells that make up animal bodies) was the evolution of the *cytoskeleton*. This is a skeleton-like collection of internal fibers whose movements can be controlled through metabolism. In particular, this makes possible the contraction of parts, and hence changes of shape in the cell as a whole. The evolution of the cytoskeleton was a landmark in the evolution of behavior. When I discussed bacteria above, there were two kinds of "output" described: making chemicals and locomotion. These are the main things bacteria can *do* (along with a couple of other things involved in predation and gene exchange – though perhaps there are phenomena I don't know about here). With the evolution of the cytoskeleton, much richer forms of behavior become possible for cells, as the whole body can change shape in finely controlled ways. (An amoeba is a good example). This is the beginning of non-trivial manipulation of objects and new kinds of locomotion. When richer forms of behavior become available, richer forms of sensing are also worth having.

5.2. Multicellularity: Animals arose perhaps 900 million years ago, as one of several independently evolving forms of multicellular life. Multicellularity makes possible differentiation and specialization of parts with respect to the proto-cognitive capacities that had been crammed previously into single cells: sensing, processing and integrating what is sensed, and acting. Some of this specialization can be achieved without a nervous system, but only a very small range of extant animals do not have nervous systems – sponges, placozoa, and a few reduced oddities – and some of those are borderline cases. So I'll move on to the next transition.

---

[13] I use "qualia" in a minimal sense, without the baggage that Dennett criticized in "Quining Qualia."

5.3. Nervous systems: Nervous systems probably arose quite quickly in the animal branch of the tree of life – perhaps 700-800 million years ago, though these dates are very uncertain. I said above that nearly all animals have nervous systems, but said that without addressing the question of what a nervous system *is*. This is not a question with a straightforward answer. In a paper I am co-authoring with Gáspár Jékely and Fred Keijzer, we find it necessary to introduce two definitions of "neuron," a broader and narrower sense.[14] In the broader sense, a neuron is an electrically excitable cell that influences another cell by means of electrical or secretory mechanisms. This functional definition includes as neurons some cells in plants and the "non-neural" animal *Tricoplax*, and also some cells in animals that would usually be called muscle cells. That is fine, we think, but for other purposes it makes sense to use a somewhat narrower definition of a neuron that includes an anatomical criterion: a neuron in the narrow sense is an electrically excitable cell that influences another cell by means of electrical or secretory mechanisms (as above), and whose morphology includes specialized projections. This is still a broader definition than some (synapses are not essential, for example). In the narrower sense, some important generalizations can be stated: all and only organisms with neurons also have muscle cells. Muscle and neurons seem to have co-evolved.

What were the first animals with neurons like and what did they do with them? Very little is known about this. They might have been jellyfish-like – filmy and soft, living either on the sea floor or in the water column. There is a good chance that a lot of what they did with them was not closely associated with what we'd now call "behavior"; early neural control systems may have had a lot to do with control of metabolism and development. And behavior, too, might have been a different affair.

The first animals whose lives we know something about from the fossil record lived in the Ediacaran period, around 635-540 mya. And once we find animals whose lives we can say something about, we see something of philosophical interest. Many animals in the Ediacaran seem to have lived on the sea floor, grazing on microbes or filter-feeding. Based on genetic evidence, it seems extremely likely that some of them had nervous systems. What were they doing with them? We can make some (defeasible)

[14] Jékely, Keijzer, and Godfrey-Smith, "An Option Space for Early Neural Evolution."

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.[15] In the paper I mentioned with Jékely and Keijzer, we distinguish *internal coordination* versus *input-output* roles for early nervous systems. I think that in the Ediacaran, a lot of it was internal coordination. What was going on that needs *reacting* to? Not much. Nervous systems in the Ediacaran may have functioned mostly in "pulling the animal" together, enabling simple locomotion and feeding, and controlling physiology and development, without complex real-time sensorimotor arcs being present.

5.4. Sensorimotor complexity and CABs: Then we reach the "Cambrian explosion," in which many new kinds of bodies appear. Again, from these bodies we can make some 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 now form a family that has particular importance here. These views hold that at least one important thing that happened in the Cambrian was a process of feedback that linked the evolution of bodies with the evolution of new kinds of behavioral interaction. Predation arose – seen clearly in the fossils – and with predation, a series of "arms races" followed, rapidly improving the senses and the means for bodily action. The evolution of image-forming eyes, means for tracking other animals at a distance, may have been particularly important, a trigger for other changes. Whether eyes were pivotal or not, for the first time, the *details* of what was going on around an animal came to matter to its life and prospects. When predators or prey are around, you must react in real time to the contingencies of your external environment, to the moment-to-moment changes in what is around you. This rather obvious feature of present-day animal life was probably not in place before the Cambrian.

---

[15] There is one possible known exception, some *Cloudina* fossils in China.

Michael Trestman outlines a useful category here: *complex active bodies* (*CAB*s).[16]

> 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).

CABs originate in the Cambrian, probably first in athropods. With these bodies, the role for nervous systems that we are familiar with – the fine-grained linking of perception and action – becomes prominent. Some evidence suggests that associative learning may also date from this time, and Ginsberg and Jablonka (2010) suggest that this itself may have been a factor in the feedback driving change in the Cambrian.

If this is right in broad outline, we reach the following picture. The first nervous systems did 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, learning and remembering. Eventually, these did become central to animal life, in a process that began 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.

I will leave the historical story there. To stop here is surely to stop well before we get to most 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 in stages just 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

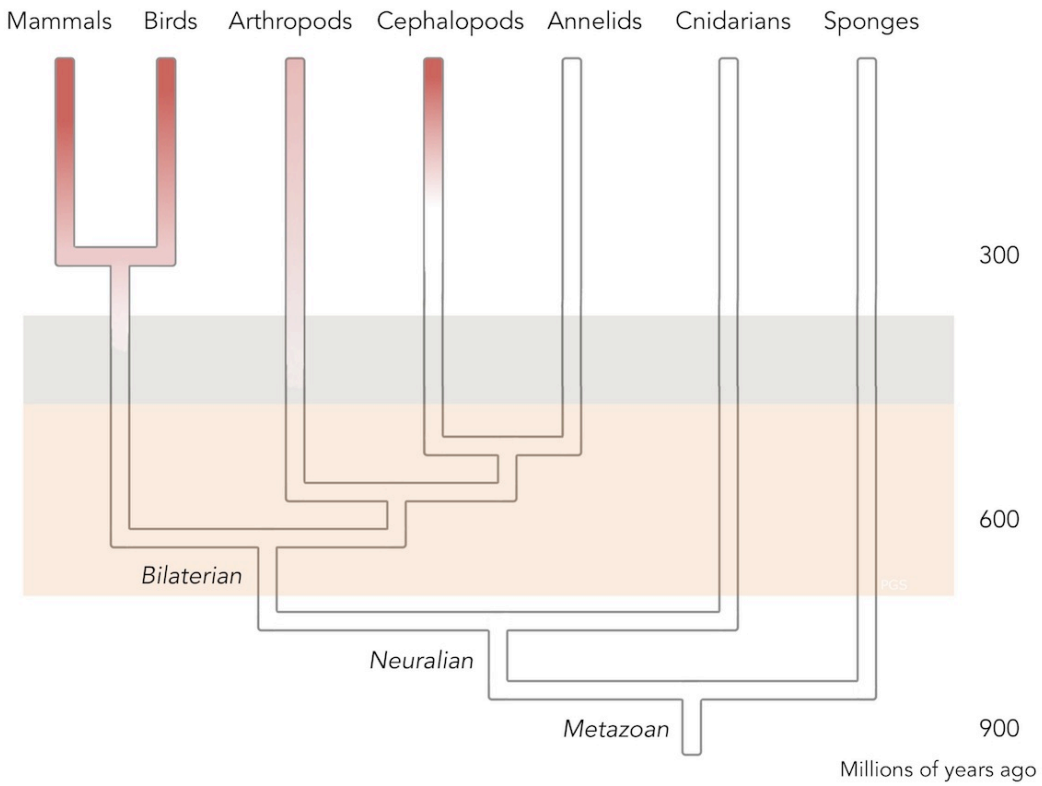
---

[16] "The Cambrian Explosion and the Origins of Embodied Cognition," *Bio. Theory*, 2013

more quantitative character: more neurons, more sophisticated learning and categorization, more complex behavior. These changes are seen along several independently evolving lines – especially in some vertebrates, some arthropods (eg., bees, spiders, mantis shrimp), and a few molluscs (cephalopods). Earlier I discussed thinking in terms of gradients of complexity for both the cognitive and qualitative sides of the mind. Whether or not that is misplaced when applied to all forms of life, from this point forward some sort of gradient view seems well-motivated. So in the sea of changes that link us with bacteria and non-living matter, I am emphasizing two in particular: the evolution of metabolic life, and the evolution of animals with rich sensorimotor arcs, including the feedback that links actions taken now with what is sensed on the next time-step.

Here below is a figure I discuss in more detail in another paper, tracing out each of the lines. The figure can work as a rough guide here, though. The figure gives one view of history of animals, with an orange band for the Ediacaran and grey-green above it for the Cambrian. The lower branchings and dates are all controversial.[17] The Metazoa are animals, the Neuralia are animals with nervous systems, and the Bilaterians are animals with left-right symmetry, like us. The red shading shows increase in *sensorimotor and cognitive capacities that I see as relevant to subjective experience.* I don't represent overall values for each group, but *high* values within each group, and I mix taxonomic levels and leave a great deal out. I'm very aware of the limitations of this kind of representation, but at a very coarse grain, I think it's OK. I think of the chart as a rough candidate map of the evolution of subjective experience.

[17] See "On the Origins and Distribution of Consciousness" for some of the controversies, especially regarding the category "Neuralian" and the placement of sponges. I assume here that nervous systems evolved once – another topic of debate at present. In the case of arthropods, complex behavior is seen in a range of different groups – spiders, bees, and mantis shrimp, in particular. Within molluscs, it is confined to cephalopods.



Part of the animal branch of the tree of life, with red shading showing the location of high levels of sensorimotor and cognitive complexity within particular groups.

To close, I'll look at a very different view of the matter. I said at the start of this section that if we work within a simple form of functionalism, it's natural to envisage a sort of "proportionality" between the cognitive and qualitative sides of the mind. I think that a lot of work during the latter part of the previous century was set up like this – my title, for some, will recall Bill Lycan's paper "Form, Function, and Feel" from 1981, a paper that exemplifies this view. An important theme of more recent work, though, sometimes explicit and sometimes implicit, is a rejection of any sort of proportionality assumption about the richness of the cognitive and qualitative. *Divergences* between these are now emphasized, and there is a lot of work that 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.

Work of this kind can motivate a view in which subjective experience is an evolutionary *latecomer*. A lot of what goes on in humans has no subjective feel, and what does have this "feel" appears to be indicative of a particular way of organizing perception and cognition, a way that features the achievement of forms of cognitive unification that many non-human animals probably not have. When those things evolved, it might be argued, so did subjective experience – and not before; vague talk of "gradients" in this area does not take seriously what we have been learning.

Complex cognitive processes that appear to have no subjective feel include much of "early vision" and unconscious language processing. An especially interesting example is a body of work on "dual stream" models of vision in mammals, especially work by Milner and Goodale (eg., *Sight Unseen*, 2005). They describe what they take to be two streams by which visual information is processed in the brain, with quite different roles. 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 does so in a way that can produce effects akin to "blindsight," where a person denies being able to see but can act on some visual information.

Another relevant body of work concerns "workspace" models of consciousness, pioneered by Bernard Baars and developed by Stanislas Dehaene and others. According to these views, what we are conscious of is information in a specific part of the brain that functions as a "global workspace," integrating information from various senses. This machinery of integration is something that many animals probably do not have, and it's certainly an *extra* piece of machinery, linked to attention and executive control. Views of consciousness that give a special role to "working memory," such as Jesse Prinz's AIR theory have a similar role. All this work shares the following picture: a lot of cognitive activity goes on in us that has no felt side, no associated subjective experience, and we need to work out which are the special pieces that do have this feature. Once we find those special functional features, and – better still – their neural correlates, we know what other animals need to have. They don't need to have *exactly* the things we have, as their

experience will not be the same as ours, but *something like* them. A further hard task will then be working out what count as the relevant kinds of similarity.[18]

Some of this recent work on "consciousness" written by scientists probably uses categories similar to mine, in which a theory of consciousness is *not* a theory of subjective experience in the broadest sense. They might think that basic forms of pain, for example, can exist without "consciousness." But other parts of this work, especially by philosophers, are written within the newer set-up and hence seem committed to the idea that recently-evolving sophistications are necessary for an animal to have any subjective experience at all. Carruthers and Prinz think something like this, as I understand them.

I agree that some earlier work made too simple an assumption about the mapping between cognitive and qualitative. A latecomer view is one response to these findings. There's an alternative, though, which I will call the *transformation* view. According to this view, special 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 aleady there, and this may include pushing some stuff 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 since 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 the ongoing role of what seem like old forms of subjective experience that appear to be *intrusions* into more organized and unified kinds of processing. Consider the intrusion of sudden pain, or the intrusion of the "primordial emotions" in Derek Denton's sense. 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 the sophisticated cognitive processing in mammals that has arisen late in evolution? I doubt it.

---

[18] Prinz (2000) claims that this latter task can never be done, so a qualified "mysterianism" must be accepted.

In reply, you might make an argument like this: for there to be subjective experience, there must be a *subject*. That requires a certain kind of psychological *unity*. The cognitive mechanisms I've been describing all have a link to unification – certainly that is true of the workspace, and Milner and Goodale make comments about ventral stream vision that suggest something like this.[19] My reply is as follows: to the extent that unity is an issue here, simpler animals do have it. An animal has it in virtue of its status as a real sensorimotor unit, the locus of coordination between perception and action. This is part of what becomes laid down, or at least greatly "firmed up," with the evolutionary changes stemming from the Cambrian. Especially if you are a biopsychist, you might add that metabolic life itself creates non-arbitrary units that are plausible subjects in a minimal sense. Later developments, including those specific to mammals and then to humans, will undoubtedly transform subjective experience, giving rise to something more deserving of the name *consciousness*. For us, it will be hard to conceive of subjective experience without these later-evolving features, but that is no argument that subjective experience only arises with them.

---

[19] Below is a quote from Milner and Goodale about the work of David Ingle. Ingle rewired the nervous systems of some frogs. By crossing some wires, he was able to produce a frog that snapped at prey in the opposite direction from where it should have, given where the prey was. But this rewiring of part of the visual system did not affect all of the frog's visual behavior. The frogs behaved normally, moving in the right directions, when they were using vision to get round a barrier. They behaved as if part of the visual world was mirror-image reversed, and part was normal. Here is Milner and Goodale's comment:

> So what did these rewired frogs "see"? There is no sensible answer to this. The question only makes sense if you believe that the brain has a single visual representation of the outside world that governs all of an animal's behavior. Ingle's experiments reveal that this cannot possibly be true. Once you accept that there are separate visuomotor modules in the brain of the frog, the puzzle disappears.

"The puzzle disappears," they say. Perhaps *a* puzzle disappears, but another one is raised. What does it feel like to be a frog in this situation? Does it not feel like anything?