Materialism, Subjectivity, and Evolution

Peter Godfrey-Smith

University of Sydney
& CUNY Graduate Center

The Jack Smart Lecture 2017, Australian National University.

1. Australian materialism

I met Jack Smart in 1984, when I spent several months as a undergraduate "vacation scholar" at the ANU.[1] This was around the time I was deciding to become a philosopher. One of the things that had already struck me was what a generous and interesting collection of people philosophers were. Even against that background, Jack's graciousness stood out.

His work ranged over many topics: scientific realism, utilitarianism, the nature of time, materialism about the mind. His 1963 classic *Philosophy and Scientific Realism* opens with a sketch of how he was thinking about philosophy at the time:

> In recent years I have been moving away from a roughly neo-Wittgensteinian conception of philosophy towards a more metaphysical one, according to which philosophy is in a much more intimate relation with the sciences. Philosophy, it now seems to me, has to do not only with unravelling conceptual muddles but also with the tentative adumbration of a world view. (*PSR*, p. vii)

---

[1] The main text here is pretty close to the lecture as given on July 12, 2017. Footnotes follow up paths and add (incomplete) references and context. Two passages that were cut from the version delivered have been added back in, the final summary and another passage marked in the notes. The biological side of this paper is followed up in more detail in "Phylogenies of Consciousness," given at SPP 2017.

Here is how Smart approached things in the rest of that wonderful book. The aim is a coherent and comprehensive picture of the world, guided by science, but also aided by forays into the assessment of hypotheses that are not empirically testable, at least for now, but are meaningful and about which we can make reasonable choices. Those choices for Smart involved frequent appeals to Occamism, to parsimony. He also spent a good part of the book showing that apparent philosophical obstacles to materialism about the mind are not genuine problems, leaving the way clear for scientific arguments and appeals to parsimony to play their role. His positive view of the mental retained a significant Rylean/Wittgensteinian element, with features close to behaviorism, but he insisted that the mental also includes phenomena, like pain, that are inner states which we can report on.[2] These are identical with brain processes. The identity is contingent, similar in status to the identity of lightning with electrical discharge; there is no reason why the material nature of thought should be apparent to an ordinary person, and things might have been different. Smart had his eye on what was later called "multiple realizability" – he accepted that robots could have conscious mental states if their inner processes have the "same pattern" as ours. Smart saw cybernetics as an important way in to some of the harder behaviors to explain mechanistically. He urged that we not trust intuitions about what cannot be explained in a materialist framework; cybernetics made vitalists of an earlier generation, who insisted on various impossibilities, look "foolish." Smart was wary about the physical itself, though – the nature of the world described by physics. He saw the two-slit experiment as motivating real unease about a realistic interpretation of modern physical theory. He saw quantum mechanics as imperfect ("an inelegant mess"), not merely in philosophical interpretation but substantively, and expected a "new and simpler" theory to come along.[3]

---

[2] These parts of his view are not entirely behaviorist: "I would prefer to say, for example, that fear is the state of a person which is the causal condition of the characteristic behaviour pattern, rather than as with Ryle that it is the behavior pattern" (p. 89). This is the sort of view developed in more detail by Armstrong.

[3] References for the quotes in this paragraph: robots and "pattern", p. 106; how cybernetics made some vitalists appear "foolish," p. 115; quantum mechanics as "inelegant mess," p. 40.

Smart's work was central to the "Australian materialism" of the 1950s and 60s. The next round of work moved towards a more detailed treatment of mental states in terms of causal roles and neural realizers (Armstrong, Lewis). Then came what can be called "Australian dualism." Frank Jackson gave arguments in defense of epiphenomenalism, and Chalmers for dualism or panpsychism. (Jackson in the late 1990s moved back towards materialism, but his criticisms continue to have influence.) The sort of straightforward progress that Smart envisaged in working out the materialist picture seems more problematic now. As I see it, there have been at least four shifts in the framing of the issue that have contributed to this.

1. The first is the embedding of the problem in a modal framework, guided by the perceived victory of Kripke and his embrace of modality over Quine's austerity. Close attention to the necessity of some mind-body relations envisaged by materialism has made the road harder.
2. Related to this is a constraining role for considerations of what is knowable *a priori* (independently of experience).
3. The third involves changes in how people think about the experiential side of the mind. There was a move away from a family of attempts to avoid an "act-object" view of

---

Smart's views about biology (in chapter 3) are notorious among philosophers of biology and are indeed maddening in some ways. Smart wants to argue that biology has no "theories" in a strong sense, and no laws. A problem is that Smart compares sophisticated formulations of physical law with things like "albinotic [albino] mice always breed true," which is not at all analogous. He also claims at one point that physics uses statistics in an "intra-theoretic" way, as part of the theoretical apparatus, whereas biology uses statistics just to deal with noise and uncertainty in observation. On the next page (p. 59) he qualifies this claim and admits that biology has a few intra-theoretic uses, and physics also has the other kind. Here Smart seems to have some sense of the existence of the body of abstract biological theory that was growing steadily from the 1930s to the 1960s, work that undermines a lot of what Smart says in the chapter. Myself, I am sympathetic to the idea that laws are not important in biology (see my *Philosophy of Biology* chapter 2) but not to the view that biology lacks theory. In addition, part of what Smart wants to do in this chapter is oppose "emergence" as a way of handling the mind-body relation, and I am entirely on board with that.

Re-reading the book, I was struck by how often Paul Feyerabend contributes to Smart's thinking.

experience via the "adverbial" tradition and related ideas (including Smart's view in *PSR*).[4] I will return to this later.

4. There have also been changes in views of the physical – a kind of "emptying-out" that I will discuss below – that make it a lesser or more problematic explainer.

These all make the mind-matter gap loom larger, and I think they are wrong turns. This talk will include both critical and positive ideas. On the critical side, I'll offer arguments against some views, but in other cases will more briefly indicate the nature of the disagreement. The shifts I will say least about – and these I'll cover now – concern modality and the *a priori*.

It is true that early materialists did not appreciate all the modal contours of their positions. They said their mind-brain identities were contingent. As with U.T. Place's example of someone's table which *is* an old packing case, things might not have been that way. OK, but if his actual table right now is *this* old packing case, then they'd better have all the same properties, and it should not make sense to imagine one of them in the absence of the other. Kripke in *Naming and Necessity* focused on claims materialists needed which could *not* be contingent, and said that in these cases there is still an appearance of separability of mental and physical. If they really are separable, we have two things, not one.

Since then, the modal side has taken on a life of its own. It is now integral to statements of physicalism itself, and arguments offered on both sides.[5] I reject this, as I think modal facts are too soft to take this weight, and modal intuitions are too unreliable about the worldly part of what they are directed on. The bottom line is the need to deal

---

[4] Smart said in *PSR*: "The man who reports a yellowish-orange after-image does so in effect as follows: 'What is going on in me is like what is going on in me when my eyes are open, the lighting is normal, etc., etc., and there really is a yellowish-orange patch on the wall.'" (p. 94).

Lycan ("Representational Theories of Consciousness" in the *Stanford Encyclopedia*) sees Smart's view as either adverbial ("as Smart seems to have intended") or representationalist.

[5] Here is Daniel Stoljar ("Physicalism and Phenomenal Concepts," 2005): "Physicalism is the thesis that the phenomenal, or experiential, truths supervene with metaphysical necessity on the physical truths." A consequence of physicalism so stated is this: if P is a statement summarizing the physical truths of the world and P* is a statement summarizing the phenomenal truths, then the conditional (1)—I will call it *the psychophysical conditional*—is necessarily true:

(1) If P then P*."

with the admitted fact of an appearance of separability of mental and physical. The modal angle is useful insofar as it sharpens our thinking about that problem, unhelpful when it adds its own murky contributions. There is an *appearance of separability,* and you either *explain it as a probable illusion,* or not. Everything else is secondary, or less than secondary.

I think the appearance of separability has been diagnosed in a way that shows it to be unreliable, with resources provided by Nagel, Hill, and McLaughlin.[6] We can engage in two different kinds of *imagining* in the context of mind-body thought-experiments. There is perceptual imagining – imagining *seeing* something, like a body or brain – and sympathetic imagining – imagining *being* something, a thinking subject. The separability of these psychological acts gives rise to an apparent separability of mental and physical in "zombie" thought experiments, and the judgments we make on this basis are unreliable. (They might reflect truth, but are not reliable).

These debates are often now also embedded in consideration of what is knowable *a priori*: If some mental-physical identities are necessary but not obvious, perhaps that is because it they are special *a posteriori* necessities. Or perhaps they will turn out to be *a priori* if we think hard enough, and that would be a better road. This framing affects how people proceed.[7] I don't deny that thinking about what is deducible from what, in

---

[6] The crucial move was made by Nagel in footnote 11 of his 1974 "What is it like to be a bat?" paper. I discuss it further in "Evolving Across the Explanatory Gap."

The footnote-11 reply to intuitions about the separability of mind and body has been criticized by Doggett and Stoljar ("Does Nagel's Footnote Eleven Solve...?" 2010). Their main point is that the allegedly unreliable form of imaginative combination described by Nagel seems often to lead us to *reliable* modal intuitions and *justified* modal claims, so it can't be so bad. My response is that once we see the processes at work, we should indeed question all intuitions of dissociability that derive from this source.

[7] The 2005 Stoljar paper I quoted from above (in note 5) provides an illustration on this point, too. Here Stoljar describes what he takes to be the importance of "phenomenal concepts," concepts tied to experiences and recognitional capacities, in replying to anti-materialist arguments: "The importance of this part of the answer to the conceivability argument is that, if the psychophysical conditional [see note 7 above] is necessary and a posteriori, arguments from conceivability to possibility with respect to that conditional will fail: being necessary, the psychophysical conditional is not such that it is possibly false; being a posteriori, the conditional is such that one can in some sense conceive of its being false." Stoljar's paper is then largely focused on the question of whether the phenomenal concept

different ways, casts light on questions of separability. But I think it is a mistake to see considerations of *a prioricity* as any source of constraint. When we think about materialist views, we are in a situation of conceptual disruption. If materialism is true, two parts of language turn out to be targeted on one thing despite their vastly different patterns of use. In these situations, concepts get dislodged and shift around. Eventually they re-settle, and it is possible to come along and describe some things as *a priori*. But this is *post hoc* commentary, after the fact.[8] The accountant of the *a priori* files at dusk.

While recent threats to materialism can be put to rest, the absence of a good positive theory is certainly a problem. There is a recalcitrantly puzzling feel to the situation, something that cannot be dissolved as fast as someone like Dennett thinks. I will try to make progress on the problem. A few of the moves take us back to the sort of set-up seen in Smart, while in other areas (e.g., the role of Occamism) I disagree with him.[9] In some places I will just indicate the package I accept, and show how the pieces will have to fit. In other places I'll give more detail.

A few words about terminology: "consciousness" is often now used broadly for the feel of our mental lives, including the simplest kinds of subjective experience. I think the word "consciousness" has misleading features in that context, but it is a term of art without a stable meaning. I will use the term broadly when discussing those who use it that way. I'll also use "materialism" and "physicalism" interchangeably.

---

strategy suffices to show that the psychophysical conditional (of note 5) is not *a priori.* He says the strategy only shows that the conditional has a weaker property (being "not a priori synthesizable").

The force of an appeal to "phenomenal concepts" is being lost. If they have a role here, it is as part of a psychological diagnosis of why the appearance of separability of mental and physical is misleading. Intuitions of non-determination and separability are not reliable. There could be just *one thing* here, though this appears not to be the case.

[8] A view of this kind has been defended in detail by Richard Brown, and a discussion during a talk he gave at CUNY was helpful to me. Brown: "by the time a priori methodology will be of any use it will be too late." ("Deprioritizing the A Priori Arguments Against Physicalism," 2010).

[9] I don't think there's any reason to think that simpler theories are more likely to be true, in virtue of their simplicity. Occamist arguments in philosophy I see as generally empty. A simplicity preference does have a role in empirical inquiry, discussed in "Popper's Philosophy of Science: Looking Ahead." It is better in science to start simple and expect to get pushed by data towards more complex views, than trying to move in the other direction. That is very different from thinking that the truth is probably simple.

2. The Primacy of Subjectivity

I think the right way in to these issues is via the idea of *subjectivity* – this I see as both central to the mental and as explicable, a good target of explanation. The term "subjective experience" fits with this orientation. Subjective experience is the experience *of a subject*. This suggests an order of explanation, from subjects to what goes on in them. Nagel's formulation of the crucial feature – there being "something it's like" – suggests the same orientation. There's something it's like *for* a particular kind of entity, a subject. All this contests, for example, with orientations that put the *qualitative* at center stage, as something that may or may not be tethered to subjects.

A picture we might start out from recognizes two concepts: *subjectivity* and *agency*. These have different emphases – subjectivity is more a matter of the input side; agency involves the output side. But from a biological point of view and perhaps others, these are largely correlative and complementary capacities; sensing and action coevolved, and each gives the other its point.[10] Initially, I'll think of subjects as having a pair of features: (i) a point of view on the world, and (ii) an agenda. Subjects act in ways that reflect both.

This section will mostly be organized around critical points, but in a way that points towards a positive view. First, what is *in* experience? What are we trying to explain in materialist terms? The sketch above, and the something-it's-like formulation, suggest an orientation to the problem that puts the entire organism's activities into the frame. In recent years, though, a narrower conception has been dominant. Theories of consciousness or subjective experience have been largely concerned with the sensory or perceptual side of the mind.

Jesse Prinz is blunt: "All consciousness is perceptual."[11] Dretske hedges just a little: "the clearest and most compelling instance of it [state consciousness] is in the domain of sensory experience and belief."[12]

---

[10] These relationships are not completely straightforward in the context of early neural evolution: see Keijzer, van Duijn, and Lyon, "What Nervous Systems Do: Early Evolution, Input–Output, and the Skin Brain Thesis," 2013.

[11] Prinz, *The Conscious Brain,* 2012, p. 336. Dretske, "Conscious Experience," 1993.

I oppose this. Dretske says that "the clearest and most compelling" cases are sensory experience and belief. No, equally clear and compelling are emotions, willings, moods, and urges. In fact, they are quite a bit *more* clear than Dretske's case of belief. Along with a list like mine, some people think there is "cognitive phenomenology." Galen Strawson argues that there is the *experience of* understanding, or remembering something, where this is not a matter of disguised sensing.[13] I agree.

Dretske and Prinz do have a try with moods. For Dretske, emotions and moods might be perception of chemical, hormonal, and other internal states of the body.

> Why can't we, following Damasio (1994), conceive of emotions, feelings, and moods as perception of chemical, hormonal, visceral, and muscuoskeletal states of the body? ... This way of thinking about pains, itches, tickles, and other bodily sensations puts them in exactly the same category as the experiences we have when we are made perceptually aware of our environment.[14]

The move is to see these in terms of *extra objects of* experience – a perceptual model is applied throughout. Your bad mood is the means by which you are conscious *of* something about your state. This contrasts with the idea that mood is an *aspect of* experience. It feels different to be driving while in a good or a bad mood, and this colors other parts of experience.

If you are a non-philosopher, you are probably wondering: why try to push everything into a sensory model? Part of the appeal (not the only reason) is the possibility of a representational theory of the feel of experience.[15] Perhaps the feel of an experience is no more than what the experience is *telling* us. The "content" of an experience – what it says – is the same thing as its subjective feel.[16]

---

[12] The full quote: *"If* one chooses to talk about state consciousness (in addition to creature consciousness) at all, the clearest and most compelling instance of it is in the domain of sensory experience and belief." ("Conscious Experience," 1993.)

[13] See Strawson's "Cognitive Phenomenology: Real Life."

[14] "The Mind's Awareness of Itself," 1999.

[15] Prinz does not take this road, in a reductive form.

[16] Dretske again: "What gives these sensations their phenomenal character, the qualities we use, subjectively, to individuate them, are the properties these experiences are experiences of." On this view, what used to be called "qualia," inherent qualities of sensations, become properties *specified*

I think it is hard to see moods in this way, but people try, so I will offer another example that pushes away from both a representational view of experience and the sensory model more broadly. The example is energy level, especially fatigue. You are driving along and your energy level shifts towards fatigue. A heaviness and dullness sets in. This feels like something. It is part of subjective experience. Is it plausible that this is a *presentation* to you of some state (of your body), as opposed to a situation where your experience just *exhibits* fatigue – that foggy, heavy feeling? I think the shift in energy level is an *aspect of your living activity*, at that moment, one that is felt.[17] (If there was a sensory pathway here, it ought to be possible to fool it with a drug, and as far as I know this is not the case. On the other hand, hypnotism may have an effect of this sort – what would that suggest?)[18]

With energy level pushing us along, I add other things, further along the output side, as elements of subjective experience: *urges*, and *surges of resolve*.[19] All these things lead us away from a sensory model. They push towards seeing subjective experience as an aspect of living activity as a whole. I think the legacy of sense-data views has distorted things here, leading to a constant focus on the case of visual sensations (and the notorious green after-image). If by "consciousness" you mean there

---

representationally. See also Tye: "These observations suggest that qualia, conceived of as the immediately 'felt' qualities *of experiences* of which we are cognizant when we attend to them introspectively, do not really exist. The qualities of which we are aware are not qualities of experiences at all, but rather qualities that, if they are qualities of anything, are qualities of things in the world (as in the case of perceptual experiences) or of regions of our bodies (as in the case of bodily sensations)." (*Stanford Encyc*., "Qualia").

[17] I am trying to put pressure on two views at once here: the stronger view, seen in Dretske and others, that qualia are representational properties, and also the sensory model itself, the idea that even if there is more to subjective experience than what is representationally specified (even if there is mental "paint," as Block has it), that subjective experience in general is a sensory matter.

[18] I am grateful to comments by Leonard Katz at the 2017 SPP about these issues – especially the idea that if the experience of fatigue involved a pathway with a receptor, then it could be fooled. For the possible role of hypnosis, see Morgan et al. "Hypnotic Perturbation of Perceived Exertion: Ventilatory Consequences," 1976.

[19] I am leaving out here some interesting cases which involve the effects of one's own actions on how one processes sensory information. "Enactivism" over-extended some real phenomena here. Rather than pushing action *into* perception (as in Noë's *Action in Perception*), we can say that action and perception both contribute, in their own way, to subjective experience.

being "something it's like...," then there is no reason at all to have an exclusively sensory orientation to the phenomenon. Subjective experience is not all a matter of being *told* things.

Non-sensory aspects of experience like mood and energy level fit better with a family of views that were once popular for the sensory cases, with a version seen in Smart's *PSR*, and are now seen as discredited. These are *adverbial* approaches to experience (and their relatives): talk of features of experience is best seen as talk of the *manner* of our experienc*ing*. In the sensory cases, adverbialism was discredited in large part by some challenging arguments given by Frank Jackson in the 1970s.[20] This helped motivate representationalism, which became the new alternative to an "act-object" account, a new way to avoid problematic mental objects. I think something like adverbialism is worth another look. But my main point at this stage is the way that aspects of experience like moods and urges push towards a view of subjective experience that involves more of the organism than the sensory side.

Suppose we followed this path and developed a biological account of subjectivity. Could this be a complete explanation from a philosophical point of view? Some argue that insofar as this is a biological and hence physical account, it could not explain the *qualitative* or *phenomenal* (even if an adverbial approach is taken to qualia).

---

[20] See Jackson's "The Existence of Mental Objects," 1976. Anti-adverbialist arguments provided extra motivation to make a reductive form of representationalism work, as this seemed the main remaining way to avoid recognizing mysterious extra entities as bearers of *qualia*. Tye (in "The Adverbial Approach to Visual Experience, 1984) defended adverbialism against Jackson's arguments but later decided that his own analysis was "unattractive and highly complicated" (2009) and switched to representationalism. Jackson eventually adopted a representationalist view himself: "The most plausible view for physicalists is that sensory experience is putative information about certain highly relational and functional properties of goings on inside us. As it is often put nowadays, its very nature is representational" ("Postscript on Qualia," 1998). The act-object relation was replaced with a different kind of directedness. But representationalism, whatever else you think of it, is dependent on a sensory model of experience that does not cover all cases.

Some versions of this challenge derive from another one of the changes to the terrain I mentioned in the first section. These are new ideas about the role of physical, and explanation in physical terms. A number of developments have tended in the direction of "emptying out" the physical, leaving it as bare structure, either in its nature, as it functions in explanations, or as it functions in explanations of the mental. I'll discuss two of these here (not the only two). The first is an argument specific to these topics in the philosophy of mind.

The way I would like the general story to go is like this: subjective experience has its intrinsic features and these are due to the material properties of subjects. An objection is that in explaining the mental, only the "functional" or "organizational" properties of a system can matter, and those properties are badly suited to explaining consciousness. These claims are sometimes based on neural replacement scenarios. Those arguments date from at least the early 1980s and have several forms. They are worth revisiting.

Here is what Chalmers calls a "fading qualia" version of the argument.[21] Start with ordinary human agent, and assume a gradual cell-by-cell replacement of neurons by silicon-based control elements that exactly mimic the role of neurons. The agent continues to behave normally, and also retains the organizational features underlying their behavior. With this much retained, can qualia fade? If not, then only the functional/organizational features can matter. Chalmers also offers another scenario, which he calls "dancing qualia." Now we imagine a human agent for whom a backup control device is built out of computer hardware. The second control device, which has the same functional organization as the natural one, is connected by radio transmitters to the body's sensors and effectors, so that when activated it can control behavior in the usual way. Now imagine a rapid switching between "natural" or brain-based control of behavior and control by the backup system. If the character of experience depends on material make-up as well as the functional properties of a system, does their experience jump between different forms as the switching is done, or perhaps flip in and out of existence entirely, despite the agent's behavior continuing uninterrupted? Chalmers says no, and this suggests "experience is wholly determined by functional organization."

---

[21] See Chalmers, "Absent Qualia, Fading Qualia, Dancing Qualia," 1995. Chalmers notes that this argument builds on earlier thought-experiments by Zenon Pylyshyn and others.

In reply I begin with a general point. Chalmers and others write in a yes-or-no way about functional equivalence between neural and artificial systems. But this is a matter of degree. Functional properties are grain-specific, and similarity with respect to those properties is a graded matter. *Timing* provides a clean example that stands in for many others. Time is a continuous variable (or near enough). A slower system is functionally equivalent to a faster one in some respects but not in a finer-grained sense.[22] If you want to discount absolute time differences, consider the relative timing of the many processes occurring in parallel (which are even more important). Timing depends on physical details. Another grain-dependent property is replication of the "same" behavior on different occasions. (You raised your arm a bit differently this time – was the same behavior repeated from before?)

Suppose we grant that in neural replacement scenarios we mean: a functional duplicate fine enough for casual observers not to see a difference in behavior.
If not, there's no reason to even suspect that experience is unaffected over the shift. Are the scenarios then possible? In the fading qualia case I think clearly no. These arguments were invented when our picture of neural activity was different. Neurons were seen as switching devices. All a neuron does is fire and influence others in a switching network (and alter its sensitivity and output "weights" as they figure in the network). In fact neurons do much more; they participate in the diffusion of small molecules through the system, are affected by blood flow, have their activities modulated by all the events that affects gene regulation inside them. They are living cells. Those activities might be modeled – simulated – in a system built from scratch, but the gradual introduction of nonliving elements doing all these things into an intact living system without behavioral consequences is more a fantasy.[23]

In "dancing" qualia scenarios there is switching between two complete controllers. Here my objection is different. Part of it involves the distinction between *simulation* and *realization* of a functional profile. Do you build a system with an *actual* role for diffusion, global electrical rhythms due to synchronous firing, and so on, or do you write a program that simulates these things, a program that updates lots of values

---

[22] See Dennett's "Fast Thinking."

[23] Discussions with Rosa Cao influenced my thinking on these issues.

assigned to variables when nitric oxide molecules would be diffusing from place to place, and so on? An separate correction we need in this area (one not specific to questions of consciousness) is a narrow conception of realization, as opposed to simulation, of functional profiles. A lot of what are imagined to be alternative computational realizations of a functional profile are not that at all. Once we are *building* a system, I deny that we have any reason to believe that very fine-grained functional similarity is possible. Functional similarity has always been a grain-specific matter, and neural replacement scenarios get less and less plausible as grain gets finer.

Philosophers are used to thinking casually in terms of a distinction between functional profile and material make-up. The distinction itself is OK, but the make-up of a system has consequences for what is *done* at finer grains. Philosophers are also used to thinking that only coarse-grained, flow-chart-like, functional properties are relevant to explaining the mental. Then the material specificities find themselves banished from relevance. That is a mistake.[24]

Another family of ideas in this area has tried to see the *physical itself* as merely formal, as relationally constituted. I won't discuss the metaphysical versions of this view here: "ontic structuralism." I take those to be extreme positions about basic physics, and who knows what to make of the world if they are true? A related but less radical view holds that physics as a body of *theory* can only describe properties of certain kind – relational properties, or "structure and dynamics" as Chalmers puts it.[25] This makes them unsuitable for explaining the qualitative or phenomenal side of the mind. Chalmers argues that from facts about structure and dynamics, only more facts about structure and

---

[24] Paul Churchland in *Matter and Consciousness* (1984) argued for a response to the qualia problem that is a bit like this.

[25] Smart in *PSR* says things along these lines: "All the properties which science ascribes to physical objects seem to be purely relational" (p. 72). See also pp. 74-5, where he entertains what would now be called a "Russellian" view, but resists analyzing qualia in terms of unknown intrinsic physical properties because (via his adverbialism) "there are no such *qualia*."

This and the next three paragraphs about structuralist views have been added back in to the paper.

dynamics can be derived, and "truths about consciousness are not truths about structure and dynamics."[26]

With "structure and dynamics" broadly understood, I deny that last premise – I deny we could now know it to be true, anyway. The argument looks for a bridging capacity and finds it absent at level of physics itself. But that is not where you should expect to find it. To make sense of consciousness you should look for features like point of view and agenda as features of a biological system, in relation to its milieu. Much of what matters is due to intrinsic properties of the biological system – what it's made of, how it works, and how what it's made of gives rise to what it does. If you look *very* closely at *each* of those capacities, and at the role of the parts that enable them, you will see physical arrangements and processes. Those are constituted, let's assume, by "structure and dynamics" at a low level. This fact about what we find when we look at a very low level does not "empty out" what was going on at the higher and relevant levels.

If the claim is that there can never be a biological explanation of consciousness in terms of a hierarchically organized set of capacities, where those capacities are based ultimately in physical properties each of which is poorly suited *on its own* to explain consciousness, then the right response is that we don't know yet whether that will work in the end, and there is no general reason to believe that it won't.[27]

All these arguments want us to see consciousness as full of paint-like immediacy and intrinsicality, while the physical is bare relational bones. But the physical is not just bare bones.

---

[26] Chalmers: "First: physical descriptions of the world characterize the world in terms of structure and dynamics. Secondly: from truths about structure and dynamics, one can deduce only further truths about structure and dynamics. And thirdly: truths about consciousness are not truths about structure and dynamics" ("Consciousness and its Place in Nature," 2003).

[27] A feature of Smart's *PSR* is a scientifically based wariness about the physical, about what it amounts to, given quantum mechanics. We cannot see electrons, for example, as localized objects, bits of stuff in an ordinary sense. Smart also wonders, following Zimmerman, whether space and time are "concepts which apply on the macroscopic level only" (p. 44). I agree with his wariness. There is an atomistic, particle-based preference in a lot of philosophical discussions of physicalism (Lewis "Reduction of Mind," 1994, is an example). A future physical theory may be more field-based and holistic, and this would be quite disruptive of some philosophical work about mind and matter. Revisionary, critical versions of materialism shade into mild forms of neutral monism (and neutral monism on its other side shades into panpsychism).

I will look at another argument in this family based on limitations inherent in the physical or in physical theory, though this augment does not use an "emptying" move. Thomas Nagel, in a way I find encouraging, thinks that subjectivity and point of view are central to the mind-body problem. But he thinks a physicalist perspective can't explain these features. Nagel recently re-asserted this view in the letters page of the *New York Review of Books*.[28]

> The difficulty is that conscious experience has an essentially subjective character—what it is like for its subject, from the inside—that purely physical processes do not share.
> ...
> I believe [that solving the problem] will require that we attribute to neurons, and perhaps to still more basic physical things and processes, some properties that in the right combination are capable of constituting subjects of experience like ourselves.... These, if they are ever discovered, will not be physical properties, because physical properties, however sophisticated and complex, characterize only the order of the world extended in space and time, not how things appear from any particular point of view.

I think there is an error here that has to do with the relation between the materialist world picture and how the means we have of specifying that picture furnish explanations. Physical *concepts* describe the world as it is in itself, and not *for* any conscious subject – let's assume that. But this in hand, we can see that what we are describing, the set-up specified, gives rise to points of view when particular physical configurations come to exist. The physical world gives us point-of-view properties "for free."

In the last pages of *Naming and Necessity*, Saul Kripke wants to get to the heart of his objections to materialism from a different direction. He sets aside the modal apparatus and asks: if we imagine God creating the world, *what would he have to do* to establish various facts – the identity of heat and molecular motion, or the putative identity of mental and physical states. I think this is a useful device in the context of Nagel's argument. Once God makes the physical, and once physical systems of a certain kind evolve, you *have* individual points of view, ways things seem to particular individual organisms, given their makeup and circumstances. Even though the language of physical theory does not call on these things, their presence is a consequence.

---

[28] "Thomas Nagel Replies," *NYRB*, June 8, 2017.

3. The Biology of Subjectivity

In the last part of the lecture I'll look more closely at some of the biological story. The way I will approach things is by exploring origins and basic forms.

I said at the start that I'll think of subjects as having a pair of features: a point of view on the world, and an agenda. That starting point is guided by parts of everyday thinking, but it's a start, and we can ask: where do these features come from? Cellular life itself gives us the beginnings of this combination – precursor forms, almost compulsorily and certainly ubiquitously. All known bacteria and other prokaryotic organisms have some sensory capacities that coordinate what they do. In simple forms of life, the "doing" is mostly chemical, though most though (not all) known prokaryotes are mobile. Though some basic features in an account of the origin of subjectivity are laid down as early as this, today I will focus on animal life, and distinctive features of animals that bear on the evolution of subjectivity.

Animals are a branch of the total "tree of life," one kind of complex multicellular life. They have a different kind of organization from other multicellular organisms. There is extensive division of labor across their parts, but what is more important is the *kind* of division of labor and how it is achieved. Animals have a trio of properties.

1. Multicellular and spatially organized sensory structures (with the retina as an example), making it possible to track and respond to macro-level environmental structure.
2. The multicellular effector system characteristic of animals: muscle.
3. The multicellular control system linking them, the nervous system.[29]

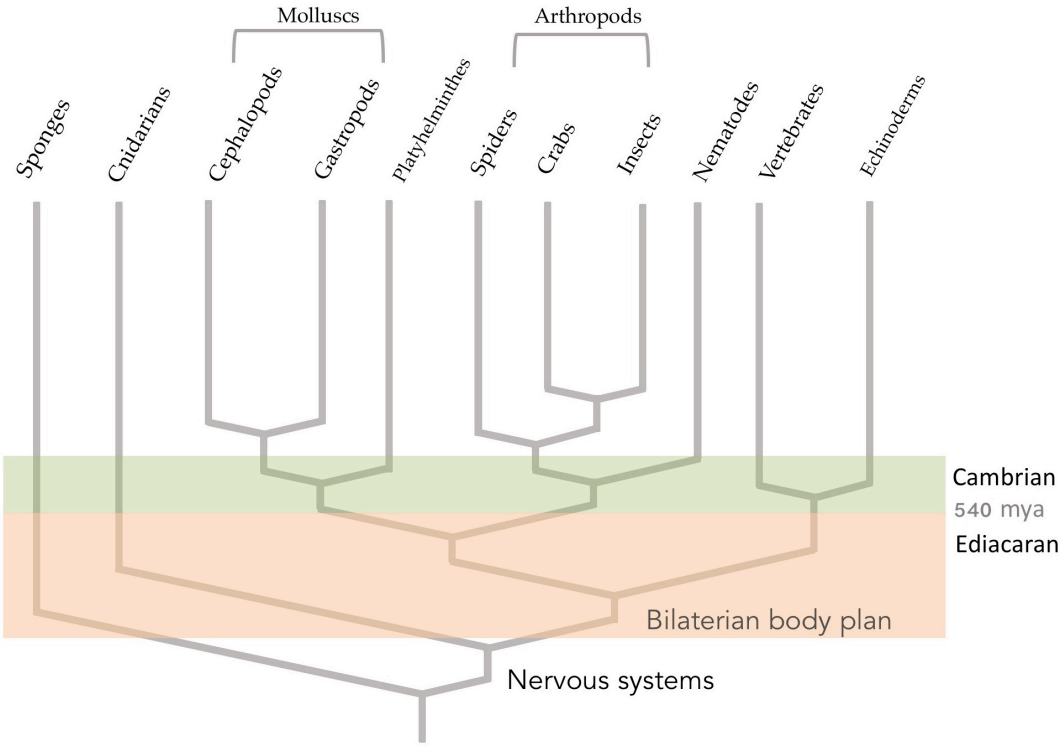
Nervous systems and muscle are found in nearly all animals, and are not found in significant form outside animals, though there are borderline cases (such as the Venus flytrap, with a quasi-nervous organization).[30]

---

[29] For treatments of these features and their significance see Arnellos and Moreno, "Multicellular Agency...," 2015; Jékely et al. "An Option Space for Early Neural Evolution," 2015; Feinberg and Mallatt, *The Ancient Origins of Consciousness*, 2016; Keijzer and Arnellos, "The Animal Sensorimotor Organization..." 2017.

[30] See Böhm et al, "The Venus Flytrap Dionaea muscipula Counts Prey-Induced Action Potentials to

When did these features arise? They arose so early that their history is very unclear, with no fossil record of the earliest forms. The fossil record of animal evolution begins in the *Ediacaran* Period, 635-450 million years ago. This was a behaviorally "quiet," time, it seems; the organisms that appear to be animals show no apparent bodily means for complex behavior and targeted action. But during this time, a lot of basic branchings between evolutionary lines occurred. Here is a chart of the animal part of the "tree of life."



Time runs up the page. The chart represents the order of branching of lineages, and hence facts about the genealogical relationships between different animals. We see a very early split between sponges and all the other animals shown, then a split between cnidarians (corals, jellyfish, anemones) and a large group of *bilaterian* animals – those with a left-right symmetry in their body plan. Nervous systems evolved before this split, as they are

---

Induce Sodium Uptake," 2016.

seen in similar form in both cnidarians and bilaterians.[31] After the bilaterian branch of the tree starts, it splits into two big groups, one that includes us, the deuterostomes, and one that includes many familiar invertebrate animals – arthropods such as insects, and molluscs such as clams and octopuses. This split probably occurred 600 million years ago or so.

A pivotal event in animal evolution was the transition between the Ediacaran and *Cambrian*, around 540 million years ago. This is the orange-to-green transition in the figure above. In the Cambrian we see the rapid appearance of sophisticated senses and various means for behavior and interaction – eyes, legs, claws. There is also evidence of predation. Animal lives in the Cambrian became more entangled, especially because of the targeted actions of one on another. Three groups gave to some species with what Trestman calls *complex active bodies* – bodies that have means for rapid motion, distance sensing, and manipulation of objects. Those three groups are arthropods, vertebrates, and one group of molluscs, the cephalopods.[32]

In these three groups – and outside them in more marginal forms – we see the evolution of more strongly subjectivity-relevant properties. These include: (i) *image-forming eyes*, eyes that enable genuine *viewing* of environmental objects. What Nilsson calls "Class IV" eyes – eyes that enable high resolution vision – appear to have independently evolved three or perhaps four times, with camera eyes in vertebrates and cephalopods and compound eyes in arthropods (perhaps four as spiders may have made a transition to high-resolution vision separately from insects). (ii) *Integration of sensory channels*. (iii) *Compensation for reafference*; compensation for the effects on one senses of one's own actions. This involves a kind of internal registration of a self-other distinction. (iv) *Instrumental learning* – learning by tracking the good and bad

---

[31] I omit ctenophores and placozoans, whose position is uncertain. The former have neurons and the latter do not. Branch lengths in the figure are arbitrary and the timing of many events is uncertain. See "Phylogenies of Consciousness" for more detail.

[32] Trestman's paper is "The Cambrian Explosion and the Origins of Embodied Cognition, 2015. A different angle on some of these facts via Glenn Northcutt: Of the 40-odd animal phyla, only four feature brains, the three listed here plus annelids, but these make up about 90% of animal *species* ("Evolution of Centralized Nervous Systems..." 2012). This is especially because of the success of arthropods.

consequences of one's actions – and other kinds of complex, non-routine handling of evaluation.

"Integration" sometimes figures as a rather hopeful gesture in discussions of the evolution of consciousness, but it has a genuine role here. The integration of sensory channels yields a more definite *point of view*. The animal acquires a perceptual locus, a point from which different events are registered. On the evaluative side, instrumental learning and its relatives bring an animal's agenda under its own control. There is a transition to *desires* from mere *drives*, in something like Sterelny's sense.[33]

An animal of this kind not only acquires a more definite point of view, but is made into a certain kind of *node* in a causal network. Such an animal can respond to Boolean combinations of events – doing X only if Y happened here *and* Z happened there... and can make changes to the world that are due to such combinations. Animals with these features are not only evolutionary products but special biological causes. An animal of this kind has a point of view, via inward-directed lines, via its scooping-up and integration of information, and is also a center of action, a locus from which directed effects emanate.

I'll now take a finer-grained perspective on some subjectivity-relevant properties.[34] I treated them as a package just above, but there are some interesting possible separations between them. I'll begin with a word about arthropods, the group that includes insects, spiders, and crustaceans. This is a big group, with a huge number of species. They set the co-evolution of animal behavior in motion in the Cambrian, very possibly. They were also the first animals to invade land, and did so repeatedly, around seven distinct times. Especially on land, they evolved complex sensorimotor skills. Many species can fly. But their lives are often short and dominated by routine. As a result, they are strong on the sensory side with respect to subjectivity-related properties – especially insects and spiders – and simpler on the evaluative, agenda-related side in many cases. Wasps and spiders are examples. Their long-lived relatives in the sea, crustaceans such as crabs, appear to be more complex on the evaluative side, at least in some respects. Crabs show wound-tending, for example, which is often (very reasonably) seen as evidence for

---

[33] See his "Situated Agency and the Descent of Desire," 2001.

[34] This material is covered in more detail in "Phylogenies of Consciousness."

felt pain. They exhibit complex trade-offs between different goals and disincentives, as studied by Robert Elwood. On land, bees are very sophisticated instrumental learners, and that is one kind of evaluative sophistication. Bees lack the obvious behavioral evidence for pain seen in crabs, but they show evidence suggestive of something like stress.[35] Insects are a puzzling case with respect to evaluative features.

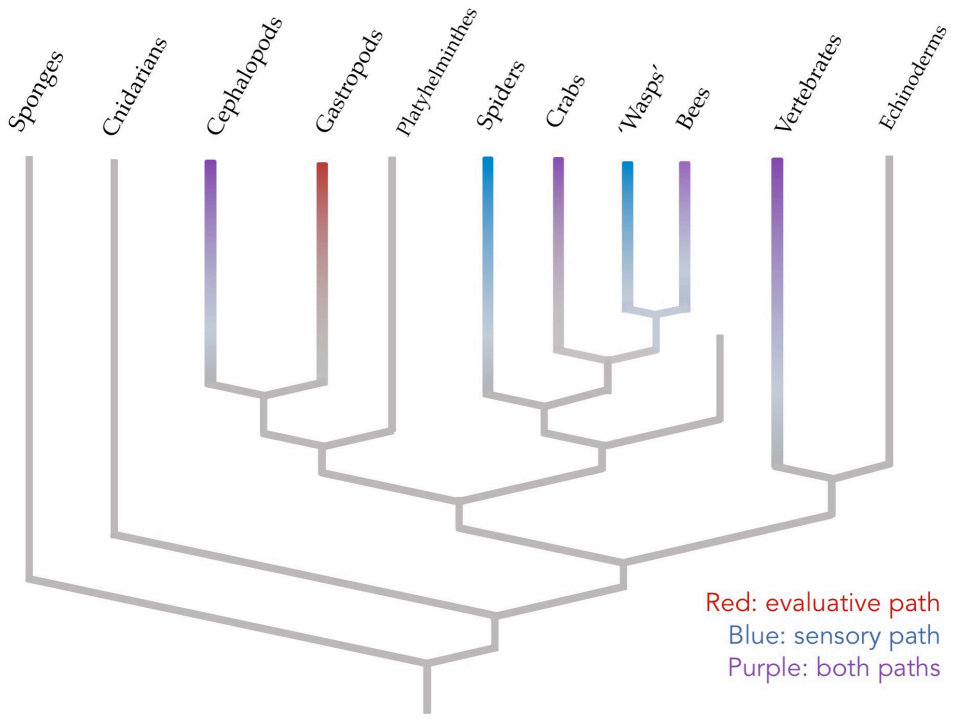
What about the flipside – more complexity on the evaluative than the sensory side? Gastropods (slugs and snails) are possible cases. They do not have high-resolution vision, but do show instrumental learning, at least of a basic kind. Perhaps they are examples, but this is much less clear than the arthropod combination, the other separation.

There are interesting questions about evolutionary paths to simple forms of subjective experience here. In the literature there are theories of "the evolution of consciousness" that are essentially sensory, and theories that are based on a notion of *feeling*, as the most plausible first form of experience. But perhaps we need not choose; perhaps there are two distinct forms of simple subjective experience, an evaluative and a sensory form. If so, can either be added first, such that the second may follow? Or is there more of an asymmetry? There is a lot to think about in relation to the evolution of subjectivity-related properties of different kinds – whether they cluster or dissociate, whether they evolve in a reliable order or not.

Below is another chart to summarize some of those ideas. It is intended very cautiously, as an illustration of some empirically motivated possibilities.

---

[35] See Elwood "Evidence for Pain in Decapod Crustaceans," 2012, and Tye's *Stressed Bees and Shell-Shocked Crabs*, 2016.



Purple indicates a combination of sensory and evaluative sophistication. In purple are vertebrates and cephalopods, the groups with the largest nervous systems, and also crabs and bees which are less clear cases. Blue indicates a predominantly sensory path, red a predominantly evaluative path.

I'll now start to work back towards the philosophical side. What I am trying to do is describe explicable biological features that, taken together, start to close the gap between mental and physical. These properties are not just kinds of "complexity," but features related to subjectivity itself. They are features that give a system a point of view in a strong sense, that give the system a sense of self-versus-other, and that bestow events with value, positive or negative. The idea is that with a filled-out story of this kind – one that I have only sketched here, but which I think we can glimpse – there is no extra question of whether there's something it's like to be an animal of these sort. Once we have subjects, subjectiv*ity* comes along. The *way it feels to be* a system of this kind comes along. These biologically established points of view are *occupied*, not merely parts of the world's layout.

Questions and challenges remain. Above I asked questions about clustering. Another, something more like a challenge, involves the role of gradual change. When people think about the evolution of consciousness they often think in terms of landmarks and crucial inventions, and they are guided by a sense of what seems a reasonable line-drawing (plants, bacteria, and jellyfish are out; cats and parrots are in). I share some common intuitions here, but these intuitions are not arguments. The way the biology is tending is a way that recognizes a lot of gradations with respect to the important properties.

On the sensory side, there is every possible intermediate and partial case between stages. An example that has been discussed less is learning. Philosophers often work with a simple distinction between organisms that can learn and inflexible or "tropistic" organisms. In the theories of people like Tye and Dretske, learning is seen as having a tight constitutive link to consciousness. Perhaps it does have this link, but the biological distribution of these features can be surprising. Almost all bilaterian organisms can be classically conditioned (the Pavlovian style of learning) and there has now been a report (without obvious problems as far as I can see) of classical conditioning in a plant (this is Monica Gagliano's work at UWA). For Dretske, instrumental learning rather than classical conditioning is the special kind – learning by reinforcement. This does seem rarer, but there are various partial cases and simulacra popping up, stretching into flatworms with tiny nervous systems, and so on.[36] As an important feature of sensing in animals, I make made much of reafference compensation – registering the effects of one's own motion, and the glimmer of a sense of self that this engenders. This may well be important, but a version of this is seen in tiny nematode worms with about 300 neurons.[37]

The situation may push towards recognizing a truly graded character to subjectivity-relevant properties. You might say: yes, but there may be some sort of *threshold* effect in how they give rise to subjective experience itself; there might be some sort of nonlinearity. That fits with our lights-switch-on intuitions, but I think it is quite problematic. It tends towards the bad kind of "emergence." It may be that the way things

---

[36] See Barron et al., "The Roles of Dopamine and Related Compounds in Reward-Seeking Behavior Across Animal Phyla," 2010.

[37] Crapse and Sommer, "Corollary Discharge Across the Animal Kingdom," 2008.

turn out will just force us to think differently about animals in particular, and life itself perhaps, and replace a presence/absence model of consciousness with a graded concept.

The result would not be a "panpsychist" view; the point of panpsychism is that we can't regard the mind as built out of anything *else*, so we have to find it lurking in the basic makeup of the world. The view I am describing (not today endorsing) is one with a different motivation, one that is positive rather than negative. The idea is that we *can* see what consciousness is built from – it is built out of subjectivity, a special sort of evolutionary product – but can see that it is built in a gradualist way and closely tied to basic features of life.

I'll finish by summarizing the package of ideas I've presented. I agree that there is an appearance of separability between the mental (experiential) and physical that requires a special kind of explanation. A deflationary psychological explanation of this appearance, and of modal intuitions that have driven recent anti-materialist arguments, is available. I reject the imposition of a sensory model on all subjective experience, and reject strong forms of representationalism. In describing experience itself, I hope for a revival of views in the family of adverbialism. These views are aided by the cases of mood, fatigue, and the like. I reject strong forms of multiple realizability, via the graded nature of "functional" similarity. There is no argument that *artifacts* are barred from having subjective experience, but what they have to do is different from what people suppose, and we need also a narrower notion of realization as opposed to simulation. I reject a family of other views that make the physical into a lesser explainer, such as Chalmers' "structure and dynamics" argument. Subjects are explicable products of animal evolution. In bringing the biological story to bear on the philosophical problems, we see interesting possibilities of dissociation between subjectivity-relevant properties, and may have to grapple with a graded conception of subjective experience and its origins.

* * *