Quantum Mechanics and the Consciousness Constraint

Philip Goff

Forthcoming in *Quantum Mechanics and Consciousness*, edited by Shan Gao

Quantum mechanics is one of the best predictive machines humankind has ever produced. Much of our modern technology, from computers to smart phones to GPS, is reliant on its predictive power. The trouble is nobody knows what quantum mechanics is telling us about reality. There are numerous proposals but no consensus on which is most probable. As things stand, the empirical data seems to underdetermine the theory.

In this kind of situation, philosophy has an important role to play, helping us to evaluate the evidential situation with respect to the various hypotheses. But it is generally assumed in this context that philosophy is not able to offer us *new data*, over and above the scientific data of observation and experiment. The usual expectation is that the philosopher of physics will contribute conceptual clarity and perhaps a cost-benefit analysis of the various interpretations of quantum mechanics in terms of theoretical virtues, such as simplicity, parsimony, non ad-hocness, etc.

In contrast to this standard assumption, I'm inclined to think that philosophy *does* have new data to offer, and that this data might have bearing on the ontology of quantum mechanics. What I have in mind is data pertaining to the reality of consciousness. Consciousness is not something that we know about through observation and experiment. If we were just going off the data of third person observation and experiment, we would have no need to postulate subjective experiences, as Daniel Dennett (2007) has argued very effectively. Nonetheless, contra Dennett, we do know that consciousness is real: we know that it's real in virtue of the immediate awareness each of us of our own feelings and experiences. Any theory of reality unable to account for the reality of consciousness is at best incomplete. In this sense, the reality of consciousness is a datum in its own right. I call the theoretical obligation to account for this datum 'the consciousness constraint.'

I believe that scientists and philosophers of the future will be baffled by the fact that their late twentieth century/early twenty first century ancestors did not make more use of the consciousness constraint. There is a certain phenomenon known to be real with something close to certainty, and yet the ontological implications of that phenomenon are completely ignored by most theoretical scientists, and even most metaphysicians. It is true that the problem of consciousness, broadly understood as the challenge of understanding 'how brains produce consciousness,' is now taken to be a serious scientific problem. However, this is generally assumed to be a problem that will go away with a bit more neuroscience. But the problem of consciousness is radically unlike any other scientific problem, not least because the fundamental datum that needs to be accounted for does not come from observation or experiment. Consciousness is something we know about *independently* of third-person empirical science; as such it is a valuable source of information to be *added* to the data of observation and experiment[1].

[1] It may turn out that the postulations we make to account for the data of observation and experiment can also account for the reality of consciousness, but this cannot be assumed from the start. Indeed, when you

The bearing of consciousness on quantum mechanics has been very little explored. Of course, a small number of heterodox thinkers have tried to make sense of the old idea that consciousness might have a role at the heart of quantum mechanics (see Chalmers and McQueen this volume). But this has never been articulated as part of a general approach of working out how the reality of consciousness constrains theory choice in this area. This paper will take a first step in rectifying this, by tentatively exploring the question of whether *wave function monism* – a popular interpretation of the ontology of quantum mechanics – is able to satisfy the consciousness constraint.

I – The consciousness constraint in more detail

I will assume that the reality of consciousness is at least as certain as anything else we know about the contingent world. But I think we can go further than this. Not only do we know that consciousness exists, but we know that it exists *as we ordinarily take it to be*. I don't mean by this that all of our reflective theoretical judgements about the nature of consciousness are correct. Rather, my claim is that we know that certain of our paradigmatic *concepts* of consciousness are an accurate representation of consciousness itself. The concepts I have in mind are those that have become known in the literature as 'phenomenal concepts.' This is just the philosophical jargon for the concepts we deploy when we think about our experience in terms of how it feels, or more broadly in terms of what it's like to have it. You're in pain, you attend to your pain, and think about it in terms of how it feels; in doing so, you form a phenomenal concept of your pain. I want to remain open to the possibility that some phenomenal concepts mis-represent; someone might, for example, mistakenly think they're having an orange experience, whereas when they attend in more detail they realise it was actually a red experience. It may be, as David Chalmers (2003) has defended, that there are a subset of beliefs involving phenomenal concepts which are in a certain sense infallible. However, it will be sufficient for my purposes to make the more modest claim that at least *some* phenomenal concepts – some instances of that general category – are satisfied.

Hence, we can define the consciousness constraint as follows:

> *The Consciousness Constraint* – Any adequate theory of reality must entail that at least some phenomenal concepts are satisfied (where a concept is satisfied just in case it corresponds to reality) (Goff 2017: Ch. 1).

This gives consciousness a unique status in metaphysics. With respect to most of our concepts – time, free will, colour – we are happy to have them moulded and revised by scientific developments. Perhaps our initial concept of time was such that time is absolute and the present has a privileged status. After reflecting on relativity, many philosophers have become persuaded that that pre-theoretical understanding of time cannot be quite accurate. This doesn't force us to a position of saying that time does not exist, but it does mean revising our concept of time to fit better with the picture of reality science has forced upon us. Similarly, many believe that our pre-theoretical understanding of free will as involving choices that lack prior causes is incompatible with our scientific knowledge of the brain, and hence must be replaced with a softer notion of freedom, one compatible with our choices being causally determined.

---

think about it, it would be strange if postulations which have been tailored to fulfil one theoretical task (accounting for observation and experiment) just happened to be suited to account for a distinct, and not obviously related, theoretical task (accounting for subjective experience).

In contrast, in the case of consciousness we have a class of concepts which we know accurately depict reality. Whilst some individual classifications of one's subjective experience may slightly misfire – I think I'm experiencing red when in fact I'm experiencing orange – the general understanding that each of us has of subjective experience is correct. When I entertain the proposition <there is something that it's like to be me>, I know that *that proposition*, and not some revised form of it, is true. In this sense, the knowledge that subjective experience is real is insulated from empirical refutation.

This is a powerful and much neglected tool for metaphysical enquiry. In the public mind, the only task that for a scientific theory to fulfil is to account for experimental data. But there's something we know about reality independently of experiments, and a theory of reality must be capable of accounting for that too. Could explicit recognition of this theoretical constraint help us make progress on the ontology of quantum mechanics? It is to this matter I now turn.

II – Can the Wave Function Ground Ordinary Objects?

Wave function monism is thought by many philosophers of physics to be the most simple and straightforward theory as to what quantum mechanics implies about reality (Albert 2013: 53). According to this view, fundamental reality consists of a complex-valued field in high-dimensional space: the wave function.[2] The space of the wave function is not the three-dimensional physical space of which we are familiar, but is rather 'configuration space,' so called because there is a correspondence between each location in configuration space and a configuration of particles in ordinary space. Such a configuration space is thought essential in order to capture the ways in which particles are constrained by facts about entanglement. Suppose, for example, we have two entangled particles in a superposition such that:

- Upon measuring the particles will be found either at location L1 or location L2, and are equally likely to be found at one as the other.
- If either is measured to be at L1, the other will be measured to be at L2, and vice versa

In the corresponding configuration space, there will be:

- A location corresponding to both particles being at L1, and a location corresponding to both particles being at L2.
- Locations corresponding to one particle being at L1 and one particle being at L2.

The wave function will have zero amplitude at the former locations and non-zero amplitude at the latter locations, corresponding to the fact that the particles could not be measured to be in the same locations but could be measured such that one is at L1 and the other at L2.

In the above description we are considering just two particles. For there to be facts about the wave function corresponding to all possible ways in which particles might be arranged in three-dimensional space, the configuration space in which it is housed must have 3 x N dimensions, where N is the total number of particles in the universe.

It might be tempting to think of configuration space as a way of *representing* what's going on in physical space. But for the wave function monist this gets things the wrong way around. The wave

[2] I found Ney 2013 an extremely helpful resource for understanding wave function monism.

function in configuration space is the fundamental reality; the particles in three-dimensional space – if they exist at all (more below) – are grounded in facts about configuration space.

The wave function is clearly a peculiar entity, and the idea that reality consists of such a thing seems in sharp contrast to the evidence of our senses. The claim of wave function monists, of course, is that we should look to science, not everyday experience, to find out what fundamental reality is really like. If wave function monism is the simplest interpretation of our best empirical theory, then wave function monism is what we ought to believe.

However, other philosophers of physics have questioned whether wave function monism could possibly be supported empirically, given its clash with experience. After all, what is empirical support other than what is known on the basis of experience? To be sure, scientific support involves experience in highly specific, controlled, repeatable experimental circumstances. Nonetheless, at the end of the day, we only know the results of experiments through using our senses, and whenever we use our senses, we seem to experience a world of objects in three-dimensional space. This has led Tim Maudlin (2007) to conclude that wave function monism is 'empirically incoherent': the theory cannot account for the very evidence upon which it supposedly relies.

The obvious response for the wave function monist is to dispute the charge that they cannot account for the three-dimensional world we perceive with our senses. Just because three-dimensional objects do not exist *at the fundamental level*, it does not follow that they do not exist *at all*. If the wave-function theorist can claim that three-dimensional objects are *grounded in* facts about the wave function, then they can thereby account for the existence of the evidence on which their theory is based. Unfortunately, this is easier said than done.

What is required in general for there to be a grounding relationship? Or to put it another way: What grounds a grounding relationship? In my book *Consciousness and Fundamental Reality*, I argue that the crucial feature of a grounding relationship is that the less fundamental entities are *nothing over and above* fundamental entities. I like party examples. Suppose Rod, Jane and Freddy are dancing and drinking one night. You've thereby got a party. But it's not as though the party is this wholly new thing that the revelling brings into being, in the way that the dances of witches may bring into being a demonic spirit. There's a clear sense in which the party is *nothing extra* to be the people having a good time.

The 'nothing over and above' relationship is prima facie paradoxical. How can X be *distinct* from Y, whilst nonetheless be *nothing more than* Y. I suggest that an account of grounding needs to resolve this paradox; I call this the 'free lunch constraint,' after David Armstrong's (1997: 12) famous term 'ontological free lunch' for an entity that is nothing over and above already postulated facts. Building on the work of others, I have argued that the 'nothing over and above' relationship should be accounted for in terms of an *analysis* of the grounded entities. For a given entity e, an analysis is a description of *what the reality of e consists in*, or, equivalently, *what is essentially required for e to be real.* In the case of a party, we can give the following analysis:

> *Party Analysis* – What is essentially required for there to be a party is for there to be people revelling.

For almost all the entities we talk about, it's practically impossible to give complete necessary and sufficient conditions for their existence; any proposal for the definition of 'revelling' will inevitably lead to a long-winded game of *counterexample leading to refining of the analysis, leading to counterexample, leading to refining of the analysis*, and so on ad infinitum. Despite the fact that we

can't give a precise definition of what is required for revelling, the very fact that we can engage in the game of 'spot the counterexample' entails that we have an implicit grasp of what is required.

Why is the party at Rod's house nothing over and above the fact that Rod, Jane and Freddy are dancing? Because all that is essentially required for there to be a party is for there to be people revelling, and the fact that Rod, Jane and Freddy are revelling logically entails that there are people revelling. The revellers are not strictly speaking identical with the party (they will go on existing when the party ends) but they provide all that is essentially required for the party to exist, and this gives us a clear sense in which the party is 'nothing over and above' the fact that Rod, Jane and Freddy are revelling; in other words, we have satisfied the free lunch constraint.

I call this way of accounting for grounding relationships 'grounding by analysis', defined as follows:

Fact X is grounded by analysis in fact Y iff:

- X is grounded in Y, and
- Y logically entails what is essentially required for the entities contained in X (including property and kind instances) to be part of reality.

Let's return to the putative grounding of three-dimensional objects in facts about the wave function. David Wallace (2012) has proposed that we should think of classical objects as *patterns* in the wave function. But what we want to know is why certain patterns in the wave function are sufficient to ground facts about three-dimensional objects. If we are thinking in terms of grounding by analysis, we need to make a case, via an analysis of, let's say, a table, that what is essentially required for that table to exist is logically entailed by facts about patterns in the wave function. In fact, I do argue in *Consciousness and Fundamental Reality* that we can analyse facts about macro-level objects in terms of patterns, specifically *patterns of penetration resistance among regions of space.* For any given three-dimensional object, it is impossible to precisely define the relevant pattern, but this is consistent with our having an implicit grasp of it (compare to the case of 'revelling' discussed above). We can say that the fact that there is a table in front of me consists in the fact that there is a table-ish region of space in front of me that resists penetration, for example, in the sense that if I put a cup down on it, it won't fall to the floor. This offers us a kind of rough and ready functionalist analysis of what it is for a table to exist.

If we were in a world of classical physics, it is plausible that arrangements of particles in three-dimensional space could entail that there is the right kind of pattern of penetration resistance for there to be a table. The trouble is that these patterns of penetration resistance in 3D space simply do not exist in the wave-function (my patterns are not the patterns Wallace points to in the wave function). For each particle in 3D space, there is a 3D subspace in the wave function corresponding to that particle. But there is nothing in the wave function corresponding to the *distances* between particles. The wave function is not a really a physical space in that sense; relative to 3D space it can be thought of as a *space of possibility*, with each location (in the wave function) corresponding to a complete configuration of particles (in 3D space).[3]

Of course, following Wallace, we can point to patterns in the wave function that do correspond to particles and objects in 3D space, and we can argue that those wave function patterns are physically salient due to the process of decoherence (Wallace 2010). But, as already noted, we need more than

[3] Of course, there is a sense in which configuration space is physical space: if wave function monism is true, it's the fundamental physical space. I just mean in the above that it's not a physical space in the sense that distances between locations in the wave function don't correspond to distances in 3D space.

a correspondence relation to secure a grounding relationship. To get a grounding relationship, we need the relevant patterns of the wave function to *logically entail* the relevant patterns of penetration resistance in 3D space. This is clearly not the case: given that 3D distances are not to be found in the wave function, a contradiction does not result from conjoining all the facts about the wave function with the denial that there are any patterns of penetration resistance in 3D space.[4]

There may be other ways of analysing what it is for a table to exist, or other ways of accounting for the grounding relationship. But then proponents of wave function monism are obliged to come up with the goods. David Albert (2013, 2015) has offered perhaps the most detailed attempt to do this. Albert effectively gives an analysis of what it is for a certain 3D system of particles to exist by specifying the Hamiltonian of the system. Without going too much into the details, we can think of a Hamiltonian as capturing how the system's behaviour across time is dependent on the masses and velocities (along three dimensions) of each of the particles involved in the system and the distances between them. Albert then argues that facts about the evolution of the wave function over time realise that functional profile.

However, as Alyssa Ney (forthcoming) has pointed out, the problem with Albert's account is that whilst the Hamiltonian of the system of particles in 3D space captures of terms of *velocities and distances in three-dimensional space*, the corresponding facts about the wave function involve *velocities and locations but not three-dimensional distances*. As a result, the relevant wave function facts do not meet the requirement specified by Albert's own account of the essential nature of a system of particles in 3D space. As in Wallace's account, all we really have are *correspondences* between particle facts and wave function facts without an adequate account of how the latter ground the former. This is a quite general challenge: it's hard to see how we could capture what is essentially required for there to be a system of particles in 3D space without mentioning *distances in three-dimensional space*, features whose existence does not logically follow from facts about the wave function.

Despite her scepticism about Albert's account, Ney is herself a wave function monist. Rather than offering a functionalist account of the grounding of 3D objects in the wave function, Ney offers a *priority monist* account. Priority monism is the ancient position pioneered in recent times by Jonathan Schaffer (2010), according to which there is one fundamental entity, usually taken to be the universe as a whole. Philosophers tend to assume that facts about big things are grounded in facts about little things, with all facts ultimately grounding in facts concerning arrangements of particles. Schaffer turns this on its head: facts about little things are grounded in facts about big things, with all facts ultimately grounded in facts about the universe.

Thus, Ney aspires to ground particles as *parts* of the wave function. One obvious difficulty with this ambition is that particles on the one hand and the wave function on the other do not share a common space. Particles exist in familiar three-dimensional space (or perhaps four-dimensional space-time); the wave function exists in a high-dimensional configuration space. Ney responds by suggesting that sharing a common space is not a general requirement for a part-whole relationship, on the grounds that some entities that stand in such a relationship do not exist in space at all: 'For example, 'egalité' is part of the national motto of France' (Ney forthcoming).

[4] Note that what we are talking about here is strict logical inconsistency. Likewise, in a particle-ontology, a complete description of the particle facts does not logically entail any facts about composite objects. The difference in this case, however, is that facts about particles (in 3D space) can logically entail the relevant patterns of penetration exists which are essentially required for there to be composite objects.

Whilst it is true that non-spatial entities can stand in a part-whole relationship, we have no examples of *spatial* entities that stand in a part-whole relationship despite the fact that they fail to share a common space. In the absence of such examples, Ney's claim that there is a part-whole relationship here seems quite obscure. She suggests analysing mereological relationships in terms of their formal features, such as being a partial ordering relation and principles such as Supplementation: If x is a proper part of y, then ∃z(z is a part of y and it's not the case that z overlaps x). However, one may be skeptical that one can give a complete analysis of the part-whole relationship in terms of these purely formal features, or one may (as seems plausible to me) want to stipulate that it's a necessary condition for X to be a part of Y that if X and Y are spatial entities, then X and Y share a common space. Without this stipulation, we leave open the possibility that entities in spatio-temporally distinct parallel universes could stand in part-whole relationships, which, to my ears, sounds incoherent.

I think there is a deeper challenge for Ney's view. Note that her account does not involve grounding by analysis; she does not account for the grounding relationship in question in terms of *what is essentially required* for 3D objects to exist. And in the absence of such an analysis, Ney has not demonstrated that she can account for a *nothing over and above* relationship between wave function facts and particles facts.

One way to see why this matters is to note that there are radical emergentist forms of priority monism, i.e. views according to which the parts of the universe depend on the universe but are nonetheless genuine additions in being relative to the universe. These views are just priority monist version of more familiar forms of radical emergence, such as were defended by the British emergentists of the 19th and early 20th centuries.[5] According to the British emergentists, chemical, biological and mental properties are fundamental entities in their own right that arise from physical properties without being in any sense reducible to them. If the British emergentists had been priority monists, they would have held that fundamental chemical, biological and mental properties arise from the physical nature of the universe without being in any sense reducible to facts about the universe.

It is standardly assumed that materialists must give some way of distinguishing their view from radical emergentism. For the materialist, all facts are nothing over and above the fundamental physical facts. Likewise, Ney is obliged to distinguish her view from a radically emergentist form of priority monism. Indeed, a view in which particles radically emerge from the wave function would *not* be a form of wave function monism but would rather be an interpretation of quantum mechanics according to which there is *both* a wave function and matter (with the latter arising from the former). To make sense of wave function monism, we need to show that all facts are *reducible* to facts about the wave function. The most obvious way to do this would be to offer an analysis of what is essentially required for systems of particles in 3D space to exist, and then to show that facts about the wave function meet that requirement. But, as Ney herself has argued, no extant account has managed to do this.

I hope to have shown that at the very least there are serious difficulties for a wave function monist wanting to account for 3D objects. John Hawthorne (2010) has suggested that this is a serious 'explanatory gap' analogous to that which seems to hold between the physical facts and the facts about consciousness. In one sense the wave function/3D objects explanatory gap looks even more challenging: in the case of the physical/experiential gap we at least know what a functionalist

[5] For examples of British emergentism, see Mill (1843), Broad (1925), and Alexander (1920). For a good discussion of British emergentism, see McLaughlin (1992).

account would look like, whether or not we find it plausible. On the other hand, in another sense, the gap is less serious, as we lack the certainty of the existence of 3D objects that we enjoy with respect to consciousness. If the wave function/3D objects gap can't be closed, it might be an option for the wave function monist to simply deny the existence of 3D objects, provided she can give some response to Maudlin's charge of empirical incoherence.

Whilst accepting that I have not here made anything like a conclusive case that that gap can't be closed, in what follows I would like to explore what follows for wave function monism if this gap cannot be closed.

III – The nature of evidence

In the last section we critiqued Alyssa Ney's attempt to close the wave function/3D objects explanatory gap and thereby to respond to Maudlin's charge that wave function monism is empirically incoherent. In an earlier paper (Ney 2015), presumably before she had formulated the account critiqued above, Ney suggests an alternative way in which a wave function monist might respond to Maudlin's challenge. After conceding, at least for the sake of discussion, that the wave function monist cannot ground objects in three dimensional space, Ney suggests that an option for the an wave function monist is simply to revise our understanding of what 'evidence' and 'confirmation' consist in, on the grounds that '[w]e should reason from what it is reasonable to believe our evidence is like given our best scientific theories, not from how our evidence pre-theoretically seems' (Ney 2015: 3120):

> On this view, confirmation is not going to involve causal relationships between spatially localized bits of a threedimensional world. Instead confirmation is going to involve the entire state of the world moving from one that is properly described (nonexhaustively) as "Theorists have constructed quantum theory T," to one that is properly described (nonexhaustively) as "Theorists have acquired evidence for theory T." For this to be viable, we should be able to connect such descriptions with elements of a wave function ontology (this is where decoherence can help), but this does not require solving the macro-object problem. Wave function realists should be open to the epistemic possibility that we may have to substantially revise how we think (in ontological terms) of things like "theorists" and processes like "constructing theories." (Ney 2015: 3122)

This seems to put the cart before the horse. Surely we need grounds for thinking the theory is true *before* we are entitled to reinterpret our evidence in terms of the theory. And to have grounds for thinking the theory is true we need *already* to have evidence that supports it. If it were permissible to re-interpret evidence in terms of that theory, then a fundamentalist Christian could interpret the pages of the bible as the unerring word of God, and thus find plenty of confirmation.

This raises the tricky question of what our basic evidential support for scientific theories consists in. Does our evidence for the standard model of particle physics consist of the experimental results we perceive with our senses? These sensory experiences (of the results of experiments) seem to represent objects in three-dimensional space and time, which would seem to imply that what they give us grounds for believing (if anything) is the existence of objects in three-dimensional space and time. And it would seem to follow that our scientific theories cannot be inconsistent with the existence of such objects, on pain of undermining the evidential support we have for those theories.

However, I agree with Ney's concern that this seems to put undue limitations on the potential of science to revise our pre-theoretical views about the underlying structure of the physical world. As

she puts it 'we should be highly skeptical that we should be able to legislate as an a priori matter what science can and cannot reveal about the spatial structure of our world' (Ney 2015: 3123). Indeed, as Ney points out, for most of the history of philosophy the dominant view about the nature of reality was idealism: the view that reality is fundamentally mental. The idealist need not – contrary to what Ney claims in this paper – think that the three-dimensional space and its contents are illusory: Berkeley believed in physical objects but thought they were constructed of ideas. However, some idealists do contend that the physical world is an illusion, and it seems wrong to think that we can rule out this view, as Samuel Johnson famously tried to do, merely by kicking a stone.

One obvious way around this is to define our evidence in terms of consciousness, and there are two ways in which we might do this. We could take our evidence simply to be our conscious states, collectively, and judge a theory in terms of its how well it explains the fact that we have the experiences we do. Obviously, this approach leaves open the possibility of idealism (which is not to say that idealism is plausible). Alternately, we can define the content of our experiences in terms of causal connections between experience and the external word. Thus, the pointer on a dial can be defined as 'whatever causes our pointer experiences.'[6] On the former approach our evidence are the conscious experiences themselves; on the latter approach our evidence consists of entities that are (or at least might be) mind-independent but which are latched onto in virtue of their relationships to experience. In terms of their evidential import, there might end up being not too much difference between the two approaches. The crucial point for our purposes here is that either approach allows a great deal more flexibility in what science can end up telling us about the physical world. We are not tied, just in virtue of our use of empirical evidence, to a universe of 3D objects in space and time. That which accounts for our conscious experience, or that which our conscious experiences latch onto, may turn out to be very different to how we ordinarily conceive of it.

However, this way of conceiving of our evidence does not leave our metaphysics entirely unconstrained. For it commits us to the reality of human conscious states, either explicitly (if our evidence just is our conscious states) or implicitly (where our connection to our evidence is fixed in virtue of our conscious states). Hence, even if Ney is right that the wave function monist is not obliged to account for a 3D world, they do end up being obliged to account for consciousness.

Can we construe evidence in a way that avoids this commitment to consciousness? Perhaps one might try to adopt a purely externalist theory of perceptual content. Thus, we might say that in perception we latch on to *something*, but that reference is fixed wholly by facts outside of the content of the perception. In this way, appeal to perceptual evidence would not presuppose a commitment to the 3D world, and nor would it presuppose a commitment to consciousness, as neither (on this view) enter into the basic content of perceptual experience.[7]

[6] Putting this terms of Chalmers' (2004) two-dimensional semantic framework, we would say that the concept's primary intention picks out whatever is causally related in the right way to consciousness. The primary intention corresponds to our epistemic situation, and hence in each epistemically possible world in which the concept refers, consciousness exists.

[7] On Chalmers' two-dimensional framework, such concepts aren't possible as they would lack primary intentions. However, many philosophers think such concepts are possible. See Goff 2017: 4.3, and Goff and Papineau 2014 for more discussion. Also, one might assume that any appeal to consciousness necessarily commits to consciousness, given that perceptual states are conscious states. However, I am imagining here a view according to which perceptual evidence is nothing more than receipt of information from the external world.

However, this approach would not leave the wave function monist unconstrained; it would constrain her to commit to *reference*.[8] And if we can't ground 3D objects in the wave function, it's hard to see how we could ground perceptual reference in the wave function. In general, naturalistic theories of reference ground reference in causal relationships between perceiver and environment.[9] If the wave function realist doesn't have either perceiver or environment, then they're going to have to give an account of reference purely in terms of the wave function. At the very least, nobody (as far as I know) has shown how this can be done.

In any case, even if there is way to make sense of evidence without a commitment to consciousness (in a way that avoids being tied to a 3D world), the fact remains that we know that consciousness is real. Even if it's possible to construe empirical data in a way that makes no reference to subjective experience, nonetheless we have at least as much justification for believing in subjective experiences as we do in trusting the empirical data. Regardless of its connection to evidence, the wave function monist is obliged to account for consciousness simply because we know that consciousness exists.

IV – Can the Wave Function Monist Account for Consciousness?

As we have seen, the crucial question for wave function monist is whether she can account for the reality of consciousness. If she can, then she can construe our evidence in terms of our conscious experiences. If she can't, then wave function monism cannot be true in any case, as there's something real that the theory can't account for. Is accounting for consciousness harder for wave function monism than for other theories?

The answer to this question may depend on whether we aspire to give a materialist or a non-materialist account of consciousness. Let us begin by focusing on materialist accounts of consciousness. There are broadly speaking two kinds of materialist theory of consciousness, which David Chalmers (2002) labelled 'type-A' and 'type-B'. Type A materialists aspire give an a priori functionalist reduction of consciousness, that is to say, to analyse conscious states in terms of their causal role, a causal role which is realised by physical states.[10] Thus, for example, the type-A physicalist might hold that, by definition of the word 'pain', all that is essentially required for someone to be a pain, is for there to be an inner state that negotiates between bodily damage and avoidance behaviour in the distinctive way that pain does. Could facts about the wave function realise these functional facts? There is no bodily damage or avoidance behaviour in the wave function, and so the best option for the wave function monist would be to ground the existence of particles in 3D space and then to suggest that systems of particles realise the functional states constitutive of mentality. But as we saw above that the wave function monist struggles to give a functionalist account of systems of particles in 3D space, as such an analysis at the very least involves reference to 3D distance relations between particles, 3D distance relations which are not

[8] I'm not claiming here that a commitment to reference is implied by the content of a perceptual state (e.g. by its primary intention). My point rather is that the theory that is being signed up to here, i.e. 'in perception we latch on to *something*, but that reference is fixed wholly by facts outside of the content of the perception', involves a commitment to reference, and hence that someone signing up to that theory is obliged to account for that commitment.

[9] There are related concerns in Braddon-Mitchell and Miller (2019), which discusses whether reference could be grounded in a world in which time is not fundamental. There are also phenomenal intentionality theories (Kriegel 2013, Mendelovici 2018) that ground facts about reference in facts about consciousness. I will not discuss these here, as the next section concerns whether the wave function monist can account for consciousness.

[10] For some examples of type-A physicalism, Putnam 1967; Lewis 1966; Armstrong 1968.

present in the wave function. Without particles, it's hard to see how the wave function monist can give a functionalist reduction of consciousness.

Type-B materialists, in contrast, do not rely on the *meaning* of mental terms in order to account for the grounding of consciousness in the physical.[11] This doesn't entail rejecting the grounding by analysis model outlined above; it just requires holding that an analysis of consciousness is to be given *a posteriori* rather than a priori. In other words, the type-B analysis holds that empirical work, rather than introspective reflection, reveals the essential nature of consciousness. Typically, the type-B materialist proposes that our phenomenal concepts refer to physical states, but that it is not a priori knowable that this is the case (because the reference is fixed outside of what we have a priori access to). In the toy example that has become standard in the literature, 'pain' refers to c-fibres firing in the brain, but one cannot know just through reflection on pain experientially conceived that this is so.

The wave function monist adopting this approach will hold that our mental terms latch on to features of the wave function (although, of course, this cannot be known a priori). The difficulty here is that some account must be given of how precisely our mental terms latch on to features of the wave function. And as discussed in the last section, accounting for reference without three-dimensional objects in space and time is not going to be easy.

In summary, the challenges faced by the wave function monist in accounting for 3D objects translate into challenges for the wave function monist desiring to give a materialist account of consciousness. What about non-materialist accounts of consciousness? Naturalistic dualists postulate basic psychophysical laws which ensure that certain facts about consciousness arise from certain physical facts. Given that there is a correspondence between facts about the wave function and the putative facts concerning familiar physical entities in 3D space, I can see no reason in principle that psycho-physical laws formulated in terms of the latter could not be 'translated' into laws concerning the former. The result would be psycho-physical laws according to which facts about consciousness arise from facts about the wave function. In the final chapter of *The Conscious Mind* (1997), David Chalmers defends an Everettian version of this view; on the Everettian view wave functions do not collapse. However, one might also combine it with the view, explored in this volume by Chalmers and Kelvin McQueen, that consciousness collapses the wave function.

However, one big challenge for this combination of wave function monism and mind-body dualism is that it's hard to see how it can give an adequate account of perception, at least if we're still working with the assumption that wave function monism cannot ground ordinary objects in 3D space. The perceptual states of a person provide information about facts about the external physical world. If wave function monism is true (and there are particles/ordinary objects in 3D space), those facts about the external world are not quite what they seem to be; they are facts about the wave function rather than facts about objects in 3D space. Nonetheless, we surely still want to say that my perception teaches me about them; as discussed above, the empirical coherence of wave function monism would seem to depend on this being so. However, in the absence of 3D objects, we can't explain this in the familiar way, in terms of objects in the environment causally impacting on my consciousness via my brain. Again, we seem to be left with no coherent way to make sense of wave function monism as a theory that can be empirically confirmed.

[11] For examples of type-B physicalism, see Loar 1990, Balog 1999, Papineau 2002, Diaz-Leon 2010, Howell 2013.

One currently popular alternative to both physicalism and dualism is Russellian monism.[12] Russellian monism starts from the thesis that physical science fails to give us the complete story of the nature of physical reality. This is put in various slightly different ways, but the basic idea is that physical science tells is confined to telling us about the *structure* of physical reality – those features that can be captured in a mathematic-causal vocabulary – leaving us in the dark about the underlying intrinsic natures of the entities that realise that structure. This is sometimes called 'the problem of intrinsic natures.' The Russellian monist proposes locating consciousness, or perhaps proto-consciousness, in the intrinsic nature of physical entities.

The attraction of this approach is that it promises to avoid the difficulties both of the materialist and of the dualist. The challenge for materialism is that there are good arguments to the conclusion that facts about subjective qualities cannot be fully accounted for in terms of kind of structural facts conveyed by physical science. The challenge for dualism starts from the putative evidence that physical reality forms a causally closed system, which seems to leave non-physical consciousness with no role to play in the generation of human or animal behaviour. By taking consciousness (or at least proto-consciousness) as basic, the Russellian monism hopes to avoid the problems with physicalism; by placing proto-consciousness within the intrinsic nature of the physical, the Russellian monism hopes to avoid the problems with dualism.

A wave function monist taking this approach would locate (proto)consciousness in the intrinsic nature of the wave function. Are there problems with this form of Russellian monism, not shared with other forms? According to Russellian monism, physical science tells us nothing of the nature of physical reality – whether it consists of a wave function or particles in 3D space – and hence difficulties won't arise in connection with the intrinsic nature of the wave function. Or at least, given that we are equally ignorant of both the intrinsic nature of the wave function and the intrinsic nature of particles in 3D space, we can't possibly have grounds for thinking one is more inhospitable to mind than the other. If it is harder to squeeze consciousness into the wave function than it is to squeeze it into other putative forms of physicality, this must be down to the only thing physical science does reveal to us about the wave function: its structure.

The wave function certainly does have a bizarre structure. To make things clear, we can take a simple scenario that we don't in fact find in the real world: one in which the amplitude of the wave function is entirely located in one location. This corresponds to a situation in which all particles have completely determinate locations. Now imagine a panpsychist Russellian monist wanting to ground facts about conscious particles in terms of this wave function fact. On such a view:

F1: a large number of conscious particles with determinate locations

Is grounded in:

F2: a high-dimensional space with a property (i.e. amplitude) instantiated at one location.

I can't see any way of definitively ruling out such a grounding relationship. Perhaps if we knew the intrinsic nature of configuration space and the intrinsic nature of amplitude, it would just be obvious that F2 grounds F1. Nonetheless, the very different structure of the two facts makes it, to say the least, puzzling how they could stand in a grounding relationship. Certainly we don't have the more familiar panpsychist picture of little conscious things combining to produce bigger conscious things, or even the cosmopsychist picture of a conscious universe fragmenting into conscious parts (at least

[12] For a good collection of essays on Russellian monism, see Alter & Nagasawa 2015. See also Goff 2017.

not parts in the same space).[13] It is true that there are also puzzles about structural mismatch in these more familiar forms of Russellian monism (Chalmers 2016, Goff 2017: Ch. 8) – the structure of consciousness seems different to the structure of the physical brain – but the structural mismatch faced by the Russellian wave function monist seem of a wholly different order.

Notice that F1 involves reference to locations in 3D space, which entails the existence of distances in 3D space. Does the Russellian wave function monist face the same difficulties grounding 3D distances that we discussed in section II? The advantage for the Russellian monist is that she is committed to locations and/or distances having an *intrinsic nature*, which opens up the possibility that an analysis of the intrinsic nature of locations/distances may yield a condition (specifying what is essentially required for there to be locations/distances) which can be satisfied by facts about the intrinsic nature of the wave function. Of course, we have no grasp of a possible intrinsic nature of location/distance that could do this, but Russellian monists tend to be comfortable with admitting ignorance about the intrinsic nature of physical reality on the grounds that we wouldn't expect as evolved creatures to have ways of accessing the underlying intrinsic nature of matter.

However, we do have access to the intrinsic nature of consciousness. And therefore, if the Russellian monist wants to ground F1 in F2, then she must offer an analysis of consciousness that 'opens it up' to this kind of reduction. To make clear what I have in mind, return to the analysis of a party. For there to be a party is for there to be people revelling; this analysis opens 'there is a party' up to being reduced to facts about revellers. As I have argued at length elsewhere (Goff 2017: Ch. 9; Goff 2019), I don't think that something analogous can be done with respect to consciousness. The only analysis we can give to the proposition <there are N subjects of experience> is the trivial one that what is essentially required for this proposition to be true is for there to be N subjects of experience. And it's hard to see how F2 could satisfy this condition. The only possibility is to identify each dimension with one of the N subjects; but, in this case, the fact that the amplitude is focused in one location would seem to have no role in the grounding of F2.[14]

I argue in chapter 9 of *Consciousness and Fundamental Reality* that given how 'thin' the analysis of consciousness is, the only hope for grounding consciousness is to do so via what I call 'grounding by sumsumption.' In cases of grounding by subsumption, a less fundamental entity is irreducibly subsumed in a more fundamental entity. To give an analogy, some Christians think about the trinity in these terms: the three persons are irreducible conscious subjects but are nonetheless subsumed in a single substance. I suggest the Russellian monist may hold that conscious minds are grounded by subsumption in the universe: although irreducible, they are subsumed in the more expansive reality of the universe. For this model to be compatible with wave function monism, human and animal conscious minds would have to be irreducibly present in the wave function (compare: the three persons are irreducibly present in the Godhead, on this model of the trinity). This would be the case with respect to F2 only if each dimension was identical with a conscious subject; but, as already stated, this would make the fact that the amplitude is focused in one location redundant in the grounding of F1.

In summary, there doesn't seem to me to be a way for the wave function monist to account for the facts of consciousness, at least on the assumption that they can't account for the facts about 3D

[13] Of course if we could ground conscious particles, we could then try a more familiar panpsychist grounding story to get to macro-level consciousness.

[14] Perhaps there could be a causal relationship between the fact that the amplitude is in one location and the fact that each of the dimensions is a subject, but this wouldn't be a grounding relationship.

objects. And if wave function monism cannot account for consciousness, then wave function monism cannot be true.

Could these challenges be avoided if one gives up wave function monism and adopts an interpretation of quantum mechanics which postulates matter or particles in addition to the wave function, such as the Bohmian view? This would be a solution for the dualist. Once we have matter in the picture, the dualist can proceed to give the familiar account of visual perception in terms of the interactions between material entities and our conscious minds. One concern for materialists or Russellian monists is that these interpretations can be understood as rendering matter epiphenomenal: something that is guided by the wave function rather than having any causal efficacy in its own right. If consciousness is grounded in matter, then consciousness too would lack causal power. The best option for the materialist or Russellian monist may be to adopt realism about matter accompanied by the thesis that the wave function has the status of a law, specifying the causal powers of matter, rather than being a physical entity in its own right.[15]

V – Much work to be done!

I have tentatively argued that wave function monism is false, on the grounds that it cannot account for the reality of consciousness. I don't take these arguments to be conclusive; rather they present challenges that future work may overcome. The crucial point is that there is much work to be done here properly assessing whether or not wave function monism, and other interpretations of the ontology of quantum mechanics, can account for the reality of consciousness. There is a desperate need for a new generation of philosophers who know the physics and who are serious about consciousness. How to interpret quantum mechanics has been one of the biggest scientific/philosophical challenges for nearly a hundred years. At the same time, the potential for our knowledge of consciousness to assist in metaphysical progress has been woefully neglected for at least as long. It's just possible that proper attention to consciousness – the one phenomenon we know with certainty to exist – might allow us to make progress on this issue.

References

Allori, Valia, Sheldon Goldstein, Roderich Tumulka, & Nino Zanghi (2007) 'On the common structure of Bohmian mechanics and the Ghirardi-Rimini-Weber theory,' *British Journal for the Philosophy of Science* 59: 353–89.

Albert, D. Z. (2013) 'Wave function realism,' in Ney & Albert (2013).

Albert, D. Z. (2015) *After Physics*, Harvard University Press.

Alexander, Samuel (1920) *Space, Time and Deity*, Macmillan.

Alter, Torin & Nagasawa, Yujin (2015) *Consciousness in the Physical World*, Oxford University Press.

Armstrong, David (1968) *A Materialist Theory of Mind*, Routledge and Kegan Paul.

Armstrong, D. M. (1997) *A World of States of Affairs*, Cambridge University Press.

[15] For an example of this kind of view, see Allori et al 2007.

Balog, Katalin (1999) 'Conceivability, possibility, and the mind-body problem,' *Philosophical Review* 109: 4, 497-528.

Braddon-Mitchell, David & Miller, Kristie (2019) 'Quantum gravity, timelessness, and the contents of thought,' *Philosophical Studies* 176: 7, 1807-1829.

Broad, C. D. (1925) *The Mind and its Place in Nature*, Routledge and Kegan Paul.

Chalmers, David J. (1997) *Consciousness and its Place in Nature*, Oxford University Press.

Chalmers, David J. (2002) 'Consciousness and its place in nature,' in Chalmers (Eds.) *Philosophy of Mind: Classical and Contemporary Readings*, Oxford University Press.

Chalmers, David J. (2003) 'The Content and Epistemology of Phenomenal Belief,' in Q. Smith & A. Jokic (Eds.), *Consciousness: New Philosophical Perspectives*, Oxford University Press. pp. 220-72.

Chalmers, David J. (2004) 'Epistemic two-dimensional semantics,' *Philosophical Studies*, 118: 153-226.

Chalmers, David J. (2016) 'The combination problem for panpsychism,' in G. Brüntrup & L. Jaskolla (Eds.) *Panpsychism*, Oxford University Press.

Chalmers, David J. & McQueen, Kelvin J (this volume) 'Consciousness and the Collapse of the Wave Function.'

Dennett, Daniel C. (2007) 'Heterophenomenology reconsidered,' *Phenomenology and the Cognitive Sciences* 6: 1-2, 247-270.

Diaz-Leon, Esa (2010) 'Can phenomenal concepts explain the explanatory gap?' *Mind* 119: 476, 933-51.

Goff, Philip (2012) *Spinoza on Monism*, Palgrave Macmillan.

Goff, Philip (2017) *Consciousness and Fundamental Reality*, Oxford University Press.

Goff, Philip (2019) 'Grounding, analysis, and Russellian monism,' in S. Coleman (Ed.) *The Knowledge Argument*, Cambridge University Press.

Goff, Philip & Papineau, David (2014) 'What's wrong with strong necessities?' *Philosophical Studies*, 167: 3, 749-62.

Hawthorne, John (2010) 'A metaphysician looks at the Everett interpretation,' in Saunders et al 2010.

Howell, Robert (2013) *Consciousness and the Limits of Objectivity: The Case for Subjective Physicalism*, Oxford University Press.

Kriegel, Uriah (Ed.) (2013) *Phenomenal Intentionality*, Oxford University Press.

Loar, Brian (1990/97) 'Phenomenal states,' originally published in J. Tomberlin (Ed.) *Philosophical Perspectives 4: Action Theory and Philosophy of Mind*, Ridgeview; reprinted in substantially revised form in N. Block, O Flanagan, & Guüzeldere (Eds.) 1997 *The Nature of Consciousness: Philosophical Debates*, MIT Press.

Lewis, David (1970) 'An argument for the identity theory,' *Journal of Philosophy*, 63: 17-25.

Maudlin, Tim W. E. (2007) 'Completeness, supervenience and ontology,' *Journal of Physics A: Mathematical and Theoretical,* 40: 12.

Mendelovici, Angela (2018) *The Phenomenal Basis of Intentionality*, Oxford University Press.

Mill, John Stuart (1843) *System of Logic*, Longmans, Green, Reader and Dyer.

McLaughlin, Brian (1992) 'The rise and fall of British emergentism,' in A. Beckerman (Ed.) *Emergence or Reductionism*, De Gruyter.

Ney, Alyssa (2013) 'Introduction' in Ney & Albert (2013).

Ney, Alyssa (2015) 'Fundamental physical ontologies and the constraint of empirical coherence: a defense of wave function realism,' *Synthese* 192: 10, 3105-3124.

Ney, Alyssa (forthcoming) 'Finding the world in the wave function: some strategies for solving the macro-object problem,' *Synthese.*

Ney, A. & Albert, D. Z. (Eds.) (2013) *The Wave Function: Essays on the Metaphysics of Quantum Mechanics*, Oxford University Press.

Papineau, David (2002) *Thinking About Consciousness*, Clarendon Press.

Putnam, Hilary (1967) 'The nature of mental states,' reprinted in his *Mind, Language and Reality*, Cambridge University Press, 1975.

Saunders, Simon, Jonathan Barrett, Adrian Kent, & David Wallace (Eds.) (2010) *Many worlds? Everett, Quantum Theory, and Reality*, Oxford University Press.

Schaffer, Jonathan (2010) 'Monism: The priority of the whole,' *Philosophical Review*, 119: 1,31-76; reprinted in Goff 2012.

Wallace, David (2010) 'Decoherence and ontology,' In Saunders et al 2010.

Wallace, David (2012) *The emergent multiverse*, Oxford University Press.