Panpsychism and Free Will

A Case Study in Liberal Naturalism

Philip Goff
Department of Philosophy
University of Durham
50 Old Elvet
Durham
DH1 3HN

Abstract
There has been a resurgence of interest in panpsychism in contemporary philosophy of mind. According to its supporters, panpsychism offers an attractive solution to the mind-body problem, avoiding the deep difficulties associated with the more conventional options of dualism and materialism. There has been little focus, however, on whether panpsychism can help with philosophical problems pertaining to free will. In this paper I will argue: (A) that it is coherent and consistent with observation to postulate a kind of libertarian agent causation at the micro-level, and (B) that *if* one if believes in libertarian agent causation at the macro-level, there are significant advantages in also postulating its existence at the micro-level.

I

Conservative and Liberal Naturalism: How do we find out what reality is like? It is broadly agreed that we should test our theory of reality against the data of observation and experiment. But, for any empirical data, there are always an infinite number of theories compatible with that data. Which one should we go for? The standard scientific approach is to go for the simplest. Thus, glossing over an awful lot of detail, we can sum up a broadly held view as to how we find out about reality as follows:

> *Conservative Naturalism*: Work out the simplest theory of reality that can account for the data of observation and experiment.

It is not obvious, however, that the only data that should inform our theory of reality are empirical. Here are some other plausible candidates:

Some Examples of Potential Non-Empirical Data

- Consciousness: Feelings and experiences cannot be observed – you can't look inside someone's head and see their experiences – and hence a strict conservative naturalist has no basis for postulating consciousness.[1] And yet we know that consciousness is real: nothing is more evident than the reality of one's own feelings

[1] Some might be inclined to call our immediate awareness of our experiences a form of 'observation.' But this seems to me a very different use of the word 'observation' from the standard use in a scientific context. In any case, we could simply define conservative naturalism in terms of *third-person observation* to remove any ambiguity.

and experiences. It is thus plausible that the need to account for consciousness is a further datum an overall theory of reality must account for, over and above the data of observation and experiment.

- Mathematical/Logical Truth: In addition to what we know about the world through perception, we also know through reason that reality is necessarily constrained in certain ways, e.g. contradictory states of affairs are impossible. It is arguable that a theory of reality must account both for the ground of these constraints and for our knowledge of them.
- Value truths: David Enoch (2011) has argued that we need to postulate objective, irreducibly normative facts in order to make sense of what we're doing when we deliberate. Some conservative naturalists may agree that irreducible normative facts are needed if deliberation is to make sense, but argue that in the absence of some independent reason to suppose that there are such facts, we must conclude that we are deluded in thinking that we can deliberate. However, a conservative naturalist responding in this way might be accused of double standards, as the empirical enquiry she *is* committed to relies on a series of anti-skeptical assumptions, e.g. that the external world exists and that careful use of our senses can tell us about it. Why is it permissible to adopt these anti-skeptical assumptions but impermissible to adopt Enoch's anti-skeptical commitment to the possibility of rational deliberation? If Enoch's argument is successful, we can take the reality of irreducible normative truths as a basic datum that our theory of reality is obliged to account for.
- Free will: E. J. Lowe (2008) argued that a commitment to libertarian agent causation is pre-supposed by the belief that humans are able to respond to reasons. If all of our decisions are the result of prior causes, he argued, then normative facts have no role to play in explaining our behaviour. Again, some conservative naturalists may be tempted to infer from this that humans are not in fact able to respond to reasons; but, again, the charge of double standards may be used in response.[2] We can see Lowe's argument as an attempt to show that our theory of reality must be able to accommodate the reality of libertarian agent causation.

We can define *liberal naturalism* as the view that some or all of the above as basic data that should inform our overall theory of reality, in addition to the data of observation and experiment. Of course, I have only gestured at the case for accepting each of the above kind of data. But I hope to have shown that a commitment to conservative naturalism cannot be assumed without argument to be the default position.

The first non-empirical datum listed above is the most difficult to deny.[3] It seems to many that the reality of their own consciousness is more evident than the reality of the external world. In that case, it seems strange to take empirical data as metaphysical data, but not the reality of experience itself. Accepting this first non-empirical datum does not in itself imply any particular theory of consciousness; it might turn out that the entities postulated to

[2] Of course, in the above two cases, many conservative naturalists will adopt naturalism about value and compatibilism about free will. All I mean to do here is gesture at the case that might be made for taking irreducible normativity and/or libertarian free will as basic data for a theory of reality.

[3] It is perhaps just as difficult to deny logical/mathematical truths (thanks to Guy Longworth for making this point). However, the question of whether logical/mathematical truths have metaphysical implications is more contentious than the question of whether experiential truths have metaphysical implications.

account for observation and experiment also happen to account for the reality of consciousness. But once we depart from conservative naturalism, this cannot be assumed at the start of enquiry.

I have found in my discussions with philosophers, both amateur and professional, that a great deal of talking at cross purposes results from these methodological assumptions not being out in the open. I defend panpsychism: the view that consciousness is a fundamental and ubiquitous feature of reality. If we assume conservative naturalism, panpsychism is absurd, as empirical science does not tell us that consciousness is everywhere (indeed, arguably conservative naturalists ought to think that consciousness is nowhere). In many of the discussions I have, it quickly becomes clear that my opponents are assuming conservative naturalism and inferring on that basis that panpsychism is absurd. But I agree with that inference! To be sure, my opponents may hold that we have very good reason to adopt conservative naturalism (although the view is often assumed and rarely defended); still, it ought to be possible for the sake of discussion to consider the following conditional:

> *If* liberal naturalism (of a certain form) is true, panpsychism is plausible.

It would be perfectly consistent for a conservative naturalist to think that panpsychism is implausible while agreeing with the above conditional. Sadly, the discussion often doesn't get that far.

In this paper, I will be considering *pan-agentialism*: the view that libertarian agent causation is a fundamental and ubiquitous feature of physical reality. Conservative naturalists will take this position to be absurd, and so they should, given their methodological starting point. But the point of this paper is not to argue that pan-agentialism is plausible. It is rather to argue for the following conditional:

> *If* the reality of libertarian agent causation and consciousness are taken as basic data for a theory of reality, then pan-agentialism is plausible.

I hope that readers, regardless of their own methodological commitments, can see the value of assessing this conditional. Afterall, even if Lowe's argument (briefly described above) for a basic commitment to libertarian free will is unsound, it is not *obviously* unsound. It is worthwhile exploring the implications of methodological assumptions that are not obviously false.

The pan-agentialist picture we will be exploring is a form of the Russellian monist theory of consciousness; in section II I will briefly outline and defend Russellian monism (a position I have defended at great length in Goff 2017 and 2019). In section III, I will outline my understanding of libertarian agent causation. In section IV, I will outline pan-agentialism, before going on in section V to argue that this view is preferable to other forms of libertarianism.

II

Panpsychism and Panprotopsychism: Feelings and experiences don't seem to show up in the story of the brain that we get from neuroscience. Some respond to this by postulating consciousness outside of the physical workings of the body and the brain. At the other extreme, some conservative naturalists (Frankish 2016) deny that consciousness exists: if consciousness doesn't show up in our third-person science of the body and brain, then we shouldn't believe in it. From the perspective of liberal naturalism, a middle way between these two extremes is desirable. We liberal naturalists accept the reality of consciousness as a basic datum over and above the data of observation and experiment, but we seek the most simple and unified account of how this datum fits together with the data of third-person science.

Given these aims, panpsychism is an attractive option; in particular the form which has become known as 'Russellian panpsychism,' so called because it is inspired by certain theses defended by Bertrand Russell (1927) in the *Analysis of Matter*.[4] The starting point of the Russellian panpsychist is that physical science doesn't really tell us what matter is. This seems at first a bizarre claim; if you read a physics textbook you seem to learn all sorts of incredible things about the nature of space, time and matter. But Russellian panpsychists point out that physics, for all its richness, is confined to telling us about the *behavioural dispositions* of matter. Physics tells us, for example, that matter has *mass* and *charge*, properties that physics characterizes in terms of behaviour: charge is defined in terms of attraction and repulsion, mass is defined in terms of gravitational attraction and resistance to acceleration. Physics tells us nothing about the categorical properties underlying these dispositions.

It is arguable, then, that there is a huge hole is our scientific story of the universe; the proposal of the Russellian panpsychist is to put consciousness in this hole. Their bold claim is that the categorical properties of matter are forms of consciousness. Physics tells us what mass and charge *do*, in the sense of capturing the behavioural dispositions that these properties endow to their bearers, but in their categorical nature mass and charge are forms of consciousness.

What this offers us is a beautifully simple and elegant way of bringing together what we know about ourselves from the inside and what science tells us about the body and brain from the outside. There is just matter, nothing supernatural or spiritual, but matter can be described from two perspectives: physical science describes matter, as it were, *from the outside*, i.e., in terms of its behaviour; but matter *from the inside*, in terms of its categorical properties, is constituted of forms of consciousness.

One might worry that it's uneconomical and unnecessary to postulate experiential properties in *all* matter. The panpsychist needn't claim that literally all things are conscious. The standard form of the view is that fundamental physical properties – mass, spin and charge – are incredibly basic forms of consciousness, and hence that the fundamental constituents of physical reality have incredibly simple forms of experience. But a panpsychist needn't hold that all arrangements of micro-level entities result in subjects of experience; she needn't hold, for example, that rocks and socks and conscious (as opposed

[4] For recent works on Russellian panpsychism, see Strawson 2006; Coleman 2014; Chalmers 2015; Goff 2017; Roelofs 2019; Mørch 2019, or the essays in Alter & Nagasawa 2015, Brüntrup & Jaskolla 2016 and Seager 2020.

to being constituted of micro-level entities that are conscious). Still, panpsychism implies that consciousness pervades the universe, and one might wonder why such a commitment is required in order to explain consciousness in brains.

Those motivated by this concern might be interested to know that there are non-panpsychist forms of this Russell-inspired approach to consciousness, according to which fundamental physical properties are forms of *proto*-consciousness. Proto-consciousness properties are categorical properties which are not themselves forms of experience but are somehow intrinsically suited to realizing consciousness when combined in certain complex forms.[5] David Chalmers (2015) has dubbed these views collectively 'panprotopsychism.'[6] For those motivated by parsimony, however, it is important to appreciate that panprotopsychism is no more economical than panpsychism: both views postulate categorical properties that go beyond the dispositional properties revealed to us by physical science. And the trouble with panprotopsychism is that it invariably collapses into mysterianism, as we have no way of accessing the categorical nature of matter other than our immediate awareness of our own conscious experience.[7]

Do we need categorical properties at all? Pan-dispositionalists argue that all fundamental properties are pure behavioural dispositions: that once you know everything there is to know what an electron does, you know everything there is to know about what an electron is.[8] I'm not convinced pan-dispositionalism is intelligible, but even if there is a possible world in which pan-dispositionalism is true, that isn't a world that contains consciousness.[9] At least, that is what the classic arguments against physicalism – the knowledge and conceivability arguments – try to show.[10] I have defended these arguments at length elsewhere and will not repeat them here. But the basic idea is that conscious experiences involve a qualitative character – what it's like to see red, or to feel pain, etc. – and that that qualitative character cannot be captured in purely causal-dispositional terms. The fact that there's consciousness entails that there's more to our world than mere causal structure.

I suppose one could have a view according to which pan-dispositionalism is true everywhere except for the brains of living organisms, which somehow grow their own categorical properties. But this view would be as ugly as dualism. Russellian panpsychism and Russellian

[5] More precisely we can say that facts about proto-phenomenal properties a priori entail facts about experience, such that this a priori entailment is not wholly dependent on the causal/dispositional features of the proto-phenomenal fact. In this way we can distinguish panprotopsychism from traditional forms of physicalism, which hold either than there is no a priori connection between physical facts and facts about experience – what David Chalmers (2002) calls 'type-B physicalism' – or that the a priori connection is dependent on functional/causal analysis of experiential facts – what Chalmers calls 'type-A physicalism.'

[6] For examples of panprotopsychism, see Stoljar 2001; Pereboom 2011; Coleman 2014.

[7] There is one form of panprotopsychism that does offer a positive characterization of the categorical properties of matter: panqualityism, prominently defended by Sam Coleman (2014). According to this view, the categorical properties of matter are *unexperienced qualities*. I raise problems for panqualityism in Goff 2017, although it's definitely a view that should be taken seriously.

[8] See Bird 2007; Ellis 2001, 2002; Molnar 2003; Mumford 2004.

[9] There is a familiar charge that pan-dispositionalism entails a kind of vicious circularity, as everything ends up being defined in terms of everything else; see Campbell 1976; Robinson 1982; Armstrong 1997, Lowe 2006, Goff 2017, Ch. 6.

[10] Jackson 1982; Chalmers 2009; Goff 2017.

panprotopsychism look to be the most elegant and parsimonious ways of fitting consciousness in to the world. The collective name for these views is 'Russellian monism.'[11]

Crosscutting the distinction between panpsychism and panprotopsychism is a distinction between strong and weak emergentist forms of Russellian monism.[12] This distinction concerns the ontological status of emergent, macro-level, states of consciousness, such as the consciousness of humans and other animals. Weak emergentists think that facts about macro-level consciousness are wholly grounded in facts about micro-level (proto)consciousness. Strong emergentists, in contrast, think that facts about macro-level consciousness are fundamental facts in their own right, causally dependent on (but not grounded in) facts about micro-level (proto)consciousness.

A couple of analogies might make the view clearer. For the weak emergentists, the relationship between macro and micro (proto)consciousness is a bit like the relationship between a party and the corresponding people partying. The fact that there is a party in a certain location is nothing over and above the fact that there are people dancing, drinking, etc. in that location. Similarly, according to weak emergentism, the fact that I am currently having an experience of red is nothing over and above the fact that some of the particles making up my brain are currently having such and such (proto)experiences and are arranged in such and such ways. For the strong emergentist, the relationship is more like the ritual dance of wizards that brings into being a demon; the demon is very much something over and above the wizards and their dancing, but the latter created the former and perhaps sustains it in existence. Similarly, according to strong emergentism, my consciousness is something over and above the (proto)consciousness of the bits of my brain, but my consciousness is nonetheless brought into being and sustained by the activities of my (proto)conscious parts.[13]

If one accepts a strong emergentist position, and hence accepts that macro-consciousness can't be reduced to micro-consciousness, why would one be a panpsychist? Why not simply hold that consciousness strongly emerges from bog standard physical stuff? In fact, if one accepts that macro-level conscious states are fundamental, then there is a powerful simplicity argument in favour of Russellian panpsychism (Goff 2017). In general, Russellian monism is a much simpler and more unified picture of nature than dualism, which gives us reason to prefer the disjunction of panpsychism and panprotopsychism over dualism. And, relative the background assumption that some conscious experiences are fundamental, anyone tempted to reject the panpsychist disjunct would need to come up with a reason for believing there is some other kind of fundamental property-determinable in addition to the determinable of consciousness. We have no way of accessing the categorical properties of matter outside of our own brains, and hence we would need a reason for supposing that matter has two kinds of fundamental categorical property – consciousness properties and some other kind – rather than just one.

To summarise, we have four forms of Russellian monism:

- Weak emergentist panpsychism
- Strong emergentist panpsychism

[11] See Alter & Nagasawa 2015 for a collection of essays on Russellian monism.
[12] See Goff 2017 and Roelofs 2019 for defences of weak emergentism; see Mørch 2019 and Brüntrup 2016 for defences of strong emergentism.
[13] For more detail, see Goff 2017, Ch. 2.

- Weak emergentist panprotopsychism
- Strong emergentist panprotopsychism

As we shall see, the panlibertarian view which is the main focus of this paper is most naturally construed as a form of strong emergentist panpsychism.

III

Libertarian Agent Causation: Perhaps the best way to introduce agent causation views is in terms of the contrast with the orthodox view of mental causation in analytic philosophy of mind. According to this orthodox view, the actions of agents are caused by their mental states. Suppose Sarah raises her hand to ask a question in the Q&A after a talk. Orthodoxy might explain this in the following way:

(A) Sarah has a desire D to ask a question
(B) Sarah has a belief B that raising her hand is the best way to ask a question
(C) D and B in conjunction cause Sarah's hand to go up.[14]

Agent causation views, in contrast, hold that, it is *Sarah* – and *not* her mental states – which is the cause of the hand raising (assuming this is a case of free action). I will understand *libertarian* agent causation to be agent causation in which the agent's causing of a given event (in a case of free action) does not have a prior cause. The libertarian is not claiming that *what happens* lacks a cause; in our example, the hand raising is caused by Sarah. The point is that the *Sarah's causing of the hand raising* does not itself have a cause (assuming this was a case of libertarian agent causation).[15]

One common objection to libertarian agent causation is the charge that there is no coherent way of distinguishing an act of uncaused agent causation from a random and senseless occurrence.[16] If Sarah's raising of her hand cannot be causally explained in terms of *any* of her beliefs, desires, character and psychology more generally, then we seem forced to

[14] See, for example, Davidson 1963; Dretske 1988.
[15] O'Connor 2000; Clarke 2003; Steward 2012. There are subtle variants on the account of libertarian agent causation I have given. Clarke 1993 defends a form of agent causation view according to which both the agent and her mental states cause the result of the action (e.g. the hand raising). This view arises from the perceived need to account for the motivational strength of 'reasons', which Clarke, like many analytic philosophers, identifies with mental states of the agent. I think Lowe 2008 is closer to the truth when he follows Dancy 2000 in thinking of reasons as external states of affairs, e.g. the fact that eating the cheese will bring me pleasure is a reason to eat it. As we will see, on my way of thinking about libertarian agent causation, the agent alone causes the result of the action (e.g. the hand raising) but is responding both to normative facts and to its conscious inclinations. O'Connor holds that the agent causes herself to have an intention to act rather than directly causing the result of the action; I'm happy to hold that sometimes the agent causes things to happen via causing intentions and sometimes brings about external events in a more direct manner (although, of course, still via some kind of chain of causation).
[16] See, for example, van Inwagan 2000. van Inwagan expresses this argument with the heuristic of imagining we 'rollback' the scenario after the decision has been made, to see what they would choose if the action were taken again. He supposes we have a situation which Alice freely decides to tell the truth. If God allows the action to be 'replayed' 1,000 times, in what percentage of the replays would Alice tell the truth and in what percentage would she lie? Answering this question, according to van Inwagan, would reveal the objective probabilities concerning how likely each possible action is. Like Lowe (2008), I see no reason to think there is a fact of the matter as to what would happen if God did this (nor that there is a modal truth expressing what would happen in an inter-world version of the rollback thought experiment).

conclude that the supposed decision to act was really just a spontaneous and meaningless event – analogous to the random decay of a radioactive isotope – and not something we can in any reasonable sense call a 'choice.'

E. J. Lowe (2008) had a very simple, and in my view effective, response to this charge. What is distinctive about genuinely rational choices, as opposed to random happenings, is that the making of a rational choice involves *responsiveness to normative facts*, or at least to what are perceived to be normative facts. Suppose the following three facts obtain:

- Sarah had an important contribution to make to the discussion (call this fact 'F1'),
- F1 counted in favour of Sarah's raising her hand to ask a question (call this fact 'F2'),
- in making her decision Sarah was responding to F2.

The fact that the initiation of a certain behavior involves responsiveness to normative facts is sufficient to make it a genuine choice rather than a random occurrence.[17]

I think Lowe's answer is good but incomplete. For not all of our choices involve responding to normative facts, or even to what we perceive to be normative facts. Sometimes our choices involve responsiveness to our conscious inclinations. By 'conscious inclination', I mean the felt desire to do something, e.g. the felt inclination to scratch an itch or to eat some chocolate. Suppose Anushka is trying to avoid cheese because she thinks veganism is morally obligatory, but she also has a strong conscious inclination directed towards eating the cheese that happens to be in front of her. After a little internal wrestling, Anushka 'gives in' to the inclination and eats the cheese. That Anushka's initiation of behaviour can be explained in terms of her responsiveness to her conscious inclination would seem to be sufficient for making it a genuine choice rather than a senseless happening.

Opponents of libertarian agent causation might argue that, in such a case, Anushka's desire *causes* her action of consuming the cheese. Whether or not this is true, it certainly doesn't phenomenologically seem to be true, at least not as a general rule. Perhaps when under immense duress, for example when undergoing torture, a person's conscious inclinations, e.g. to avoid pain, compel their actions. But most of the time when we feel hunger, thirst or sexual desire, it seems as though we can *choose* whether or not to act on the basis of it. I am not arguing that this phenomenological impression is accurate; we are simply discussing whether or not libertarian agent causation is coherent. In order to demonstrate its incoherence, the opponent of libertarian agent causation would need an argument to show that how things seem, in this regard, makes no sense.

I will call actions in which one entirely responds to one's conscious inclinations, cases of 'arational agency.' Human agency is a complex mixture of rational and arational agency, but some non-human animals may be purely arational agents: conscious subjects who are unaware of normative facts but are able to act in responsiveness to their conscious inclinations.

[17] Isn't the claim that Sara's choice involved responsiveness to a normative fact equivalent to the claim that that normative fact caused the resulting action (a challenge raised to me by Guy Longworth)? The view that reasons cause actions is normally associated with the view that reasons are identical with mental states such as beliefs and desires. But on the libertarian view I am considering here, normative facts are not constituted by mental states but are rather mind-independent facts outside of the causal realm. Normative facts impact the world, on this view, in virtue of rational agents recognising and responding to them.

The philosophy of mind literature tends to lump together many different things under the heading of 'desire': on the one hand, felt inclinations such as hunger and thirst, and on the other hand a very general notion of psychological aiming, e.g. Anushka's being psychologically aimed at not eating cheese because she thinks veganism is morally obligatory. In doing this, we conflate what seem prima facie to be two very different kinds of free action. When Anushka 'gives in' to her craving and consumes the cheese in an animal fury, she is responding to her conscious inclinations. When David calmly and rationally decides to eat the cheese because he knows it will give him pleasure, he is responding to a normative fact: the fact that a certain non-normative fact – namely, *that the cheese will give him pleasure* – counts in favour of eating the cheese. At least, careful attention to our ordinary ways of thinking about ourselves suggests that there are these two quite different forms of responsiveness, and argument would be needed to undermine the coherence of this distinction.

This paper is not arguing for or against the reality of libertarian agent causation; the claim is rather the conditional that *if* one is motivated to believe in libertarian agent causation, then one has grounds for taking pan-agentialism very seriously. Hence, I will assume in what follows the correctness of the libertarian agent causation position outlined above, in order to consider the attractiveness of pan-agentialism conditional on a commitment to libertarian agent causation.

IV

Pan-Agentialism: Pan-agentialism is the view that libertarian agent causation is a fundamental and ubiquitous feature of nature. I will begin by outlining a specific version of pan-agentialism, before going on to argue that it has significant advantages over the standard metaphysical framing of a commitment to libertarian agent causation.

Pan-agentialism is a form of powers realism, which I understand to be the view that causal facts are grounded in the causal powers of objects. On a standard form of powers realism, a power is defined by its *stimulus* and its *manifestation*. The manifestation of *fragility*, for example, is breaking, which is stimulated when the bearer of fragility is impacted with sufficient force. Powers realism tends to go along with what we might call *causation by compulsion*, whereby the manifestations of powers at earlier times stimulate powers at later times, a bit like a chain of dominoes. On the pan-agential view, in contrast, a power is definitionally associated with a manifestation but *not* a stimulus; instead, the power manifests when and only when its bearer chooses. According to pan-agentialism, this is the entire story of causation, right down to the micro-level where we find purely arational agents. Nothing compels a particle to act, on this view, the particle chooses for itself.[18]

One might wonder how this is consistent with observation. If fundamental particles are freely choosing when to manifest their causal powers, why do we not have more of a chaotic world? A pan-agentialist response to this worry can build on the following principle:

> *Desire Magnetism* – If there is a purely arational conscious subject S, such that (A) S has a strong conscious inclination to Φ, (B) S has no conscious inclination not to Φ (or to do something inconsistent with Φing), and (C) S is able to Φ in response to its strong conscious inclination to Φ and is aware that it is able to do so, then S is

[18] I will think of micro-level entities as particles, for the sake of simplicity.

inevitably going to Φ (unless some of (A)-(C) are removed before S has a chance to Φ).[19]

Mature human beings don't always act on their conscious inclinations. Even if Anushka has a strong conscious inclination to eat the cheese, she may refrain from doing so because of the perceived normative fact that veganism is morally obligatory. In this case, Anushka responds to a perceived normative fact – the fact that veganism is morally obligatory – over her conscious inclination to eat the cheese. But if a very young child has a strong conscious inclination to eat some cheese, knows it is able to eat it and has no desire to the contrary, and is too young to rationally deliberate, then the child is certainly going to try to eat the cheese. Everything in the child's psychology inclines her to eat the cheese and nothing inclines her not to. Of course, something could get in the way and stop the child responding to its inclination; a lightning bolt could hit before she has time to choose. But if her capacity to choose is left free to run its course, she's going to eat the cheese. Why would she not?

Desire Magnetism is the arational analogue of a principle Richard Swinburne (2004) holds is true of rational agents:

> *Reasons Magnetism* – If there's a conscious subject S such that (A) S is aware that she has an all things considered very strong reason to Φ, (B) S is able to Φ in response to this reason and is aware that she is able to do so, and (C) S has no conscious inclination not to Φ, then S will inevitably Φ (unless some of (A)-(C) are removed before S has a chance to Φ).[20]

Ordinary human adults regularly suffer from weakness of the will. Anushka strongly believes that she shouldn't eat the cheese but has an incredibly strong conscious inclination to eat it; in such situations, it's a wide-open question whether she will choose to follow her inclinations or to follow what, by her lights, she has strong reason to do. But if Anushka judges that she has overwhelming reason not to eat the cheese and has not the slightest conscious inclination to do so, then of course she will refrain from eating it. In the description of the situation, there is nothing that could possibly explain why she would choose any other course of action. On the basis of something like the above principle, Swinburne argues that God's perfect goodness follows from Her omniscience and omnipotence. God knows what She ought to do (i.e. what she has on balance most reason to do) and has no conscious inclination to do otherwise; it is inevitable, on this basis, that God will do what She ought.

With Desire Magnetism in place, the pan-agentialist is able to account for the predictability of nature. She can hold that micro-level entities at earlier times cause micro-level entities at later times to have strong conscious inclinations to act in a certain way, and that the later micro-level entities are able to act on those inclinations and are aware of that fact. Given Desire Magnetism, it will follow that the later particles will inevitably act on their inclinations and hence behave predictably (assuming they have no conscious inclinations to

[19] I am building into 'S is able to Φ' not only that S has the capacity, but that the right conditions are in place to enable S to act upon that capacity.

[20] I present the principle here in a slightly modified form.

the contrary, which could be built into the theory). Strictly speaking, particles are free, but they're too stupid to do anything other than follow their inclinations.[21]

In what sense is it *inevitable* that particles will follow their inclinations? I suggest the pan-agentialist should take the simple view that this word expresses *metaphysical necessity*: in any possible world in which particles have the same inclinations and the same causal powers, they behave in exactly the same manner.

An obvious worry arises at this point. If how particles will behave at later times is metaphysically necessitated by facts about earlier times, one might reasonably wonder in what sense they *could have done otherwise*. The pan-agentialist can respond that a particle is able to behave otherwise in the sense that it is part of the essential nature of its causal powers that it could have done otherwise. A theological analogy may help here. Could God have tortured children for fun? The answer is surely 'yes': God is omnipotent, and it is part of the essential nature of omnipotence that an omnipotent being can do anything, which logically entails being able torture children for fun. And yet, there is no possible world in which God tortures children for fun, as this would be inconsistent with God's perfect goodness (which, if Swinburne is right, follows from Her omniscience and omnipotence, given Reasons Magnetism). Similarly, the pan-agentialist may hold that the causal capacities of particles are essentially such that the particles could act differently; at the very least they are able not to manifest their powers. And yet, one can plausibly argue, on the basis of Desire Magnetism, that it is inevitable – metaphysically necessary – that they will manifest their powers.

A sense of inconsistency remains; it just seems contradictory to claim both:

(A) event E makes it necessarily the case that X will Φ

and yet also:

(B) X is capable of not Φing, even after E has occurred.

I think this sense of inconsistency arises from implicitly understanding relationships of metaphysical necessity between contingent events as cases of causation by compulsion. If E1 necessitates E2, we read that as E1 compelling E2 to occur. However, the pan-agentialist can hold that E1 metaphysically necessitates E2 not by compelling E2 into being but by making it inevitable that a free agent will bring about E2.

V

Advantages: Even supposing pan-agentialism is coherent, what reason does a believer in libertarian agent causation have to take it seriously (recall we are assessing the plausibility of pan-agentialism conditional on a commitment to libertarian agent causation)? The most significant advantage of adopting pan-agentialism is that it enables a libertarian agent causalist to avoid what Randolph Clark (2003) has called 'the uniformity of causal power'

[21] How should a pan-agentialist make sense of probabilistic causation? One option is to build the probabilistic element into the manifestation. To take a toy example, suppose C-type particles cause E1-type events 90% of the time and E2-type events 10% of the time. The pan-agentialist can explain this with the hypothesis that C-type particles have a power to create an event that will either be an E1-type events or a E2-type event, such that there is a 90% change that it will be an E1-type events 10% it will be a E2-type.

objection. In most agent causation views, causal relations in the inanimate world are radically different from causal relations involving agents. In the animate world, it is generally assumed, there is causation by compulsion: prior events compel later events to occur. Whilst in humans (and perhaps some other animals), according to libertarian agent causalism, we find the emergence of a radically new form of causation, whereby an entity freely chooses to initiate forms of behaviour. For naturalists who seek as simple and unified a view of reality as possible, this is not a good picture.

I suspect this is at the root of much hostility towards libertarian agent causation views. Many philosophers have a hunch that libertarian agent causation is in some way at odds with contemporary science, despite the fact that there is no observational evidence against it.[22] This sense of a tension between science and agent causation may be due to the deep disunity in the standard libertarian agent causation picture, according to which the causal powers associated with humans are radically different from those we find in the scientific story more generally. The pan-agentialist, however, avoids this problem altogether. On her view, there is only one form of fundamental causal interaction: free agential causation.

But does the pan-agentialist really avoid disunity in nature? Even if particles 'activate' their manifestations, this seems to have little to do with choice as it occurs in humans, as the latter is essentially directed at achieving certain goals. When I choose to go to the fridge to get a beer, I'm consciously aware of the state of affairs I'm aiming at bringing about. Particles, in contrast, do not have the mental sophistication to understand the things they make happen. If particles really do choose, then in some sense they must represent the options available to them, but a plausible version of pan-agentialism would hold that a particle's options are represented to it simply as brute options: at the minimum simply 'DO!' (where the particle, in principle, has the option of not doing). This seems on the face of it to have little in common with the goal-directed choice of humans, and hence we seem not to have avoided a sharp division between the causal powers of humans and the causal powers of particles.

Clearly the free choices of humans are very different from those of particles. I believe, however, that the pan-agentialist can tell a story as to how the kind of 'choice' adult humans have arises from the kind of 'choice' particles have:

> There is something the agency of a human infant has in common with the agency of a particle: in both cases, possible choices are represented to the agent as brute options (although, of course, the possible choices available to the infant are vastly more complex than those available to the particle). The libertarian capacities of the infant are powers to make changes in its brain, although obviously the infant doesn't conceive of these powers as such; the infant just knows that it can do *this* or *that*, and that these options correspond in some way with what it feels inclined to do. As the infant develops, those brute options come to be represented in terms of their distal effects (e.g. raising an arm) just as the 'buzzing blooming confusion' of experience – to use William James's expression – comes to be represented as objects in an environment.[23]

[22] There are the much discussed Libet (1985) experiments and more recent experiments (Soon et al 2008) along similar lines. Like many philosophers, I think the implications of these experiments are grossly exaggerated; I discuss this in Goff 2019: Ch. 5. There are also causal closure considerations, discussed below.
[23] James 1890/1981, p. 488.

Of course, this story is, to an extent, subject to empirical confirmation/disconfirmation, but it seems no less empirically plausible than the kind of story of development that opponents of libertarian agent causation might tell. By adopting this story, the pan-agentialist can plausibly hold that adult human choice intelligibly arises from something not dissimilar to particle choice (although of much greater complexity).

Like any libertarian, a pan-agentialist will hold that human choices could not, even in principle, be predicted from knowing the kind of facts physical science teaches us about the body and the brain. To put it vividly, Laplace's demon could know every position of every particle in Anushka's body, and have complete knowledge of the laws of physics, but would still not be able to predict whether or not Anushka will eat the cheese that is tempting her. For the pan-agentialist, this is because the laws of physics track the free choices of purely arational micro-level entities in response to their conscious inclinations. However, unlike an electron, Anushka believes that there are objective reasons to act in certain ways; she may choose to respond to her inclinations, but she may also choose to respond to what, by her lights, she has reason to do. In so far as Anushka is responding to (what she believes to be) objective reasons rather than inclinations, her behaviour will depart from the causal processes that are tracked by physics (i.e. those grounded in the free choices of arational micro-level entities acting in response to their inclinations).[24]

Doesn't it follow that a new type of causal relationship has emerged, one that involves responsiveness to normative facts? Yes and no; the pan-agentialist should hold that particles do not respond to normative facts but only because they have no understanding of normative facts. Of course, it is only when we have creatures with an understanding of normativity that we get creatures able to respond to normativity. But the pan-agentialist can hold that the only new thing at play here is the understanding of normativity: free choice in conjunction with understanding of normativity yields responsiveness to normativity. Of course, we still have the massive challenge of explaining how human beings come to have knowledge of the normative, but this is a challenge on any view.[25]

The second advantage of pan-agentialist arises from its adoption of the more general panpsychist position. Most libertarian agent causalists are *dualists*, holding either that the mind/person is distinct from the brain/body (substance dualism), or that certain properties of the mind/person are distinct from the physical properties of the brain/body (property dualism). Dualism is an inelegant, disunified and unparsimonious picture of nature, with the mental properties of humans (and perhaps animals) radically different from the physical properties we find in the rest of nature. For the panpsychist, in contrast, the agent and her mental states are part of the intrinsic nature of the brain and its physical states. Human persons are simply a highly evolved form of what we find in nature more generally.

Libertarian agent causalists hold that the free choices of humans involve irreducible causal interactions. This commitment almost certainly entails the existence of fundamental objects and properties at the macro-level, as the physical processes associated with human choice occur at the macro, not the micro, level. Hence, the pan-agentialist will probably adopt a

[24] Given that we're working within a strong emergentist framework, there will presumably be strongly emergent conscious inclinations at the macro-level, and in so far as the strongly emergent macro-level subject is responding to these inclinations, this may be a further reason why its behaviour departs from what would be predicted by basic physics (which ex hypothesi tracks the inclinations of fields/particles).

[25] At least our knowledge of the normative is a challenge on any robustly realist view of normativity, but the libertarian agent causalist view we are working with assumes such robust realism.

strong emergentist form of panpsychism, according to which the part of the brain associated with human consciousness is a fundamental entity and the states of the brain associated with human choice are fundamental properties. Reality, for the pan-agentialist, is layered rather than flat. Despite this, the pan-agentialist picture is significantly more simple, elegant and unified than that of the dualist.

As a form of panpsychism, pan-agentialist is also consistent with *physical causal closure*, the principle that every (caused) physical event has a sufficient physical cause. The actions of my mind just are the actions of my brain. Pan-agentialist is not consistent with *micro*-physical causal closure, the principle that every (caused) physical event has a sufficient physical cause *at the micro-level* (or is grounded in an event that has a sufficient physical cause at the micro-level). As discussed above, pan-agentialist is committed to the view that human choices, at least those involving responsiveness to reasons, cannot be predicted from causal goings on at the micro-level. But it's not clear that we have good reason to believe micro-level causal closure. There are no peer reviewed scientific papers offering observational grounds for thinking that the causal powers of any large part of the brain are completely determined by the causal powers of the neurons and neurotransmitters which compose it.[26] At any rate, if there is observational support for micro-level causal closure, this would give us grounds for doubting any form of libertarian agent causalism.

IV

Conclusion: I have not in this paper given grounds for thinking that pan-agentialist is true, or even that it is plausible. I have rather tried to argue for a conditional: *if* we start from a commitment to libertarian agent causation, one has reason to take pan-agentialism very seriously. If we are to suppose that libertarian agent causation exists, it would be better to suppose that it is a quite general feature of matter rather than being something that magically popped up when human beings came on the scene. To put it another way, if we are radically free, then probably particles are too.[27]

References

Alter, Torin & Nagasawa, Yujin 2015: *Consciousness in the Physical World: Essays in Russellian Monism*, New York: Oxford University Press.

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

[26] Perhaps the best argument for something like micro-level causal closure is David Papineau's 2001 paper 'The rise of physicalism.' Papineau offers an inductive argument from the premise that we have never found events in the brain that cannot be explained in terms of the laws of physics to the conclusion that there are no events in the brain that cannot be explained in terms of the laws of physics. What seems to me lacking from this paper, however, is an empirically detailed account of what observations we would expect if, say, the cerebellum had emergent causal powers.

[27] This paper builds on the section entitled 'Are we really free?' in chapter 5 of my book *Galileo's Error.* I would like to thank Louse Hanson for encouraging be to attempt to turn this rather experimental brief discussion into an academic article. I would also like to thank Helen Steward, Guy Longworth and audiences at the Durham Research Seminar I did on this topic, the Nottingham Postgraduate Conference I gave this paper as a keynote for, and, of course, my Aristotelian Society talk on this topic.

Bird, Alexander 2007: *Nature's Metaphysics: Laws and Properties*, Oxford: Oxford University Press.

Brüntrup, Godehard 2016: 'Emergent panpsychism'. in G. Brüntrup & L. Jaskolla (eds.), *Panpsychism*, New York: Oxford University Press.XXX

Campbell, K. 1976: *Metaphysics: An Introduction*, CA: Dickenson.

Clarke, Randolph 1993: 'Toward a Credible Agent-Causal Account of Free Will,' *Noûs* 27(2), pp. 191-203.

Clarke, Randolph 2003: *Libertarian Accounts of Free Will*, Oxford University Press.

Coleman, Sam 2014: 'The real combination problem: Panpsychism, micro-subjects and emergence,' *Erkenntnis* 79(1), pp. 19-44.

Ellis, Brian 2001: *Scientific Essentialism*, Cambridge: Cambridge University Press.

Chalmers, David J. 2002: 'Consciousness and its Place in Nature.' In Chalmers, D. J. (ed.), *Philosophy of Mind: Classical and Contemporary Readings*. New York: Oxford University Press.

Chalmers, D. J. 2009: 'The two-dimensional argument against materialism,' in B. McLaughlin (Ed.) *Oxford Handbook of the Philosophy of Mind*, Oxford University Press.

Clarke, Randolph 1993: 'Toward a Credible Agent-Causal Account of Free Will'. *Noûs* 27(2), pp. 191-203.

Dancy, Jonathan 2000: *Practical Reality*, Oxford: Oxford University Press.

Davidson, Donald 1963: 'Actions, Reasons, and Causes'. Reprinted in D. Davidson (ed.) 1980 *Essays on Actions and Events*, Oxford: Clarendon Press, 3-20.

Dretske, Fred 1988: *Explaining Behavior: Reasons in a World of Causes*, Cambridge, MA: MIT Press.

Ellis, Brian 2002: *The Philosophy of Nature: A Guide to the New Essentialism*, Montreal: McGill-Queen's University Press.

Enoch, David 2011: *Taking Morality Seriously: A Defence of Robust Realism*, New York: Oxford University Press.

Frankish, Keith 2016: 'Illusionism as a theory of consciousness'. *Journal of Consciousness Studies*, 23 (11-12), pp. 11-39.

Goff, Philip 2017: *Consciousness and Fundamental Reality*, New York: Oxford University Press.

Goff, Philip 2019: *Galileo's Error: Foundations for a New Science of Consciousness*, London: Rider, New York: Pantheon.

Jackson, Frank 1982: 'Epiphenomenal qualia'. *Philosophical Quarterly*, 32, pp. 127-36.

James, William (1890/1981) *Principles of Psychology*, Cambridge MA: Harvard University Press.

Libet, Benjamin, W. 1985: 'Unconscious cerebral initiative and the role of conscious will in voluntary action'. *Behavioral Brain Sciences* 8(4), pp. 529-66.

Lowe, E. J. 2006: *The Four-Category Ontology: A Metaphysical Foundation for Natural Science*, Oxford: Oxford University Press.

Lowe, E. J. 2008: *Personal Agency: The Metaphysics of Mind and Action*, Oxford: Oxford University Press.

Papineau, David 2001 'The rise of physicalism'. In C. Gillett & B. M. Loewer (eds.), *Physicalism and its Discontents*, Cambridge: Cambridge University Press.

Roelofs, Luke 2019: *Combining Minds: How to Think about Composite Subjectivity*, New York: Oxford University Press.

Molnar, George 2003: *Powers: A Study in Metaphysics*, Oxford: Oxford University Press.

Mørch, Hedda Hassel 2019: 'Is Consciousness Intrinsic?: A Problem for the Integrated Information Theory'. *Journal of Consciousness Studies* 26 (1-2), pp. 133-162.

Mumford, Stephen 2004: *Laws of Nature*, Abingdon: Routledge.

O'Connor, Timothy 2000: *Persons and Causes: The Metaphysics of Free Will*, New York: Oxford University Press.

Pereboom, Derk 2011: *Consciousness and the Prospects for Physicalism*, New York: Oxford University Press.

Robinson, Howard 1982: *Matter and Sense*, Cambridge: Cambridge University Press.

Russell, Bertrand 1927: *The Analysis of Matter*, London: Kegan Paul.

Seager, William (eds.), 2020: *The Routledge Handbook of Panpsychism*, Abingdon: Routledge.

Swinburne, Richard 2004: *The Existence of God*, 2nd edn. Oxford: Oxford University Press.

Soon, Chun Siong, et al. 2008: 'Unconscious determinants of free decision in the brain'. *Nature Neuroscience II*, pp. 543-45.

Steward, Helen 2012: *A Metaphysics of Freedom*, Oxford: Oxford University Press.

Stoljar, Daniel 2001: 'Two conceptions of the physical'. *Philosophy and Phenomenological Research*, 62(2), pp. 253-81.

Strawson, Galen 2006: 'Realistic materialism: Why physicalism entails panpsychism'. *Journal of Consciousness Studies* 13 (10-11): 3-31.

Van Inwagen, Peter 2000): 'Free Will Remains a Mystery'. *Philosophical Perspectives* 14, pp. 1-20.