

ON THE ORIGINS OF THE COSMOS AND CONSCIOUSNESS

**Interview with Philip Goff, Philosopher and Consciousness Researcher,
Associate Professor/Deputy Director at Durham University (UK)**

(Episode 3, March 11, 2021)

Welcome to 'Sciences of the Origin' where we delve into the philosophical and methodological foundations of the scientific quest for the origins of the universe, life, and mind. The main aim of this project is to discuss common methodological challenges of cosmology, biology, and archaeology. The 'Sciences of the Origin' interviews are supported by the University of Oxford project 'New Horizons for Science and Religion in Central and Eastern Europe' funded by John Templeton Foundation.

We bring you an interview with Philip Goff, philosopher and consciousness researcher who is currently working as an Associate Professor and Deputy Director at Durham University (UK). He explores new approaches to the problem of consciousness and defends versions of panpsychism and Russellian monism. Philip Goff has written many philosophical articles and two books, 'Consciousness and Fundamental Reality' (2017), aimed at an academic audience, and 'Galileo's Error: Foundations for a New Science of Consciousness' (2019). The interview is hosted by Janko Nešić, postdoctoral researcher at the Institute for Philosophy, University of Belgrade.

Janko Nešić: Welcome, Philip, and thank you for agreeing to do this interview with me. How are you doing? How is everything in England?

Philip Goff: It's very good, considering it's a bit tricky at the moment with lockdowns and negotiating parental commitments and work with two young children. But I can't complain, really. I love my job and my kids so it's nice to be less busy. How are you doing?

JN: So-so. It's a bit better at the moment, with the vaccination process starting. So to begin, could you tell us a little bit about your current thoughts on the problem of consciousness,

how to solve the problem of consciousness, and generally the hard problem of consciousness? And can you tell us more about why you defend panpsychism and Russellian monism, and why are they significant today? Why are they a better solution to materialism, dualism, and other approaches to the problem of consciousness?

PG: Maybe we could start with the problem. What I'm always keen to emphasize from the very start of these discussions is that I don't think the problem of consciousness is just another scientific problem, that we just need to carry on with our standard ways of investigating the brain and we'll crack it. Although this problem is taken very seriously now, it's still a very common reaction to just say, 'Oh, it's just another scientific problem'. I don't think that's right. And one way of saying that quite straightforwardly is that consciousness is not publicly observable. A scientist can't look in someone's head and see their feelings and their experiences. Science is used to dealing with unobservables, but there's something very different here. In all other cases, scientists postulate unobservables in order to explain what we can observe. In the unique case of consciousness, the thing we are trying to explain is not publicly observable. So it's a totally different explanatory project. And I think because of that, there are limits to what we can do experimentally.

So I think the problem sort of divides up into a scientific bit or an experimental bit, and a kind of theoretical or philosophical bit. So the experimental bit is the project of mapping out what are referred to as the neural correlates of consciousness, trying to work out which processes in the brain go along with which kinds of experiences. How do we do that if we can't observe consciousness? Well, we can observe people's brains and ask them what they're feeling, and in the best cases, we can even stimulate a bit of the brain artificially and ask people what it felt like. So that is an experiment, and that's already got lots of philosophical difficulties. But at least it's a kind of experimental scientific project.

But that's not the end of the story. Even if we can work out which brain states go along with which experiences, we've still got the question: why? Why does certain brain activity go along with experience? Why should that be? And I don't think you can answer that question with an experiment, because consciousness is not publicly observable. If you're doing more experiments, you're just going to get more correlations, you'll never get to the 'why' question. And at that point I think we need to turn to philosophy, and philosophers have offered a number of different

proposals for explaining why it is. I don't mean 'why' in a sort of meaning of life sense, just why it is; an explanation of why brain activity and consciousness go together, certain kinds of brain activity, and we just have to assess those proposals and try and distinguish between them.

So that's the more theoretical task. And I've got various thoughts on how we do that. I think when you just do that, it's plain that the view I prefer, the panpsychist or more generally the Russellian monist options, just look to win hands down over the more traditional explanations of either materialism on the one hand, that consciousness is just electrochemical signaling, or dualism on the other, that consciousness is somehow non-physical outside of the physical workings of the body and brain. The panpsychist option just avoids deep difficulties present in those more traditional options, and hence is the one to be preferred. So you've got the scientific data, or how consciousness goes along with brain activity, and all of these three philosophical theories are neutral in that they all have their different explanations of that experimental data. And we just have to choose between them.

JN: So what do you think about, for example, the integrated information theory, would that be like a philosophical, metaphysical theory of consciousness that you are talking about, or is it more like a scientific theory of consciousness? So it's not enough to accommodate the hard problem of consciousness?

PG: I think it's a bit of both, actually. So one aspect of IIT, integrated information theory, is a proposal for what kinds of physical states go along with consciousness, a proposal for the neural correlates of consciousness, and that proposal is that consciousness is correlated with maximal integrated information, and we could talk perhaps about what that means. That's just a kind of straightforward experimental hypothesis, so that's on the scientific side. But Tononi builds in a lot of philosophical background and vocabulary. He talks about intrinsic existence. He proposes an identity between consciousness and maximal integrated information, so that's going more into the philosophical bit.

We had a conference on IIT in New York a few years ago at NYU, and a lot of philosophers were saying 'We could distinguish thin IIT...', which is something like the scientific proposal, '...from thick IIT', which is the philosophical framework, and Tononi wasn't too keen on separating them. In terms of the philosophical aspects, they're very interesting but I don't myself buy it. I've written a review of Christof Koch's book on panpsychism from 2019, 'The Feeling of

Life Itself', and I just reject that identity claim; because the core of the problem with materialism in general, in my view, is that physical science works with a purely quantitative vocabulary, whereas consciousness involves qualities—the redness of a red experience, the smell of coffee, the taste of mint—and I don't think you can capture these kinds of qualities in a purely quantitative vocabulary of physical science. And so if your theory of the brain is framed in a purely quantitative vocabulary, you inevitably just leave out these qualities and hence leave out consciousness itself. This is another way of saying why it's not just a straightforward scientific project, because we're not trying to explain publicly accessible data, we're trying to explain subjective qualities that are not publicly observable, but immediately apprehended in our experience.

So, that's why I reject materialism, because of this gap between the qualitative and quantitative. The reason for the title of my book, 'Galileo's Error', is that this is well understood by Galileo. When he set up physical science, he said, we want a purely quantitative mathematical science. If we want that, we've got to take consciousness outside of science. So he designed science to ignore consciousness and now people say, 'Oh, physical science has gone so well, of course it will one day explain consciousness'. Well, that's a misunderstanding of the history of science. It's gone so well because Galileo designed it to ignore consciousness.

Anyway, coming back to IIT, I think that just has the same problems as materialism, because, as I discussed in my review of Koch's book, they're trying to identify a purely quantitative state, namely maximal integrated information that's described in purely quantitative terms, with the qualitative reality of consciousness; and I just think those identities are unintelligible. So I don't agree with the philosophical bit, but that still leaves thin IIT the proposal as regards the neural correlates of consciousness. I'm open to that and I often use it in my papers because it gives us a nice, clear possibility of what might be the neural correlate of consciousness. But basically I think it's such early days in the science of consciousness. People get very excited by brain scans, but you've got to remember, every pixel on a brain scan corresponds to five point five million neurons and we're only 70 percent way through understanding a maggot brain which has much fewer neurons than the 86 billion in the human brain. So I think we're such early days in understanding actually how the brain works. I think we have a basic grip of the basic chemistry, neural firings and action potentials, and somewhat of a grip on the large scale functions. So

we've got some grip on the top and the bottom, but it's the in-between we're almost totally clueless about—how those large scale functions are realized at the cellular level. And until we have more of a grip on that, I just think it's so hard to assess these different neuroscientific proposals, like the integrated information theory, global workspace theory, and so on. So I'm somewhat agnostic.

What I try to do—I think of myself as more on the theoretical end of consciousness science—is to try and answer the philosophical question, work out a proposal of how to explain why consciousness is correlated with brain activity, a general proposal that could fit with any specific scientific proposal of what the neural correlates are. And often I talk about IIT as an example, but I hope my theory could fit with any in principle. And I would look to the experimental consciousness researchers to fill in that bit. I think we're used to in physics to thinking of the experimental bit and the theoretical bit. But I think with consciousness science, people think it's all kind of experimental. And because the unique nature of consciousness is not publicly accessible, I think we really need to get more serious about the theoretical end.

JN: I would agree with that. So I asked because some philosophers think that IIT could be combined with some form of panpsychism or Russellian monism. How should we think about IIT for start? Should we think about it in metaphysical terms, as a physicalist theory of consciousness, or could it be a panpsychist theory? Koch, I believe, would support a panpsychist IIT or something similar. That's why I posed the question, to see if there are panpsychist inclinations in such a theory, or if it can simply be viewed as another physicalist, functionalist, or materialist theory of consciousness.

PG: Insofar as we're thinking of thin IIT as just the proposal about the neural correlates of consciousness, that consciousness is correlated with maximal integrated information, I think like any neuroscientific theory, it's just neutral between all the philosophical views. Or if you are a David Chalmers style property dualist, and you are persuaded of the truth of IIT, you can explain that as Chalmers does. Chalmers is a naturalistic dualist, he thinks brain activity and conscious states are distinct, but they're tied together by natural law, by what he calls psychophysical laws. And he thinks the psychophysical laws are over and above the laws of physics—so if it were just the laws of physics, there'd be no consciousness—but because there are these psychophysical laws, they ensure that with a certain brain activity conscious experience emerges. So you could

combine naturalistic dualism with thin IIT; you could say that the reason maximal integrated information goes along with consciousness is because the fundamental laws, psychophysical laws, tie them together; or you could adopt a materialistic proposal and say: ‘No, there's just that, just an identity’, ‘Consciousness just is maximally integrated information, just as water just is H₂O’, ‘There aren't two things’; or you could develop a panpsychist or Russellian monist approach. Hedda Hassel Mørch at the Inland Norway University of Applied Sciences has done really interesting work connecting up the two, and she spent some time at Tononi's lab, and so the Russellian monist would say that consciousness is the intrinsic nature of integrated information. So the general Russellian monist or Russellian panpsychist proposal starts from this gap in physics, that physics just tells us what stuff does, how it behaves; things like mass and charge are just characterized in physics in terms of what stuff does. It leaves us completely in the dark on the intrinsic nature of stuff, what it is in itself, considered independently of what it does; because when you're talking about what stuff does, you're talking about its relationships, how the particle impacts other particles, you're not saying anything about the electron in itself. The Russellian panpsychist puts consciousness in that hole, says consciousness is the intrinsic nature of matter; so this is not dualistic. To come back to IIT, the proposed idea would be it's not that we've got integrated information on the one hand and consciousness on the other—consciousness is the intrinsic nature of integrated information.

So there's three different philosophical ways of interpreting IIT. You've got the same data—consciousness goes with the maximum integrated information—but three ways of explaining that they're different but tied together by natural law. I guess the materialist proposal is more like ‘Consciousness is reducible to maximal integrated information’, whereas the panpsychists proposal is ‘Maximal integrated information is reducible to consciousness’. So there are three different proposals. I mean, the way Tononi and Koch talk, it looks like a straightforward materialist proposal; that's how it looks to me as a philosopher on the face of it, but maybe philosophers are interpreting it in a way they didn't really intended. It's also panpsychist in the sense that it implies that consciousness is more pervasive in the universe, but that's different from panpsychism as a sort of philosophical theory of consciousness. And I think IIT is compatible with all these views, but so is any other theory of consciousness. So insofar as we're just thinking of the thin IIT, the proposal as to the neural correlate of consciousness, I think that's compatible with any of these theories, and any of the other theories are compatible with all the

theories as well, so I don't think IIT necessarily helps us make progress, particularly more than other theories of consciousness. But it's a cool, interesting theory nonetheless.

JN: If this is like the experimental neurological, physiological side of the exploration of consciousness, and there is also the phenomenal science aspect of consciousness that we want to explore, could any ideas from the phenomenological tradition be relevant for today's analytic philosophers and analytic metaphysical theories of consciousness?

PG: I think what analytic philosophers are very bad at is characterizing consciousness as it appears to us, from the first person perspective, consciousness itself. Often when philosophers of mind talk about consciousness, they talk just about colours and sounds and smells and tastes. And in fact, a lot of philosophers, people like Jesse Prinz and Michael Tye, think that that's all there is to consciousness; whereas I'm one of a growing movement of analytic philosophers who think that consciousness is much richer than that. We see things as faces; if I'm looking at you now, I see a person and a nose and a jumper and a cupboard. And that's part of the character of my experience. In other words, aspects of the character of my experience involve the deployment of concepts. And if you see something as a cupboard, that involves the use of a concept. Susanna Sieglar's is well known for arguing for cognitive penetration; well, cognitive penetration is connected to that, but cognitive penetration could mean merely a causal thing, that your cognition affects the conscious experience. So, Michael Tye could think experience is just colours and sounds and shapes, but sometimes there's a causal impact from cognition. Whereas this view that I'm inclined to is more that it's not just that cognition causally affects conscious experience—it's involved in conscious experience. In fact, I'm inclined to think that thought itself is a kind of conscious experience, sometimes called cognitive phenomenology. And what phenomenologists certainly are much better at than us, than analytic philosophers historically, is characterizing. The wonderful characterizations of consciousness we get from Husserl, for example, the talk of the horizon and affordances. And I think this is being rediscovered in analytic philosophy; people like Dan Zahavi are very good at this.

There was a conference about a decade ago on analytic philosophy of phenomenology, and I found actually many of the phenomenologists had a kind of antirealist take on things, not believing that there is in any straightforward sense a mind-independent world. I don't like that aspect of phenomenology myself. Although I'm a panpsychist, I think there is an external world

out there independent of our minds and our thoughts about it. But in terms of benefiting from the insights of phenomenologists, in terms of the characterization of consciousness, that's certainly very important and much neglected in my philosophical tradition.

JN: Yes, I agree. Also, things that are common themes in phenomenology, like subjectivity and pre-reflective self-awareness, the subjective side of consciousness, are rarely a part of discussions in analytic philosophy of mind. So I wanted to ask about the status of individuals or subjects of experience in panpsychism. Could you tell us more about your preferred version of panpsychism? Or Russellian monism? Where you are right now and how should we understand subjects of experience in that theory?

PG: As for panpsychists in general, I would say they have lots of different views on this. Some are very reductionist about subjects of experience, and think really that the subject of the individual is just a sort of bundle of experiences; Barry Dainton, for example, defended this kind of view, and Annaka Harris in her recent book on consciousness. She thinks all of the problems—the classic problem of panpsychism, the combination problem, how do you get lots of little conscious things adding up to a big conscious thing—are all rooted in the mistaken view that there are subjects of experience as opposed to just experience itself. She's writing something for a special issue of the 'Journal of Consciousness Studies' coming out on in October on my book 'Galileo's Error' as some response to that. I'm more of a hardcore believer in subjects, and I suppose I am increasingly inclined to think that subjects of experience are in some sense irreducible, that you can't account for the existence of a conscious subject in more fundamental terms, and so we have to take them as in some sense irreducible aspects of reality.

So, broadly speaking, panpsychists split between strong emergentists and weak emergentists, in terms of the connection between particle level consciousness and systems level consciousness, so biological consciousness. The strong emergentists postulate fundamental laws of nature to bridge that gap. It might just be a basic law of nature that when conscious particles are arranged in a certain specific way, maybe maximal integrated information if you like IIT, you just get consciousness associated with the whole emerging, just because of this fundamental law of nature. Whereas other panpsychists, people like Luke Roelofs, for example, are much more reductionist. These are the weak emergentists. They want to say that 'We don't need extra laws of nature; once you've got conscious particles arranged in the right way, that's all it takes for

there to be a conscious system'. I always like to give party examples—if you've got people dancing and drinking and having a good time, you've got a party. The being of the party wholly consists in the fact that there's people dancing and drinking. You don't need extra laws of nature to get a party. Similarly, for a reductionist panpsychist, the fact that there's a conscious system wholly consists in the fact that there are particles arranged in a certain specific way.

So what I've defended in my most recent paper called 'How Exactly Does Panpsychism Explain Consciousness' on my website, is the hybrid of the two. So we divide between subjects of experience or the things that have consciousness on the one hand, and conscious experience on the other. My hybrid view involves a strong emergentism about subjects, but weak emergentism about their conscious experience. So there are special laws of nature that ensure that in certain specific circumstances local irreducible subjects of experience emerge. Again, you might think that's where you get enough integrated information if you like IIT, or something else if you like another scientific theory, but these new irreducible subjects of experience emerge, and I'm one of them and you're one of them. But in my view, they don't come with their own new forms of consciousness; rather, they inherit streams of consciousness from the fundamental level, from the level of basic physics. So we're emergentists about the subject, but reductionists about its experience. And I suggest this perhaps solves some of the problems facing both the strong emergentist and weak emergentist options. The big problem with the weak emergentist view is that it seems to have a kind of explanatory gap it can't bridge. The big problem with the strong emergentist option is that you might worry if all these new forms of consciousness are popping up, as this would be empirically implausible. So this view, I hope, solves both of these problems. We bridge explanatory gaps with fundamental laws, but because the forms of consciousness are inherited from the fundamental level, we avoid the empirical worries with other strong emergentist views.

So I end up being pretty hardcore about subjects of experience. Coming back to the party, all it is for there to be a party is for people dancing and drinking. I find it hard that you could give that kind of analysis of a subject of experience. Like 'All it is for Phillip to be feeling anxious, is for his particle's to be doing something'; I don't think you can reduce claims about a particular individual having particular experiences to a more fundamental kind of description. So, pretty hardcore about subjects, individuals; maybe not necessarily about their experiences, though.

JN: I saw you wrote a paper with Luke Roelofs on phenomenal sharing? Related to the part about experiences being inherited.

PG: Yeah, we defend the coherence that two subjects of experience could share a single token experience. So if you imagine maybe two conjoined twins, maybe parts of their brains overlap, there might be one headache that they're both experiencing. So it's not merely that they have a qualitatively similar experiences, they have numerically one and the same experience. So even though my mind is irreducible, my mind might share experience of my particles. Although the view I ended up defining in my paper that I referred to earlier is slightly less reductions; that idea is more what I call inheritance, that consciousness ceases to belong to the level of physics and comes to be inherited, comes to belong to the biological subject. And then when the biological subject ceases to be, those streams of consciousness go back to the fundamental level; so it's sort of going up and down. In my paper, I run it in a cosmopsychist version where it's actually the universe that is the fundamental thing. So it's the streams of consciousness from the universe descending to the biological level and then returning to the level of the universe. That sounds a bit mystical, but the paper is just a really sort of cold-blooded, naturalistic proposal.

JN: It's good that you mention that, because I wanted to come to the issue of cosmopsychism. Recently you've written some more popular papers on the fine-tuning of the universe. So we find evidence for-fine tuning, but it's philosophically challenging what to make of it. What would be the best explanation for fine-tuning? Is it a theistic or multiverse theory, or possibly some form of panpsychism or cosmopsychism? What is your current stance on this? Is it related to your cosmopsychism theory, because you have an article in 'Aeon' magazine where you discuss cosmopsychism and fine-tuning, or is it something else?

PG: So I've published just this last week an article in 'Scientific American' arguing that there's a fallacy; in moving from the fine-tuning to the multiverse, there's just a straightforward, fallacious inference, the what's known as inverse gambler's fallacy. And actually this has been discussed for decades by philosophers, but in these very dense Bayesian articles, and philosophers don't communicate enough. There's huge interest in fine-tuning among scientists and the public, but no one has any idea of this outside of academic philosophy. So I was very happy to get that idea out to a broader audience. And I'm writing an academic version of that where I actually link the

philosophical discussion to the scientific discussion, which actually no philosophers have done in this particular discussion of the inverse gambler's fallacy. So I don't like the multiverse explanation because I think there's a fallacy in there. I mean, if we had independent evidence for a multiverse of the right kind, it might solve the problem but—and this is what I talk about in the article—I don't think you can use the fine-tuning as evidence for a multiverse. But I don't like the theistic hypothesis either, because I'm totally persuaded by the problem of evil, at least for the most traditional conceptions of God; I don't think a loving, all-powerful God would create a universe like this with so much suffering, or create intelligent life through such a long-winded, torturous process like natural selection. So I don't like that either, and was actually very reluctant to give up the multiverse hypothesis. I'd always assumed that for a long time, but I was totally persuaded that there was this logical fallacy in that position.

So, coming directly to your question—what does the fine-tuning tell us? I think fundamentally it is strong evidence that considerations of value somehow shaped the early stage of the universe. And that's weird and that's not what we expect, but I think we have to follow the evidence where it leads—and that's the spirit of the Enlightenment, you should just follow the evidence where it leads. But I think human beings aren't very good at that because they get in a certain conception of how science ought to be, and it doesn't involve teleology or value. But then it turns out we have this very strong evidence that considerations of value have played a role somehow in shaping the universe. People say, 'Oh the fine-tuning is improbable'. That's not it; any values of those constants that come up would be equally improbable. What's surprising is that they are against all the odds, and exactly the values necessary to have a universe of great value with intelligent life. Many other combinations are values of the strong nuclear force; for example, you just have hydrogen, the simplest element. This shouldn't be as contested as it is, because we now have mathematically precise ways of understanding evidence from Bayes' theorem. We could just do a straightforward Bayesian inference to reach a strong confirmation for the hypothesis that considerations of value have played a role somehow in shaping the early state of the universe. So we just have to face up to that and try and theorize how it could have happened. But I don't think we need to posit anything supernatural.

So, yeah, I've worked out a kind of cosmos. I think on reflection it doesn't actually need to be cosmopsychist; just a view in which part of the workings of the universe involved responsiveness

to value. The way I worked it out is just drawing on the observation of David Hume that science just tells us how things behave, but it doesn't tell us why they behave. Newton gave this law of gravity, mathematical law, describing how objects behave. And people said to him, why does that happen? And he said, in Latin, I don't frame hypotheses. Physicists just give us these mathematical laws; they don't explain why. So I proposed that it's a coherent, empirically adequate proposal that what is really driving the show is the universe trying to maximize value. You might think, okay, if the universe is trying to do that, why are things not better and what are the laws of physics doing? Well, in this proposal, the laws of physics record the constraints, the limitations of the universe. So this is not an all-powerful deity. This is something that's trying to maximize value, but under certain constraints, not imposed from outside; it's just limited in what or what it's able to do. And the thought is that—and here I cheekily borrow from multiverse theorists—there would be some flexibility in the earlier stages of the universe for shaping those constants if we bring in a kind of string theory; the constants are thought of as a sort of a phase of space, and multiverse theorists hypothesized that that was kind of up for grabs in the very early stages of the universe. So we've got a universe that's maximizing value, but under constraints, and those constraints are a little bit less settled in the early stage of the universe. And we explain the fine tuning that way. So lots of more details need to be filled out. But I mean, I find increasingly odd many of the arguments for God. I think there's some force to this argument, but I don't see why we have to posit anything supernatural. I don't like either extreme sort of ignoring these things that seem obviously to have force, or retreating to the supernatural. I think there's clearly a strong argument here that value is playing a role in shaping the early universe; but we don't need anything supernatural, we just need to try and work out how that could happen. So that's the approach I currently favour.

JN: It's still sounds like a cosmopsychist version of the theory.

PG: So you don't really necessarily need the universe to be conscious, just something that's responding to value. Maybe you think that has to involve consciousness, not necessarily though.

JN: Could you just elaborate a bit on that? What does it mean to have a position like cosmopsychism? Is the Cosmos a subject of experience or something similar? How should we understand that? But not in theistic, individualistic terms?

PG: Panpsychists don't necessarily think that literally everything is conscious. The idea is that the fundamental building blocks of the physical world are conscious. So if you're thinking in terms of fundamental particles, if you think the world is built out of particles like electrons and quarks, then the panpsychist view would be that these fundamental particles have incredibly simple forms of experience. And then the complex experience of the human or animal brain is somehow built up from those. But many theoretical physicists prefer to think in terms of universe wide fields rather than particles, it fits better with quantum field theory. So if we think the fundamental building blocks are these universe wide fields, and then particles are just sort of local excitations in those fields, if we combine that with panpsychism, then the fundamental forms of consciousness would turn out to be the intrinsic nature of those fields; and the bearer of those fundamental forms of consciousness—if there is one, maybe we go back to this point, some panpsychists don't like subjects—but if you do think there are subjects where you have consciousness, then the bearer, the subject that has the fundamental consciousness, is going to be the universe itself, the bearer of those fields. So you get a kind of cosmopsychism.

I defended a form of cosmopsychism in my academic book, 'Consciousness and Fundamental Reality', but there this wasn't an intelligent agent or anything like that, it wasn't good, it was just a mess. I assumed in writing that book that you have to be subject to millions of years of natural selection to become an intelligent agent. This is just the kind of messy consciousness; it was the fundamental form of consciousness, but it was just a mess. So I think that's all you'd get if you're just trying to explain consciousness. But if you then bring in the fine-tuning, then you might have reason to attribute some more sophisticated mentality to the cosmos, responsiveness to value. Awareness in some sense of the consequences of actions because...

JN: At one point, sorry for interrupting, you call it agentive cosmopsychism?

PG: Yeah, I call it agentive cosmopsychism. You know, maybe that word can put certain people off. I mean, the key thing is responsiveness to value. So you needn't think of this as an agent-like being, in many ways it may be much more limited than us—we're kind of flexible and we've got these very flexible mental representations. This could be a thing that just responds to value. And it's not sort of thinking, 'Oh, what should I do today?' It is just responding to value. I was thinking of an agent as something that responds to value, but perhaps the word agent can be a little bit misleading.

JN: So how do scientists, physicists, and philosophers of science respond to ideas about fine-tuning and the similar?

PG: I haven't talked about this too much. I got a very angry blog post from Jerry Coyne, who's written about 13 angry blog posts about me to this date, and one of them is about this stuff. But I mean, he's a very ideological sort of individual. If you think about the spirit of the Enlightenment, what was the Enlightenment all about? On the one hand, it was about just following the evidence and the arguments where they lead, but on the other hand, it involved a certain conception of what science should be, what the universe looks like according to science, that it's sort of mechanistic and not teleological. And I think since the middle of the 20th century, with the fine-tuning, they've come apart because the evidence is sort of pointing towards something teleological, whereas the picture of science that we've had for 400 years is pointing in the other direction. I think people are actually conservative, and it's very hard to just look dispassionately at the evidence, it gets into people's identity and their sense of who they are, 'This is the truth' and 'We're not like those idiots'. So, you do get hostile reactions, but then other people like Sean Carroll, who I'm sure couldn't disagree with me more, is very happy to discuss in an open-minded spirit. And who knows who's right, ultimately. And so many of my closest friends, great philosophers of physics like Barry Loewer, are again totally the opposite view, but we have mutual admiration for each other's work and wonderful open-minded debates and discussions, and that's the way it should be.

JN: Yeah, I agree. So perhaps for the final question, I wanted to ask how do you see the future of metaphysics? You said true post-Galilean metaphysics hasn't begun yet. So what do you think the future holds for the metaphysics of consciousness and in general?

PG: Great question. Big question. So I think we're going through a phase of history where people are 'I've been so blown away by the success of physical science and the incredible technology that's produced', that they're inclined to think, 'This is it, this is everything, we found the truth'. But in my view, if you trace things back to the start of the scientific revolution, I think the reason it's been so successful is because it was designed for a quite focused, specific task, namely accounting for publicly observable data with a quantitative vocabulary.

But I think there are lots of things we know to be real that can't be accounted for in that way, that are just a totally different exploratory project. Consciousness, for example, is not about

accounting the publicly observable data, it's about accounting for the invisible, subjective qualities that we immediately apprehend in our experience. So consciousness is the most obvious one, and I think that's the one that's going to persuade the world in the first instance, but other things like facts about value, facts possibly about free agency, facts about what grounds mathematical and logical truth. There are many things that need to be accounted for, many questions we need to answer to inform our theory of reality, which aren't straightforward scientific questions. So I think we need to get out of this scientific phase to appreciate that there are things we know to be real. Not on the basis of observation of experiments; that's our current world view—the only things we should really believe in are the things we know on the basis of observation experiments. If you religiously followed that, you wouldn't believe in consciousness. Daniel Dennett is wonderfully consistent on this because consciousness is not known about in that way.

So I think we're getting there, I think we're gone from people denying the existence of consciousness and much of the 20th century pretending it doesn't exist, to people now taking it very seriously because you can't really pretend it doesn't exist, but still thinking, 'Oh, just do more science'. I think people are now starting to see the philosophical underpinnings of the problem, see that it's not a straightforward scientific problem. I think that consciousness will, has to, ultimately break us out of this scientific phase of history and perhaps open us up to the fact that there are other things we know to be real that also are not straightforward scientific phenomena. And that will be a radical shift in the way we think about science, the way we think about reality, and I think at that stage, we can start for the first time in history doing metaphysics properly. I think we've never been in a phase in history where we've had mature natural science, serious metaphysics, and people taking consciousness seriously. I think the last 30 years we've started to get serious metaphysics. We're starting to get people taking consciousness seriously again. I think when we have at least those things in position, we could start for the first time in history really doing metaphysics properly. So people say, 'Oh, metaphysics never got anywhere'. I just think, we haven't really started it yet. And it's about time we got on with it and it might work.

JN: We'll finish the interview on that optimistic note. So thank you again for taking time to talk to me today. It was a great pleasure.

PG: Thank you.

To find out more about the ‘Sciences of the Origin’ project, make sure to follow us on Twitter, subscribe to our YouTube channel, or visit our webpage.