

Tim Crane and Huw Price Interview Transcription

TC: I'm speaking here with Prof. Huw Price who is the new Bertrand Russell Professor in the Faculty of Philosophy, so welcome Huw.

HP: Thank you Tim, it's really wonderful to be here.

TC: Well it's great, and I should say welcome back, because many years ago you did study in Cambridge. Would you like to tell us a bit about your time here and how it affected your philosophical development?

HP: Yes, I came here I think it was the Autumn term in 1977. And the reason I came here, really, was because I had the good fortune to bump into Hugh Mellor a couple of years before that, when I was still finishing my undergraduate degree in Australia. And at that stage I was more or less set on going on in pure maths, and it was really some encouraging remarks from Hugh which convinced me that I should change my mind and do more philosophy; and that's how a couple of years later I ended up coming to Cambridge to work under his supervision for a PhD. And that, in turn, turned out to be very influential indeed, in that the main things that I've worked on since then can all be traced back in one way or another to things that Hugh was interested in at that time.

For example, one of hats I wear, as it were, is an interest in truth and pragmatism and really that all goes back to the rather pragmatic approach to the philosophy of probability that I worked on for my PhD thesis. And the reason I got into that particular topic was really that probability was famously a topic that Hugh had worked on and written on a few years before that.

And similarly, the other kind of hat I wear is an interest in philosophy of time and philosophy of physics and again, that goes back in an obvious way to some things that Hugh was thinking about in the 1970s when I met him, and which later appeared in his book *Real Time*. So although my own approach has been rather different from Hugh's in some ways, it's very much true that the particular kinds of problems which interested me derived in some sense or other from things that Hugh was doing at that stage.

TC: That's very interesting. Do you see other links with other parts of the Cambridge tradition in the 20th century or before?

HP: Well yes, certainly, as my views developed after that, one of the things I began working on- perhaps ten years or so after I was at Cambridge. I was back in Sydney at this stage – well during the 80s I worked partly in Sydney and partly at ANU. I talked to a number of people at that stage. One person who was very influential was Peter Menzies. Through him, I think, I moved from an interest in probability to an interest in causation and problems of decision theory – the debate between evidential and causal decision theory – and what that led me into was an attraction to an approach to the philosophy of causation which is very much a Cambridge one: an approach to causation in terms of agency. That's an approach which we can find in the very late work of Ramsey and then later in the 20th century in Cambridge figures or Cambridge trained figures such as von Wright and Gaskin and people of that kind.

TC: That's interesting too. You also have of course, a strong interest in philosophy of science and your book on time: *Time's Arrow and Archimedes Point* which was very influential and well received has been of interest to physicists and philosophers of physics too.

HP: Yes, and when it came out at that stage Jeremy Butterfield and Michael Redhead had their highly successful research group and graduate training group in philosophy of physics, between Philosophy and HPS here in

Cambridge. And there was some interest in my book at the time. They had a reading group on it and I came to Cambridge at some stage in the late 90s and talked to some of the people there about it.

One of the attractions in coming back here is that I can now work with people like Jeremy, who himself is back in Cambridge at Trinity, after some years at Oxford. Jeremy has links and I also have some links in my own right with some of the people in DAMTP – in Quantum Information and Foundations there, for example. So one of the exciting things, from my point of view, in coming back here, is the opportunity to work with those kinds of people on those issues in the philosophy of physics.

TC: Do you see the Faculty building new links with other departments in the University?

HP: Yes – I think it's a really excellent time to build, or perhaps *reinforce*, the link with HPS in particular, which has been there to varying degrees over the last 30 years or so, probably before that. As you know, there's a new professor in HPS as well, Hasok Chang. He and I and Jeremy and others have talked over the last year about trying to do something to raise the profile of philosophy of science in general at Cambridge. It's a rather kind of Cambridge thing that there are these immense strengths, but they're rather diversely spread —some people in the faculty, some people in HPS, people like JB in colleges, and lots of individuals within scientific departments who have an interests in the philosophy of their own subject. Like some of the people in DAMTP I mentioned before.

What we'd like to do is to provide a webpage and a network structure which makes all of that vast strength in philosophy of science visible both inside and outside the University, so that people who are interested in particular bits of the topic or people who might be interested in coming here as students to the Faculty or to HPS can see just how rich what we have here is. We have a name. We're going to call it CAMPOS (Cambridge Philosophy of Science) which conveniently means 'fields' in Spanish and Portuguese— so there's a nice little pun there. We're just working on getting our webpage together and having a seminar series. We have a conference coming up at the end of November in conjunction with a large European funded Philosophy of Science in Europe network. So that's a wonderful opportunity to do some things even better than has been done in the past, in part just by making them more visible than before.

TC: That's wonderful. So you have interest in philosophy of time, philosophy of science and an interest in pragmatism. You're also organising a big workshop on Cambridge pragmatism in May aren't you?

HP: Yes, well one of the things that struck me after I'd been offered and accepted the job was – again thinking about things I could do when I came here – I was thinking about the fact that some of the interests that I've developed in these rather pragmatic approaches to the philosophy of causation, for example, fitted in so nicely with themes that existed in Cambridge philosophy over the last century or so. It occurred to me it was really broader than that. The case of causation was just one example of a certain kind of practical concern that you could find in all sorts of Cambridge figures, approaching a wide range of topics. For example, Hugh's work on time and tense: what he's interested in doing is explaining how creatures in our situation naturally come to talk in tensed terms, without tense being in any sense a kind of fundamental metaphysical feature of reality.

So that's an example of what I would call— I'm not sure whether Hugh himself would be swayed, but what I would call — a kind of *pragmatic* approach to the problem of the nature of tense. You find similar themes in Anscombe's work on indexicality, for example. So these people, who are not in a general sense people that you would think of as philosophical pragmatists, but in particular parts of their work you find this way of approaching philosophical problems in terms of the practical role of the notions concerned in the lives of creatures like us. That has longed seemed to me to be the most interesting way of characterising philosophical

pragmatism, and what I was struck by was how much it was a Cambridge tradition. And I wanted to do something to call attention to that, and along the way to call attention to one or two figures in the past century who get a little bit less attention than they should, and I thought this would be a good way of celebrating some of those people .

TC: Great. But it's not just Cambridge people speaking at conference?

HP: No the other reason for having the conference when it's actually going to happen, at the end of May, was that there is a Sellars centenary meeting happening in Dublin the following week. And some people that I know through my own pragmatism networks in Australia over the last 10 years – people like Bob Brandom from Pittsburgh, and Michael Williams from John Hopkins, who are very good on these sorts of topics – were going to be in Dublin for that meeting

TC: That's Wilfrid Sellars, American pragmatist philosopher isn't it?

HP: Yes. So it seemed like an opportunity to snap them up and get them here to talk about this sort of stuff. Another person who's coming is Cheryl Misak from the University of Toronto – a very well known figure in pragmatist circles. I was talking to her just last weekend and she was telling me what an interesting project she was finding this whole idea of Cambridge pragmatism. I'm sure there will be Cambridge Pragmatism book from Cheryl within the next few years.

TC: Excellent. Would you like to tell us about any other plans you have for the Faculty, or for your research, or philosophy in the University that you'd like to develop?

HP: Well there's one thing that is not core philosophy but... Last summer when I was on my way here I was thinking about other things I might be able to do here, looking for opportunities which in some sense used the kind of background that I had in things like philosophy of time and pragmatism or whatever. While I was physically on my way here I went to several conferences in Europe in September. At one of them, in Copenhagen I met a man called Jaan Tallinn, who was one of the people who set up Skype. He turned out to be very interested in what he calls 'existential risk'. That is, cataclysmic risk arising from human technology – in particular in his case, from developments in AI. And I got talking to him and was very interested in what he had to say about those issues. But I also saw that there might be an opportunity for me to play a role as a kind of catalyst bringing him and his contacts and his interest in these issues into contact with Cambridge, in order to set up some kind of centre for research on these things here. I already knew of some people in Cambridge who were interested in these topics, in particular, Martin Rees, Master of Trinity, who I knew a little through philosophy of cosmology circles already, and I knew of his work on these issues of risk arising from technology.

Since I arrived here in October, I've been following that up. I've talked to Martin, who turned out to be very interested in collaborating on this project, and trying to bring Jaan Tallinn and his contacts to Cambridge. Jaan came at the beginning of February and gave a very well received public lecture organised by CSaP (the Centre for Science and Policy) and Jaan, Martin and myself and others are now working on trying to find funding for this project. I'm pretty sure we'll be able to do that. We now have an advisory board including some very distinguished academic figures elsewhere, and the whole thing is starting to take shape. And so I regard that as a very interesting and potentially extremely important project, which in some sense draws on my kind of background in philosophy of time and so on, but also takes philosophy out to make connections with much broader issues.

TC: That's very exciting. Could you say in just a few sentences, what these issues about existential risk might be?

HP: The general idea we're using the phrase 'existential risk' for, is that it's possible that developments in human technology may lead to risks which are so serious that they could threaten the survival of the species. Some people have thought that there are potential risks of those kinds in biology, and perhaps especially in artificial biology. If we created life forms which operated on something other than DNA then the idea is that our own defence mechanisms might have no defence against those things—there might be runaway effects which we simply couldn't control.

In the case of artificial intelligence, it's associated with the idea that developments in machine intelligence might get to the point where in some sense, the machines get out of control. The kinds of cases that Jaan Tallinn talks about for example, are cases where you have a system which is basically set up as an optimiser — a rather smart optimiser— to run some particular process. If that goes to the stage where in effect, it can modify its own code, then it's doubtful whether it would be controllable at that stage and you get some runaway process. Not 'intelligence' in the sense that we think of as intelligent —this thing wouldn't have anything like human goals or even a human breadth of knowledge— but it could nevertheless be a runaway intelligence which could be operated on a timescale vastly faster than our own thought processes. Jaan thinks it could lead in a very short time to what would, in effect, be a major species-threatening ecological catastrophe.

Now some people are sceptical about that. I'm not at all convinced myself, but I do think that there's a very strong case for thinking about these issues well in advance. I do think there's a strong case that some time over the next century or two we are going to encounter a major transition when we do really have AI which is potentially, in some senses, a lot smarter than we are. I think that Jaan is right about that. That's a major transition for our species and it's one which we really do need to think about.

TC: It sounds like Philosophers should be thinking about this?

HP: Yes, in conjunction with other people. Since I've started talking to these people, one group of people I've met have been some extremely smart philosophy graduate students from places we know of as some of the top ranked schools in the world, like Rutgers and MIT and so on, who are now spending a large part of their time thinking about these issues and they're incredibly smart and articulate people and easy to talk to for people like us — as philosophers we find it very easy to talk to them. One of the things I've been enjoying is those sorts of conversations and I'm hoping to involve some of those people in what we have here.

TC: One final question. What are your impressions of the Faculty after having come back after so long?

HP: Well my impressions have been what a pleasant and congenial place it is to work. It's true that famously Sydney philosophy hasn't been an entirely a coherent and harmonious place over the period that I've been there in the last 20 years, and I came in at the tail end of things much worse. And so it could be said that my perspective isn't entirely a normal one, but even leaving that aside, I've been struck by what a pleasant bunch of colleagues I now have.

TC IT's great to have you here and we look forward to working together for many, many years I hope.

HP: Thank you Tim.