The Phases and Future of the Noosphere: Segment 4

David Sloan Wilson: So what a perfect combination to talk about Pierre Teilhard de Chardin's concept of the Noosphere. He himself began with the origin of our species and then ran it all the way up to the future, including technology basically, he has a huge interest in technology, he was quite prophetic in terms of where technology was going, way back in the first half of the 20th century. And the expansion of the Noosphere, which is like this thinking envelope of human society, larger and larger scales, ultimately encompassing the entire earth, which of course could not take place without technology. And so regardless of how much you know about Teilhard, both of you are experts in the topic that he was addressing. And so Terry, why don't you begin by just outlining the map here, especially with respect to technology and the future, with the goal of global cooperation in mind, global governance and cooperation in mind, lay it out for us, and then Stuart can add to that.

Terrence Deacon: All right. Well, so one of the big issues of course, about being human is that we are dependent on each other for cognition. That is, even though we have nervous systems like chimpanzees, and in fact, there's no real fundamentally new parts to human brains, nevertheless, all of our cognitive capacities depend upon being embedded in this culture, held together by language and other forms of communication. So that in a large respect, we are not islands of cognition, we're this linked cognition.

DSW: I just wanted to insert language is itself a technology, and so we don't have to be talking about hardware. Language, spoken language is one of the first technological inventions of our species. So anyhow, now I promise I will not interrupt you, so please continue.

TD: All right. So the issue is that, of course, it's a technology, both of thinking and of communication, and communication in some respects is a kind of thinking, that it's a collective mental process. But in any case, what Teilhard realized back in the 20s and 30s when he was writing, was that in effect the growth of technologies and particularly communicative technologies and knowledge technologies in the world, is bringing us human beings closer and closer together, making us more and more interdependent, making human thought more and more a collective phenomenon. And so this analogy of the Noosphere as a thinking sphere analogous to the way the biosphere is a sphere of life on the earth, for Teilhard what it meant was that in effect over the course of time, and particularly up till the present and now into the future, there would be more and more convergence of human thought. And so his notion of the Noosphere was that over time, this convergence would grow and grow and grow. And we're very much interested in that trajectory, in this discussion.

The key innovation that's happened really just within the past few decades, of course, as a result of computer technology and communications, added to that, artificial intelligence that is offloading not just work, physical work, but now cognitive work onto devices, is intermingled with and entangled with this process of human communication, playing more and more of a role. And since Stuart is an expert in thinking about the future of AI, what AI has done, what it could do and where it's going, it's appropriate to pursue this question about what it's going to be doing, not just in and of itself as the technology, but what it's doing to us as now being in a sense cyborgs within that technology. So that our thinking process is not just distributed by language between us, but is now mediated by this elaborate web of communication that is amplifying this process by orders of magnitude.

My particular interest in conversing with Stuart about this, is not just to understand how that works, but also to understand the limits and the potential problems, because Stuart has played a great deal of the role of worrying about the control issue, how do we control this? And in particular, how do we control it when we are using it for means that might actually be destructive, for example, in terms of weaponry, in terms of sales, in terms of controlling people's thinking processes and so on. That's my hope that we can get to that Stuart.

Stuart Russell: Okay. Well, so there's a lot of different topics, different roads we could take in this conversation, so I think I'll just mention a few of them. So one of them is a topic that I think pervades much of my recent thinking about AI and its role, and that's what human preferences are. And so version zero of the theory is that roughly speaking, humans do have preferences about the future. There are futures we want to avoid, such as extinction and enslavement and various other dystopias, and futures that we would like to bring about.

And this concept of the Noosphere is really important to that because we're not born with these preferences, they result from our immersion in the Noosphere. And so understanding the dynamics of that is extremely important because to some extent, our preferences about the future end up determining what future we get. We may not get the one we actually prefer, but our preferences control the way we act and one way or another, that ends up controlling the future. So that's the first point.

I think another interesting point about the Noosphere viewed as this global thought process, as you put it, Terry, is that at the moment, I don't think AI is directly contributing as additional thinkers in the Noosphere, but it's dramatically affecting the connectivity of the thought processes that happen in the Noosphere, and it's doing that primarily through recommended systems operating in social media. And it's interesting to think that Stalin and Kim Il Sung and Pol Pot had far less control over their subject's cognitive intake than Facebook and Twitter do today.

They control the cognitive intake of billions of people, for hours and hours a day, and probably the majority of their non familial and non-work related cognitive intake. And so that's incredibly important and we have absolutely no idea what effect that's going to have, but it's going to have an enormous effect. One way of thinking about it is like, think about material science and what makes something a magnet. It's to do with lining up all of the spins of the atoms in the magnetic material, and to some extent what AI could be doing, what these recommended systems could be doing, is creating these magnetic domains where all the spins are lined up in the same direction because of the mass effect of running the same algorithm, billions of copies of the same algorithm affecting people all over the world, and that can't do anything but have a huge effect on how the Noosphere operates.

Just as when you take a normal material and turn it into a magnet, it's now a completely different thing and has all these new macroscopic properties that it didn't have before. So that's a real concern, I think, and we're just getting to grips with acknowledging that it's already happened and the revelations from Facebook and others saying, "Oh yeah, we knew that we were ripping apart societies and destroying democracy and creating these huge islands of completely misinformed people. But we were making so much money and having so much fun, that we just carried on doing it." So this, we really have to get to grips with.

But then most of my recent work on thinking about the implications of AI is not about current AI, which I think is actually still quite rudimentary, and I think we're probably going to go through several more boom and bust cycles in AI, because I think there's way too much hype going on right now, just as there was in the 80s with expert systems and the bubble will burst. But what happens when we solve those problems, when we achieve the additional breakthroughs that are required to produce general purpose AI, which would be AI systems that are capable of quickly learning to do anything that human beings can do, and probably a lot more beside. And so the concerns I have there, one of them is the problem of control. So if you're making systems that are more powerful than human beings, how do human being retain power over those systems forever? And that almost sounds like an impossibility.

And then I think the other big set of questions has to do with how do we find purpose in such a world? So even if we do retain control, we are then faced with what Kane's called our real, our permanent problem, which is how to live wisely, agreeably and well in the absence of any forces that arise from economic necessity. And those forces have shaped our society since before we were human beings, and the economic would include the need to find food, the need to ensure safety of yourself and your family

and your tribe, and shelter, and so on. This is what's been driving human life and has given people a reason to get out of bed in the morning for hundreds of thousands of years, and what happens when that goes away?