AI and You

Transcript

Welcome to episode 26, and part two of my interview with Thomas Homer-Dixon. Thomas is director of the new Cascade Institute at Royal Roads University in Victoria, British Columbia. He holds a University Research Chair in the Faculty of Environment at the University of Waterloo where he was founding director of their Institute for Complexity and Innovation. He has a PhD from MIT in international relations, defense and arms control policy, and conflict theory, which gives you a good idea of what the Cascade Institute is all about. They are a research center "… addressing the full range of humanity's converging environmental, economic, political, technological, and health crises."

In last week's episode, we talked about the work of the Institute, and Thomas' book *Commanding Hope*, which deconstructs hope as a mechanism for global change, and that's obviously right in line with the Institute's purpose.

We pick up the interview right after Tad was talking last episode about systems with nonlinear behavior, self-reinforcing feedback loops, and how they can reach tipping points that cause cascades of behavior, hence the name of the Cascade Institute. That triggered something in me and so I open the continuation of the interview talking about those models and bringing up **Systems Thinking**. You'll hear me make a note to myself to explain that during the introduction, so here it is. Many years ago I read a book called *The Fifth Discipline,* by Peter Senge, which introduced me to Systems Thinking, and it completely resonated with me. In the time I have now I'm not going to do justice to it, but it's about seeing and understanding a situation or an environment as a system or group of systems, where a system is the relationships between the components. So it's about focusing on how the whole is bigger than the sum of the parts, and what makes it so. Do by all means read the book, *The Fifth Disicpline*, it's quite digestible and you'll end up with a much better understanding of Systems Thinking.

So here we go with part 2 of the interview with Thomas Homer-Dixon.

And some of the models you were talking about there, is that also from systems thinking? If I were to insert a reference to Peter Senge and *The Fifth Discipline* before this, that would be a good point for people to investigate?

Yeah. So we've been in touch with Peter, and especially we're interested in his educational program. So one of the things we're doing within the Cascade Institute is developing, you might say, complex systems curricula for explaining these ideas simply and easily to people in ways that they can use them. So the larger framing, or what we call it philosophy, the ontology, the basic theory of reality, is a systems theory, and we have a very specific way of thinking about systems. And then there are systems that are sort of mechanical, which are linear in their behavior. And then there are systems that are complex and nonlinear in their behavior. And we're interested in the latter set because they have this ability to respond disproportionately to causes,

and that's the basic. In some sense, the basic definition of complexity is that there's disproportional causation. Sometimes small causes can produce huge effects, and sometimes really big changes in a complex system don't seem to do much at all. So the size of cause and effect become disconnected. Whereas within mechanical systems of cause and effect, you have a big cause [that produces a] big effect, you have a small cause that produces a small effect. So I mean, from a purely intellectual point of view, complex systems are way more interesting than mechanical systems.

And there will be a distinction here between systems which are the complex and that we understand the complexity of and systems whose complexity we're not aware of.

Right.

The Johari window, and I've thought often about artificial intelligence and exponential technology is that we could face an existential risk from complexity alone, that we could be done in by something we didn't understand.

Yes. And I think that's actually happening now. It was one of the basic arguments of my book, *The Ingenuity Gap*, that we are creating degrees of complexity that have rendered the world around us opaque, highly unpredictable, and perhaps very dangerous. And we're not responding to those problems with adequate solutions as fast as we need to. So a gap is opening up between the difficulty of the problems we're creating for ourselves and our ability to solve them. So when I think artificial intelligence-- I mention AI in *The Ingenuity Gap*, and I think I mentioned it again in *The Upside of Down*. It's one of those trajectories or pathways or developments, pathways of development that is potentially alarming. I'm actually more alarmed by relatively stupid AI used badly than really smart AI used in whatever way. I think that the kinds of artificial intelligence that we're liable to see, people are going to think they can do more than they can do and end up causing crises with technology that we're relying on too much that turns out to go awry.

And one of the things about the complexity in AI is, as AI systems become more complex and more capable, their failure modes rise to higher levels of abstraction. So they're not just a random bit flipping causes something on a screen to show the wrong pixel, but it could cause it to infect other machines, for instance, like the Stuxnet virus did.

Yes.

So the $64,000 question, studying these effects, and producing this research and this knowledge, how does, could, or should the results of that be used to effect change?

Well, there is a risk here and I bet it's implicit in your question, which is, you know, these are very powerful intellectual and conceptual tools, very powerful models that could potentially use for both good and for bad. And our intention is to apply these tools as much as possible, find those positive cascades, those possibilities for positive cascades, and avoid the negative ones. It's interesting, there's been this kind of transition between before and after the pandemic. Before the pandemic, when I first set up the Cascade Institute, or first introduced the idea, it was like the

world systems and the social systems around us, so the worldviews of institutions of technologies were sort of locked in place; they seemed extraordinarily rigid. And so the Cascade Institute's intent was to try to find places or ways of unlocking those systems to produce change - high-leverage intervention points. You can sort of think of sticking a lever into a crack in a system and trying to open up that crack and actually free the pieces from each other so that they can move around a bit. Then with a pandemic, it was all of a sudden as if all the pieces had suddenly disconnected from each other, and everything was moving simultaneously. So all of a sudden, we had cascades, but we didn't have any sense for which were the positive ones, which were the negative ones, how we could influence one set or the other, try to make the negative ones more positive, try to find positive ones and make them happen. And so the Cascade Institute then became very much focused on a series of projects, one of which we call the Inter-Systemic Cascades Project, which is looking at how the pandemic has affected other systems, using our tools to try to identify things like positive feedbacks that can be either exploited or need to be broken because they're driving us in bad directions. And that project, the Inter-Systemic Cascades Project has moved forward quite quickly, and we were really pleased with some of the results. Now, the question then is, are these tools potentially usable for malicious purposes? And the answer is, yes, absolutely. But most, and maybe I'm going out on a limb here, but most malicious people think less systemically. I think one of the things that is characteristic of a systems approach is you think highly relationally in terms of the connections between things. And most of, for instance, of the right-wing ideologies in the world that are inclined towards violence, are extremely divisive and individualistic in their underlying orientation towards the world. It's about breaking relationships, it's about building walls, it's about reinforcing the capacity for communities and even individuals to break free, to maintain their liberty from constraint, and to protect themselves from external threats. So it's all about trying to break down relationships and reduce the systemic connectivity in systems. So my guess is that, at least currently, these more malicious groups won't be interested in these tools at all. They're actually not even grounded in an ontology and a framework that would allow them to see the world as effectively this way. So as I say, that may be purely wishful thinking on my part, but I think we are, at least initially, so first-order approximation, putting a lot of knowledge onto the table through the Cascade Institute and just saying, "Use this if you can, and here are ways you can use it." I would expect that only certain groups are going to pick up that information and use it. And those will be ones that will have, at least within our value framework, more benign intentions, and positive intentions.

Pronouns here are kind of important when we're talking about "we", or "you", as in, "who will do this?" It actually reminds me of once when I was starting a course that was in a personal transformation organization, and someone else there said he became aware that it was going to take a different kind of person to finish that course from who he was at that point, which was rather insightful. And I often think that it's going to take a different kind of human race to get to the other side of where we are now. But I want to look at the question of where can the pressure points be exploited and by who if we have our choice. If on the other end of the scale, the malicious, like Osama bin Laden for half a million dollars changes the course of history. Donald Trump and Steve Bannon for a tiny investment throw the world into chaos by exploiting

100 to one—well, by exploiting the American psychology, but pulling off a hundred to one shot there. And it seems to me that the kind of intellectual firepower that the Cascade Institute and others are able to bring to bear here should be able to identify ways in which we can have that kind of impact positively.

Yes, that's right. And one has to wonder whether to what extent Osama bin Laden was informed by, say, preferential attachment theory within social network and network analysis and complexity science. I think not. It turns out that the World Trade Center was a very valuable target, I mean, psychologically valuable, not just economically valuable. Of course, it was economically valuable, but it had enormous psychological impact. To what extent did they see that as kind of a hub in a larger network that if they took it out, they could start to unravel larger networks? Probably they had some intuition for it, but it wasn't analytically grounded. I think the same would be true for Steve Bannon and Donald Trump that these folks were operating with an intuition, not with analysis, and they may not have actually even been terribly interested in a more precise, analytical approach. So again, the Cascade Institute is thinking in those terms, but thinking about being much more scientifically grounded and using tools that can actually see farther into the future than just in terms of the initial move, and what the downstream consequences might be for sustaining say, a positive cascade, positive outcome. You said something just in the transition to that question that I think is really important, this idea that we have to be almost a different human race. And one of the points I make in *Commanding Hope* is that I think we are at kind of cusp moments in the deep character of our species. There was a period of time that was highlighted by the German existential historian Karl Jaspers in his work in the 1930s '40s and '50s. He called it the Axial Age. It was from about 600 BC to 200 BCE, before the current era, when five human civilizations more or less simultaneously, all went through a kind of cosmological transition, a deep transition in their understanding of reality. And the really interesting thing is those five civilizations really weren't communicating very much with each other at the same time. And yet, they all made this move during this period of time, in a very similar direction that kind of laid the groundwork, for instance, among other things, for monotheism across these civilizations. And this isn't a novel idea, many people have said this, but I think we are on the cusp of something potentially similar. And the "we" at the end of the process is going to be different from the "we" at the beginning of the process. And I use "we", by the way, quite intentionally usually, to refer to the entire species, as I mentioned in the book. This is going to be a process of extraordinary maturation, or we're going to die, it's going to be simple as that. Right now, we're behaving like adolescents. As a species, we think we're invulnerable, we're not thinking about the future, and we're trashing the place. But it's becoming very clear that we're going to kill ourselves. And so, is that going to provide us with the new framework in which we learn to live on this planet as a mature species, as opposed to an adolescent one? Maybe that's the beginning of what you could call the Second Axial Age, and that would happen over the rest of the century and into the next.

To what extent do you aim your output at policymakers? Thinking about climate change, for instance, as much as individuals can recycle, go to renewable energy, and so forth, it's not enough without the Paris accords, without international agreements, without getting governments involved at a global level to set policy and regulation around industry for one

thing. So to what extent can what you do, or to what extent do you take your output and try and get that into policy conversation?

That's very much part of our agenda, our ambition, or those are definitely within our set of target audiences. I mentioned this just briefly before, we have what we call a WIT framework that underpins our work. The "W" "I" and the "T" come from the words "Worldview Institution Technology." And based on some ideas that were developed in the late 2000s out of Columbia University, we look at interlocking sets of worldviews, institutions, and technologies, how they reinforce each other these things. In the book, I talk about where one intervenes, whether one focuses on technology, on institutions, or on worldviews, or you have to kind of intervene simultaneously across all three to produce a really profound effect. But I make a pitch in the book for suggesting that the most important intervention in that cycle, the place to break in is at the worldview level. So to go directly to your point about policymakers and politicians, I think what we're finding in the world today is that, except in extraordinary circumstances, most of our leaders are not very far ahead of the underlying attitudes and values and beliefs and concerns of their larger constituencies. And so the issue then is how do you shift those constituencies and their concerns in a very rapid way. So one of the projects that we're focusing on in the Cascade Institute, in addition to the Inter-Systemic Cascades Project I mentioned earlier, is what we call our Norm Cascades Project. And this is about seeing if we can identify mechanisms by which people's beliefs and attitudes about the world, about problems like climate change, about our sense of shared fate on this planet, about attitudes toward science, for example, their attitudes in terms of their sense of responsibility to each other, and their norms of responsibility to each other, how quickly some of those can be moved in a direction that would support much more aggressive action on something like climate change. You know, if somebody in 2017 had said that a girl of 15, who's on the spectrum, was going to sit on the steps of the Swedish Parliament with a little sign saying, "School strike", and her backpack, and that she would galvanize a movement of tens, if not hundreds of millions of young people around the world in climate action, we would have said that's a fantasy, it's a ridiculous idea. But that was a very sudden and substantial normative shift. So what are the possibilities, especially in the context of the pandemic, for larger normative shifts that would then change the incentive structure for your policymakers and politicians, and this and also within the financial markets? I'm more optimistic about moving financial markets fast than I am about enterprises like the Paris accord. If we can get trillions of dollars moving out of fossil fuels into renewable energy, and some of the key, key components of renewable energy that aren't receiving sufficient investment right now, such as Ultra-Deep Geothermal, that will do more to change our prospects, to change the "T" and the WIT, and the institutions in the WIT, than anything that the governments are doing in places like Paris. That's where you get cascades. In these financial markets, you can see really dramatic cascades and the transformation of investment patterns within a matter of months even. And that depends, I think, on the underlying attitudes and beliefs and values and norms of these enormous populations around the world who stand up and say, "Enough is enough. This is crazy. We're killing ourselves. This has to stop." In the amoral financial markets, not immoral, but amoral, view their risk-benefit calculations and say, "These old investment patterns aren't working

anymore. We have to move in a new direction." And some of that's already happening in a very large way around the world. But the question is, can it be accelerated?

Wow, there's so much more that we could talk about for hours, which we don't have there, because you've started to open up for me there the idea of the science of constructing a movement. But instead, I'll ask here, there could be all kinds of people listening to this podcast and some embarking on the beginnings of careers or education, some of them thinking that they're not as intellectually inclined or capable of pursuing this kind of research, but they want to make a difference. So with that broader possible audience in mind, what would you say to people who are faced with choices at this phase in their life, and want to choose to have a positive impact?

So a couple of things. I start from the premise, and this is why I've written three major trade books now, that most of the stuff can be understood by any reasonably intelligent and thoughtful person who wants to work a little bit to try to bring the ideas onboard. I actually don't think that most of this is rocket science and actually, I'm quite anti-elitist in this respect. I figure that educated elites and scientific elites spend a lot of time building themselves barriers of concepts and theories that render themselves, render their work impenetrable, and most of the key ideas can be communicated in very powerful ways that can be understood by an audience not informed in theory and the jargon. And that's a big part, that's the third pillar of the Cascade Institute. We have the Inter-Systemic Cascades Project and the Norm Cascades Project and then we have our Curriculum Project, which is intended to take a lot of these concepts and ideas and make them usable by people in their everyday lives and in the field of political activity and political action. I mean, just as an aside, one thing I find really interesting is once you understand these tools a lot better, you can use them, for instance, in your interpersonal relationships to figure out why things keep going off the rails with a friend or neighbor or maybe even a spouse or something like that. What's the dynamic here? And it turns out that they're manifoldly useful, not just in political action, but in one's day to day life. So I like to see that knowledge as much as possible democratized. Now, to get to your question about what should people do, I think first of all, given that most people are pretty smart, that most people have the capacity to learn effectively about a lot of these challenges, to have a sense for the realistic possibilities going forward, to identify which are credible sources of information and which our incredible sources or not reliable sources of information. Most people can do that, to break out of their social media bubbles, to challenge themselves with alternative ideas, and apply themselves to think through some of these issues. And then I tell people all the time that it doesn't matter what you want to do in terms of your profession, whether you want to be an engineer or an accountant, a contractor, working in the healthcare industry, as an artist, as a philosopher, as a psychologist, as somebody doing housing retrofitting or whatever, there is a role for everybody in this set of challenges. There is absolutely infinite proliferation of economic niches for people going forward. Now, it doesn't mean you're going to be enormously prosperous, depending on the niche you identify, but we can all have our own hero project, to use the ideas of Ernest Becker whom I site in the book. The idea that we have something that gives meaning to our lives in these incredibly threatening circumstances, that gives us a sense of purpose and possibility going forward. And I provide this advice to young people all the time. And then keep in mind that sometimes one person can make

a world of difference. So I have that story in the book, maybe this is a good way for us to end today, the other key narrative arc in the book, as you know, is the story of Stephanie May, who is a Connecticut housewife in the 1950s, who single-handedly starts a national and international movement to stop the testing of nuclear weapons in the atmosphere. She has no particular training, but she has a particular kind of hope, and she's very smart about how she does things. And she starts by phoning members of the clergy in her hometown in Connecticut and getting them to distribute her petition to parishioners. And then a few months later, she's in Washington with her petition, and another housewife and she are going from politician to politician to express their concerns. And within two years, she's a key member of a national movement of mothers, probably the leader of a national movement of mothers to stop the testing of nuclear weapons in the atmosphere that was poisoning our atmosphere and making kids sick all over the world. And it's very clear that that ultimately global movement of mothers was instrumental in causing the superpowers, the United States and Soviet Union, to sign the Partial Test Ban Treaty in the early 1960s that put nuclear testing underground. And within 20 years rates of radioactivity in the atmosphere had falling back to normal. So we all have to remember that we have the capacity within ourselves, if we keep our imaginations and thinking and try to figure out avenues to respond, to produce extraordinary things. And ordinary people can do extraordinary things. And so that, I think, is what I tell young people. If we lose our hope, if we capitulate to despair, then it is truly finished. If we don't, we still have a very, very significant chance of producing a humane future for our children and our grandchildren.

What a perfect way to end. And I remember the story from the book of Stephanie May, who was doing all this with a young child in tow at her side as much as you and I are doing that, and I'm sure that you're the hero in your children's eyes for this.

That's the end of our interview. I feel very aligned with Thomas because he started his book when his children were seven and four and I finished my book when my children were seven and three. And we were both motivated to do what we do by our children.

I think that when we recognize that not enough has been done about some condition, like climate change, and that collides with the realization that we personally are doing all we are capable of about that, then hopelessness breeds. It's what makes it hard to deal with existential crisis, as we were saying in the interview. And what I try to do as much as I can is show people how they actually can do something, that there are ways they can contribute to this huge task that they may not have realized, and most importantly, to create the hope that can fuel their action.

This discussion reminds me of a story from the Apollo 11 mission. The spacecraft that took Neil Armstrong and Buzz Aldrin to the Moon had redundant parts as far as possible, so that if any component failed – and there were millions of them – another one could take over. But there was one key component that didn't and couldn't have a backup. That was the lunar module ascent engine, the motor that would blast them off from the Moon to get back to Michael Collins in lunar orbit in the Service Module. If that engine failed, they would be stuck on the Moon. We would be talking with them until they ran out of air.

Now, that possibility was known, of course, and it's catnip to the press, so at one conference a reporter asked Armstrong and Aldrin what it would be like if that engine failed. How would they feel, what would

they say, what would they do? And obviously they were expecting something moving about reflecting on their lives, talking with their families, and so on. The answer they got from Armstrong was, "I'm pretty sure we would be trying to get that engine working." That of course is why they were chosen for the mission, because they would keep working on the mission with their best selves no matter how dire the situation got. And we got an idea of how bad it could get during Apollo 13.

I always want to be careful when I start down this road that I'm not painting AI as purely an existential threat, because it's far more than that. If it were that simple then the answer would be, "Don't do that. Don't write that software. Why would anyone want to do that?" And of course, it's because the benefits of AI are just as great as the threats and certainly more immediate and easier to understand that we don't just focus on stopping it, the way we look at, say, climate change.

But the threatening aspects of AI, when we contemplate them, can be as depressing as those from climate change - and what happens to us when we think about the other ways we're putting ourselves in peril? Microplastics in the ocean, for instance, were getting bigger in the news until Covid-19 came along and now we only seem to have room to talk about one crisis at a time. This is one of the limitations of human brains, of course, that we can only think about a limited number of things at one time. There's actually a name for that number, it's Miller's Magic Number, postulated by George Miller of Princeton University, who said the number is 7 plus or minus 2. Which is why it's easier for people to remember someone's telephone number without the area code than if they need to remember that as well. Although modern research suggests the number is actually more like 4. That's basically as many different things as you can hold in your head simultaneously. If you want a quick glimpse of just how different an artificial superintelligence could be from us, get a sense of what that limitation means for your brain, and then try and imagine the capabilities of an AI that can think about four hundred things at once. Or four million.

So if humanity can face a dozen existential threats at a time but human brains can only think about four of them at a time, we've got to split the workload. What gets me upset more than just about anything else is knowing how much more we could do, with only our current resources – no new technology, no extra capital, no magical political structures – what we could do right now. If we only worked together instead of against each other 99% of the time these problems would not exist. Yes, I know, that's hopelessly idealistic. But everyone's got to have a vision, and that's mine.

What we need is to be able to look at all the threats that humanity is facing, and evaluate them all objectively for how we can respond to them without making some worse at the expense of the others. Yes, places like the Cascade Institute are working along those lines. I'm sure there are a dozen or more institutes around the world doing something like that with maybe $100,000 or a million dollars of donor funding, but that's nowhere close to what we need. Where is our global scorecard where you can go to see what the state of our danger and response is with respect to everything from coronal mass ejections to structural unemployment? Maybe one of those disconnected institutes has done something like that, but do you know where it is? Have you seen it? We need something that people know the address of as much as they know where to find Facebook and Amazon, and visit just as frequently.

For starters, we need not $100,000 but a billion dollars. The dashboard for the health and future of planet Earth. And there is one organization that could actually pull this off. Google. People already go there all the time, they have the spare capital in both cash and talent, and most importantly, making the planet safer and extending human welfare is their core goal. So if you're listening, Google, how about it?

In today's AI headlines, you know what deep fakes are, right? Training AI to understand enough about what people look like on video that you can alter a video to substitute someone else's face or have them say something different. That technology is getting better by leaps and bounds, as you might imagine, and is approaching the point where you can't tell it's being used. There are good uses for it, like fixing mistakes in video that you've shot for, say, a commercial; and you can easily tell what the bad uses are.

But now, AI is being used to fight back. Maneesh Agrawala and others at Stanford University and UC Berkeley have created a program that can spot the use of deep fake lip sync technology by recognizing minute mismatches between the sounds people make and the shapes of their mouths. In technical terms, they looked for inconsistencies between phonemes, which are the basic sounds we make with our voice, and *visemes*, which are the mouth formations associated with those sounds. It's hard for a person to track these fast enough to spot the differences, but their AI can do it.

But Agrawala who is also affiliated with the Stanford Institute of Human-Centered Artificial Intelligence, warns that there is no long-term technical solution to deep fakes. He says, "As the technology to manipulate video gets better and better, the capability of technology to detect manipulation will get worse and worse. We need to focus on non-technical ways to identify and reduce disinformation and misinformation."

I'm somewhat surprised, but relieved, that deepfakes didn't show up more in the 2020 US Presidential Election, although one was used to distort a video of Joe Biden.

Next week we'll have something a little different, since it's the holidays. I'll be talking about my TEDx experiences, as a speaker and as a curator evaluating applicants, to give you a peek behind the curtain and an idea of what it's like on the red dot. That's next week on *AI and You*.

Until then, remember: no matter how much computers learn how to do, it's how we come together as *humans* that matters.

http://aiandyou.net