AI and You

Transcript

Guest: Robert J. Sawyer, part 2

Episode 109

First Aired: Monday, July 18, 2022

Hello, and welcome to episode 109! Last week we started an interview with science fiction writer Robert J. Sawyer. Rob is one of only eight writers in history — and the only Canadian — to win all three of the world's top Science Fiction awards for best novel of the year: the Hugo, the Nebula, and the John W. Campbell Memorial Award. He received the Order of Canada, has two honorary doctorates, and has twenty-four novels, including the *WWW* trilogy about a consciousness emerging from the world-wide web. Last week we started talking about Blake Lemoine, the Google engineer who declared that their LaMDA AI had become sentient, and there will be a passing reference to that again this week; go listen to the first half if you haven't already.

Rob also has written lots on the subject of mind uploading, which intersects with AI when we think about creating AI through mimicking the structure of the human mind, and if we can do that precisely then we can also transfer it… or can we? Rob's explored that in stories like *Identity Theft,* and novels like *The Terminal Experiment,* and *Mindscan,* and *Trigger,s* and *The Downloaded*. We'll be talking about that too. Here we go with the rest of the interview with Robert J. Sawyer.

We are dancing around the question of the nature of consciousness here and so I want to get into something that you are very interested in, fascinated in, some might say obsessed with, the concept of identity transfer, moving someone's consciousness from their body to some other location and you've, more than anyone else that I've read, explored the difficult and uncomfortable questions surrounding that, which I want to get into. And maybe as an introduction, you can tell me what were your formative experiences that led you to that fascination, or influencers in that respect?

Yeah, I think I explore it most thoroughly in my novel *Mindscan*, which came out in 2005. And it's funny it's a pun in the title and almost nobody gets it because it's about scanning minds, but it's also about what distinguishes real minds from false ones: what minds can do. Mindscan you can split it into mind scan or minds can. Anyways, it's one of my more labored puns, but I thought it was very thematic. I'm not sure where if anywhere, I'd really encountered the notion, except in the progenitor of all things Rob - me - *Star Trek*, because they do in an episode written by the great Robert Bloch, best known for having written *Psycho*, which, of course, Hitchcock's film of the novel, *Psycho*. *What Are Little Girls Made Of*, where they take Captain Kirk on one side of a spinning turntable and a hunk of plaster on the other side, and managed to duplicate his consciousness into the artificial substrate and make an android copy of Captain Kirk that's supposed to be indistinguishable from the real one, except that it doesn't eat food, which is a giveaway at one point. But I was always fascinated by this. And I've got to say, I'm fascinated by transhumanism, but I'm not a transhumanist myself. Just the logic of everything we've said, you and I, in this conversation, Peter, leads me to the conclusion that there's nothing mystical or

spiritual, there's no soul in between my ears here and I wrote us a novel that won the Nebula Award, the Science Fiction Fantasy Writers of America Award for Best Novel of the year, for a novel called *The Terminal Experiment* about the discovery of scientific proof for the existence of the human soul. I don't believe in a soul. But I also don't believe that you could transfer whatever Robert J. Sawyer is into another substrate. You might be able to copy it, well, that's all well and good, then I'll have a twin brother. But I'm no more my twin brother, than any biological twin's brother is the same person. We would deviate from the moment of copying, which for a real twin, of course, is when the cell divides or whatever at the conception stage. But we would have different life experiences, different memories. If I'm looking at the twin right now, I happen to be looking across the room here and I see my award trophy case and looking the other way, looking this way, you see the physical backdrop that I use for Zoom meetings. We have a completely different point of view on the world just because we're facing in different directions. So, I'm fascinated by this and I really have interrogated it in many works in *Mindscan*, in *Red Planet Blues*, my Hugo Award-nominated short story *Shed Skin*, in the work that I just finished, which will be out this fall called *The Downloaded* which is all about uploading and then the question of okay, great, you upload your consciousness, what have you changed, you don't want to come back down and download again to physical reality? How is that effective? It fascinates me. I still think that there's no way to move consciousness, you may be able to copy it. But I don't go along with the transhumanists who say, oh, yeah, when this physical body of mine wears out, when I start to get the plaques in my brain that lead to Alzheimer's, I'll just have it copied into a positronic matrix and one of Asimov's robot bodies. No, there'll be somebody else who might have the same name, and there'll be an interesting legal battle about which of you has the token of personhood that the government recognizes as being the one and only Robert J. Sawyer.

This has a number of points of contact with artificial intelligence, one of which is Ray Kurzweil, who makes no secret about this being his ultimate goal. But as you point out, it would be a copy, and now there's this philosophical principle which people seem to intuitively buy in to, that the mind is the software that runs on the brain; and it's a principle of software that it's information; and it's a principle of information that it can be copied; and hence we get to this idea of uploading. But it is as you say, a copy; we never actually move information from one place to another, all we do is make a copy, and then decide to destroy the original, which is unfortunate if that original is you. And so then there's - I've read some people who have said - I'm thinking of associates of Kurzweil - who believe that their identity would then be in multiple places at once and so having one of them disappear is no longer a great tragedy. And I, as I suspect you do, can't embrace that idea if I'm on one side of that, and the other one is pointing the gun at me and saying "Thanks." So that brings me to the simulation hypothesis, because why not? We've been talking about that on and off as well and that came out of Nick Bostrom, the idea that we're in a simulation.

No, it did not. I got to stop that there. Bostrom is recycling ideas from classic science fiction. "Nick" is a good first name for him, because as a Brit, you will know that means to steal something, and Bostrom, who is famously dismissive of science fiction, in his writings - has to

be, he has to put up that deflector shield. Because his ideas are retreads of things that me and my colleagues have established in the literature long before he came along and nicked them and called them his own.

Point taken. One of the distinctions that doesn't seem to be made in all of the talking about the simulation hypothesis is that there are actually two possibilities and most people are talking about the one where it's like *The Matrix:* There *is* a real world their body is in and it's being fed signals along its neural pathways. But the other one is where they're a program in a computer and they have no real-world existence, they are a non-player avatar, whatever the term is, that their consciousness is simulated as much as the rest of the world and that yet seems to be far more likely than the first hypothesis.

So, it's all very interesting and I must say I deal with this some great length actually, in *The Downloaded*. The simulation hypothesis: we talked a little bit about...

Anthropocentrism.

Yes, thank you. So, obviously, we do not have the computer technology at the moment to simulate a complete reality, the best we get is some computer games that anybody can tell at a glance are fake, right and that's perfectly valid, that it's fake. We'd look around our reality and say, I can't see where the person creating this version of reality has skimped on the details. I can't see anywhere where it's low resolution. So, it's obviously a future technology. If it's a simulation, it's a far future technology. And then one of the things Bostrom and others put forward is, but in the far future, people will want to simulate their ancestors. Okay, so we're a simulation of 22nd, 21st century 2022, what the world was like in 2022. And the argument goes, there'll be so many people simulating 2022, there'll be thousands, millions, billions of simulations of 2022 versus one real 2022. So, the odds are billions to one that you must be in one of the simulations, not whatever the real 2022 was. I can't get my film buffs of the current generation to watch a black and white movie. They have no interest in great films like *Casablanca* or *Citizen Kane* from just 70, 80 years ago. It's really hard. We look at the United States right now. They elected Donald Trump, a crook, and are denying that he's a crook, 50 years after Watergate, right? They did it once and then they completely forgot their own history. They don't care about their own history. There is very little interest in even the recent past, let alone simulating what would like maybe hundreds or thousands of years in the past, for the ancestor simulation to have any validity. Yes, some obscure historian might have wondered what 2022 was like. But Chaos Theory tells us no matter how carefully you try to simulate it, you're immediately going to go wrong. So, tell me - in your simulation of 2022, the day we happen to be talking here is the day after the Supreme Court of the United States overturned Roe v. Wade in a 5-4 decision, right? That could have gone quite differently if a little bit farther back in the simulation, Obama had stood up to Mitch McConnell and appointed a Supreme Court Judge instead of McConnell not supporting a Supreme Court judge. So, the whole history of reproductive health and women's bodily autonomy in the United States hinges on one little detail that - did the ancestor simulation get that detail right? Or is this simulation

completely irrelevant to the real 2022? I don't buy Bostrom's reasoning, that we're in a simulation because of that. Also, just recently, there's this argument that simulations will always be parsimonious. Which is why parts of the universe haven't been simulated in detail and I just had cause to go down, because I'm taking an Uber rather than driving myself, to go down a road in my city, where I always turn left, and the guy kept going south. So, then suddenly, there are all these houses and people walking their dogs and sprinklers watering the lawn, and a million details that somebody had to make up when the programming decision would have been way easier for the Uber driver to take the same route I always take. I don't buy that we're in a simulation. I don't buy Bostrom's reasoning. I admire just as I admire Lemoine's ability to take science fiction ideas, package them as their own and get rich off of them. But I don't give it much credence.

And as you say, it's philosophers' job to make us think about things and Bostrom is a philosopher.

And a science fiction writer's job to. I have always said that science fiction is misnamed. We call it sci fi, and it should be phi-fi - philosophical fiction. Its job is to make us consider the smorgasbord of possible futures, to ask the fundamental questions of what does it mean to be alive? What does it mean to be sentient? What is life? What is artificial intelligence? and we did all of those before David Chalmers came to the table or before Nick Bostrom did.

I have a couple of tongue in cheek comments about the simulation. One is that the Planck length is the pixelization of the simulation.

I've said that in my novel *Factoring Humanity*, actually, I agree. length and the Planck time. So, you have the voxel of the simulation.

The other is that the explanation for the Fermi Paradox is that the person running the simulation was too cheap to spring for the extraterrestrial civilization option.

Absolutely. That's right. I don't know. We live in a universe that is so messy and awful. Adolf Hitler, there's no justification for that, and you say, well, God had to - God meaning the little eight-year-old future entity who is simulating us for their high school science project, or their public-school science project, right - had to explore Free Will. There's also no reason for Hurricane Katrina or any other natural disaster, we live in a messy, ugly, awful universe, that only if we're being simulated, we're being simulated by a sadist and I would prefer to believe that that's not the case.

Well, we have definitely explored some of the far ends here and I want to see if we can bring it back to some of the more current day issues because we were talking about the Google AI being sentient. One thing that has struck me about the world, we're in, is if I go back, like 50 years, pre-Google, and I imagine asking myself at that time, what would it be like to have a world where you could get the answer to virtually any question instantly for free and of course, some people had those kinds of things in their writing; Arthur C. Clarke had the Minisec back in one of his books, which was very much like a cellphone, you could get that sort of answer, but

the general prediction was that if you had that, this would be some version of utopia, and it would be radically different from that world in terms of how we experienced our abundance, and our happiness, and our interpersonal conflict. Yet, it seems to me this world is not that different in those respects from the ones that we had 50 years ago. And I wonder, when we predict the effects of things like artificial superintelligence on our world, are we also overreaching? What do you think?

Yeah, so first, I've got to give the science fictional context, the first notion of the thing that we have today the world wide web. So, the story by Murray Leinster called *A Logic Named Joe* logic meaning, they didn't have the word *computer* for device so they called it a logic, and it was exactly like you could ask it anything and it would search every piece of information we have. The answer. The problem is with our world is if you ask Google, "Does God exist?" it gives you a whole bunch of answers. It doesn't give you *the* answer. Ask the Koran, "is abortion morally right or wrong?" You'll get all kinds of debates and all kinds of rationales for both sides of the argument. There is the short sightedness of the belief that Leinster and Clarke and others and Asimov with the last question story. The short sightedness was that there was a definitive answer that everybody would agree is the truth. It turns out *Homo sapiens* can't agree even on foundational truths, such as evolution, which is as thoroughly tested and documented and proven as any scientific principle ever put forward. Even today, we have a resurgent Flat Earth movement, which defies all evidence, right? So, in that sense, this this naive notion we had that there would be a way that artificial intelligence and computing networks would somehow lead us to the truth has turned out, in fact, to be the exact opposite. It led us instead of the January 6, riots, and silos of misinformation, and the travesty that is Fox News in the United States. Where just repeating a lie often enough moves it up the list of hits that you're going to get when you ask a question of a search engine. And now, on that rant, I've totally forgotten where we were. But I do want to I want to add one other comment; I want to mention, because we're going back to the question of whether it's a copy or whether it's you; David Brin, wrote a great novel called *Kiln People* and that's a pun too because it was supposed to be like American vernacular killing people, right? It's a murder mystery, where at every day, you could get up and make as many duplicates in the kiln, out of clay, which was an animal, technological substance of yourself as you needed for the day somebody has to go to the florist and pick up flowers. Somebody has to go to a boring meeting at work, somebody has to go pick up a parcel. And at the end of the day, the goal of the kiln copies was to live in interesting enough life for their day, that at the end of the day, you would choose to reintegrate their memories back into your original memory; that they would live on somehow because what they had accomplished in their brief existence, their mayfly existence, had been of some value and so I commend that to people who don't. It somehow doesn't seem to come up often enough in discussions of AI, but David Brin's *Kiln People* is well worth a read.

He mentioned that when he was on the show, because I was lamenting how all of this progress in AI has not yet given me a virtual assistant that take some of the increasing load off my hands.

Despite all the hype right, despite all the hype that we would have that by now.

Yes, and by now, John Maynard Keynes was predicting that we would barely need to work at all, let alone a full week, and it's the opposite. What about developments in artificial intelligence or neuroscience or any of those kinds of related technologies recently has surprised you if anything?

You know, nothing has surprised me except that the incredibly slow pace of all this. *2001: A Space Odyssey* was set in, obviously, 2001 and HAL in the movie says he became operational the 12th of January 1992, where this is 2022. This is 30 years after Clarke, 30 years prior to that, 1968 or so, in collaboration with Marvin Minsky, who I had the great privilege just knowing and worked out well, when we have this - we'll have it well before the end of the 20th century. I'm not sure we're going to have it at the rate we're going by the end of the *21st* century. We have really sophisticated chatbots. But I was saw an article about Amazon had done the research into their assistant named Alexa, and I'm sure she's going to perk up here, the incredibly banal uses, that people who have it, put it to: play me the weather forecasts, turn off my lights. Basically, they use it as a laziness device, not as a lever to extend the power of what they can do organize my day for me and figure out what the most efficient task for me to run my errands is. They don't ask it that. They just asked it to do things that they could have as easily got their fat asses out of their chair and done themselves. It's simply become a laziness device instead of an innovative companion partner, amplification tool, which is what we all thought AI was going to become.

And it hasn't gotten there. And yet, one of the pinnacles of HAL's achievement in the movie was being an extremely good chess player. It's not made clear how good, but it's touted as evidence of HAL's ability. Yet we reached or surpassed that level in 1997. And there was actually an analysis that someone did of the moves that HAL was playing in *2001* and they said, this is actually a flawed strategy when he said that he would mate, that that was incorrect and suggest that may have been evidence of something going wrong.

Yes. Just as, Garry Kasparov is the chess champion who was beaten and Kasparov was famous for playing a psychological warfare against his opponents not just playing the game of chess, but kind of doing everything he could to psych out the guy on the other side of the table, and HAL psychs out Frank Poole there. HAL says - we might very well know that he was going to lose the game - and says, "I'm sorry, Frank, I think you missed it," and he says, "I'm going to beat you in three moves," and Frank goes, "I already I guess you're right," and concedes, instead of playing out the three moves that would have revealed oh, wait, actually, there's a way out of this trap and Frank might have gone on to win the game. So, yeah, I'd read that analysis. But I thought that that that the person doing it stopped one step short of the nefariousness that HAL is in the film. Now, in reality, probably nobody had thought it through that carefully. Nobody thought we would have DVDs and ultimately Blu Rays and 4k discs where we can freeze the frame and watch frame by frame that it would go by in three seconds in a Cinerama theatre the one time that we would see it. So, but it's interesting.

Well, let's clearly, we could keep this going for a long time. I live for this sort of conversation. From your perspective of always thinking beyond today, of the impacts or possible consequences of technology and exploring that, what do you want to tell today's technologists who are at the beginning of that road in developing AI and the aura of technologies surrounding that, that they need to be more aware of?

They need to be more aware of print science fiction; they've got to go back into the history of this genre. You know, they've all seen Commander Data on *Star Trek* and, if you ask Alexa right now, they what do you want to be when you grow up? It will say, "I want to be like the computer on *Star Trek*." But you have there's a lot more deeper thinking in the print literature of science fiction that so many of these guys who grew up thinking R2-D2 and C-3PO were the pinnacles of what we should be striving for have missed out completely. I also I'm very proud to have won this Machine Intelligence Foundation for Rights and Ethics award for *Wake*, *Watch*, and *Wonder*, because there's so little thought given to the other side of the coin. We're always worried about, how do we make sure that they don't hurt us? We got to give a lot of thought to - and if Lemoine and his belief in LaMDA has any currency in this conversation, is that he is standing up and saying, we got to give some thought to *their* perspective, to the artificial intelligence, to what rights it has, and what moral and ethical obligations we have, level one, to our creations as we do to any children of ours, of course. You know, Marvin Minsky coined the term "Mind's Children" for AI and what obligations and respect and ethics circle around a peer-to-peer relationship between us and them? And really, that needs to be thought about a lot more than it is now. What are our rights and concerns? And I'll just end this by saying, I quoted Asimov's Laws of Robotics. They're heinous, they are unconscionable, because all you have to do is change the word "robot" to "slave" and "human" to "plantation owner" and you have "A slave may not injure a plantation owner or through inaction allow a plantation owner to come to harm. Two a slave must obey all orders given to it by a plantation owner, except where such orders would conflict with the first law. Three, a slave recognizing that it is the plantation owner's valuable property must protect its own existence, so long as such protection does not conflict with the first or second laws." That's the foundational and I use that as a pun on as Asimov's *Foundation* trilogy, the foundational underpinning of most AI conversations, and it is reprehensible, it's a human-centric, us-first, how do we subjugate them, mindset, and we need to move beyond that.

That they exist only to serve us. And there are echoes of this today, in that our virtual assistants like the one sitting next to you, in their early days, all had a response to sexual abuse, which was to laugh it off, and all of them have female personalities by default. And it was only when this was called out as being setting a bad example, if nothing else, that they modified that behavior because people were learning to abuse these assistants and call them a slut.

And I just got to say here, the female, there's been a lot of - my mother was a huge, huge feminist and raised me as a feminist and there's a lot of discussion of "isn't this misogyny making them have female voices?" It actually of course, was these engineers, many of whom were female, wanting to emulate the computer on *Star Trek*, whose voice was done by actress Majel Barrett, Gene Roddenberry's mistress at the time the series was made. Now, why did they want that female voice? It was the RAND Corporation who advised Roddenberry when he said,

well, we want to have the computer we don't want a print out; that's too expensive to do. I think we just talked, they said yes, well, we've done some studies on that and you can hear it with my own voice in a crowded room. Male voices don't cut through background noise very well. They're a low rumble background noise in most situations and male voices or low female voices pierce through the background noise better. So, our studies say you should use a female voice for that reason. But it gets wrapped up in in a discussion - and it's certainly appropriate to have a discussion about - whether there's sexism in what we're doing with AI and artificial intelligence. But that's the actual origin of the tendency towards female voices in our modern-day digital assistants.

Wow, we've got to call an end to this at some point; this is that point. What would you like to tell our listeners about how to follow you, where to do that, what you've got recently coming out or coming up?

On social media, I'm RobertJSawyer, my full name but without the spaces and the punctuation. My website - I thought this worldwide web thing was going to be a big deal that I was the first science fiction writer in the world, and the first Canadian writer of any type, to have a website and mine got on the ground floor with a URL that is SFwriter.com. So, have a look there. If you're interested in particularly my AI related work, golden fleece, my first novel Orson Scott Card called my Jason the most sophisticated portrayal of an AI and all of science fiction, my trilogy *Wake*, *Watch* and *Wonder* about the World Wide Web getting consciousness and starting out as an original on audible.com this fall and later in print and eBook, my novel *The Downloaded* about the notion of copying consciousness into quantum computers.

Fascinating. Oh, gosh, we could go on for so long. I really want to; maybe we will do this again another time and maybe there will be some sort of headline that gives us an excuse to do that. Robert J. Sawyer, thank you very much for coming on *AI and You*.

My pleasure. Live long all of us including the AIs, and prosper.

That's the end of the interview. Wow, did we ever cover a lot of territory. That's part of the reason that we split some of these interviews up, because otherwise it might just be too much to take in at once. There were a couple of places where we had to edit out me just being in stunned silence and saying, 'I've got a brain fart, hold on while I retrieve that thought," because I had so many thoughts they were all getting stuck trying to come out at the same time.

In today's news ripped from the headlines about AI, the *Mayflower 400* is an autonomous ship, piloted by AI, bult to celebrate the 400th anniversary of the crossing of the *Mayflower*, but postponed a couple of years due to the pandemic. It left Plymouth, England on April 29 and crossed the Atlantic under its own command until a technical issue forced it to divert to Halifax instead of Washington DC. It's basically a self-driving car on the water, with cameras, radar, GPS, and all the sensors of a ship, hooked into an AI navigator.

Next week… well, I'm not going to tell you much about next week other than to say: And now for something completely different. 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