Zombies - Can robots be conscious?

Transcript of the Human-Robot Interaction Podcast Episode 23. Published on 28 September 2022 by Christoph Bartneck at:

https://www.human-robot-interaction.org/2022/09/28/zombies/

[00:00:00] **Christoph:** Are robots, zombies? This might seem like a strange question at first, but it leads to one of the most important questions in science today: what is consciousness and can robot's become conscious? These questions fascinate many people. And when Blake Lemoine suggested that Google's latest AI Lambda had become sentient, it triggered a worldwide media frenzy. In this HRI podcast episode, I talk with Jack Copeland about machine consciousness.

[00:00:57] **Christoph:** Jack, what do you do at the university?

[00:01:00] **Jack:** I teach philosophy of computing, philosophy of mind, logic, philosophical logic and a few bits of history of philosophy.

[00:01:08] **Christoph:** What is consciousness?

[00:01:09] **Jack:** A very good question, Christoph. If only we knew the answer. To spoil the plot: we don't know. It's too early in the trajectory of science to have an answer to that question. Just as how it was too early to have an answer to the question, what is life 500 years ago or a thousand years ago. We have to wait for some, significant scientific progress before we can give the answer to that question. But nevertheless, there are some very interesting things that we can say about it, right here and now and a preliminary observation is that it's an ambiguous question.

[00:01:43] What is consciousness. Simply because the word conscious is ambiguous in English. It has different meanings, like the word bank, it could mean at least two different things. And with consciousness, if you say that some creature is conscious. You may just be saying it's awake and functioning, there's the conscious unconscious distinction and, somebody is unconscious because they're asleep and then they're alarm clock goes and they're conscious, they're awake and functioning. So that's one sense of conscious, but I suspect that wasn't what you were asking about.

[00:02:14] **Christoph:** No, I think I've got the more complicated concept in mind.

[00:02:19] **Jack:** There's a range actually of more complicated concepts. One that we all talk about all the time is self-conscious. If you say something is conscious, you may be meaning that it's self-conscious. That not only observes and understands the world, but it also observes and understands itself.

[00:02:37] And obviously humans are self-conscious and maybe some other primates. Perhaps even dogs and cats, but, as you go further down the tree bats or swans or raccoons or something, they may well be conscious without being self-conscious. They may perceive the world without kind of perceiving themselves as agents interacting with the world.

[00:03:00] Self consciousness, consciousness in that sense kind of Peters out as you get further and further away from human beings. But maybe that wasn't what you were asking about either.

[00:03:10] **Christoph:** No, I suspect there's an even deeper meaning.

[00:03:14] **Jack:** When the question is, is it possible to build an artificial intelligence, a robot say that's conscious? I think probably when people ask that question, they don't mean either of the sences that we've just talked about. Because obviously a robot could easily be described as being asleep or being, fully awake and fully functioning. Maybe when it plugs itself into its niche to regenerate overnight, it powers down most of its perceptual and cognitive system.

[00:03:42] There's some analog of sleeping. And then, when it boots itself up again in the morning, it's fully functioning. So it is now conscious. It would be easy enough to imagine a future, not so distant in which there are robots that behave in that way.

[00:03:55] In one sense they would be conscious, but that's this first sense of consciousness that we talked about. The much more interesting sense of conscious, which is much harder to see how an artificially intelligent robot could be conscious in this sense is having sensations.

[00:04:12] That's the third sense of consciousness and it's different from the other two. And we are all conscious in that sense. We know from the inside the phenomenon of consciousness in this third sense. If you just look up into a clear blue sky there's that unmistakable sensation the perceiving of blueness. Or if you hold your hand over a candle flame, then you get that mild feeling of pain that, we all know so well. And it's in the third sense, it's the ability to have feelings, sensations like that, that constitutes consciousness. Psychologists and philosophers have a sort of fancy technical term for these sensations.

[00:04:53] They call them Qualia. It's a Latin word, I think And so you can talk about the Qualia of pain and the Qualia of blueness. There can be smell Qualia and taste Qualia. So consciousness in this sense is to have Qualia. You might well ask, could a robot be designed to have Qualia. So I suspect that when people ask whether an AI can be conscious, it's this third sense that they've usually got in mind.

[00:05:20] **Christoph:** By coming back to the question about consciousness itself. You mentioned there's this strange situation where we all intuitively know what we mean but we still can't define it.

[00:05:34] **Jack:** People get a bit hung up on definitions. We can't define most things. Definitions are a luxury. Some words you can define perfectly well. An example that often gets used in the literature is the concept of being a bachelor.

[00:05:47] You can give a nice crisp definition of that. It's an unmarried person of marriageable age. But definitions like that, crisp sharp definitions, are quite hard to come by. For example, try and come up with a definition of a game. This is a well known philosophical challenge that goes back to a guy called Wittgenstein.

[00:06:07] Back in the middle of the 20th century. He was making the point that, I'm repeating now that definitions are a bit of a luxury, and there are many words that we understand perfectly well that you can't define in a nice crisp way. And his favorite example was "game". And it's a challenge, you think you can define the word game, then come up with a definition that catches all and only those things that we're naturally inclined to call games. And I think you'll probably find you can't do it. I wouldn't worry too much about defining consciousness. It's no doubt one of those very many extremely useful terms that actually we can't define.

[00:06:44] But stepping aside from the sort of definition issue, we all know what consciousness in the third of the senses that I staked out. Because we all have sensations. When you have your lunch, you have, a zillion sensations. There's the taste of the soup and the taste of the coffee and the sound of the cutlery chinking against your plate.

[00:07:03] The waiter's voice. And the brightness and the feeling of warmth from the sun and so on. We all know this stuff intimately. It's just every day and every day accompaniment everything we do. So they're Qualia, they're conscious experiences, we know what they are.

[00:07:18] We don't have a scientific theory of them. To take an analogy, if you think of fire and you go back kind of 600 years. They all knew what fire was perfectly well. But they didn't have a scientific theory of what fire was. In that sense, they didn't know what fire was, they could point to fires and talk about fires and distinguish between fires and non fires and describe what fire is useful for and think up new uses for fire and so on.

[00:07:47] They could do all of that. But in some sense, they didn't know what fire was because they didn't have a scientific theory of fire. The theory of gases came along later and by the time of kind of Isaac Newton, they knew very well what fire was. Consciousness is like that we know what it is from the inside, but we don't have a scientific theory of it yet.

[00:08:09] **Christoph:** Is consciousness intelligence.

[00:08:12] **Jack:** No, no intelligence and consciousness are different. You can imagine a robot that is intelligent maybe marketed as an intelligent robot and it maybe people take them home and they work as a Butler, they do the cleaning and do the shopping and entertain the kids and, et cetera, et cetera.

[00:08:30] So they're intelligent. But you can imagine that the lights not on inside, they don't actually have sensations. They're not aware of the world in the way that we are. These imaginary robots, they have sensors, they can find their way around in the world.

[00:08:46] They can detect sounds and respond to you, telling it to cook a Curry or something. They don't bump into the walls. They've got laser sensors that enable them to navigate around and perform very delicate operations.

[00:09:00] In some sense they can definitely sense the world. But for these robots that I'm imagining what's happening is that kind of light flows into their optical sensors and they process it and pressure information flows from their tactile sensors and they process it. It's all just number crunching in the end.

[00:09:19] These robots that I'm imagining don't actually have sensations in the way that we do. They don't have Qualia to bring back that term I used earlier. So they are intelligent. Maybe they can even prove novel theorems in mathematics and so on. Maybe one of these robots proves a theorum, that's been puzzling mathematicians for centuries.

[00:09:38] You definitely wanna say they're intelligent, but it's another question, whether they're conscious. It's a completely different question. Do they have these Qualia, these sensations that we have and are so familiar with? Maybe not. You can easily imagine a robot that is intelligent and, maybe even behaves indistinguishably from a human being, but nevertheless doesn't have Qualia.

[00:10:02] **Christoph:** And what are the current most promising models for consciousness?

[00:10:07] **Jack:** Again, you've got to be aware of the ambiguity of consciousness. If you pick the first or the second sense that I mentioned awake versus asleep and self-conscious, self-aware, there are certainly models of these two.

[00:10:21] As I was saying, it's the third one, that's the interesting one. So I guess maybe you're asking what are the most promising models for the having of Qualia. And the answer is we don't really have any models yet. It's too early in the science of this subject, even to have models.

[00:10:36] We are really pretty ignorant of what consciousness in this third sense. In, in the scientific sense and, we utterly lack a theory of it. There just aren't any models. We don't we don't even know to what extent consciousness is physical in this third sense.

[00:10:53] People divide on whether Qualia are some kind of non-physical process, which is caused by physical processes. There are other physical processes going on in the brain. Currents moving around and ions passing across synapses and so on.

[00:11:10] According to some people, all this kind of physical chemical activity in the brain and in the neurons of the brain causes Qualia. But the Qualia themselves are non-physical, they're totally different from the physical goings on, in the brain they're caused by the physical goings on.

[00:11:26] But they're different from the physical goings on. So that's one, you can't really even call it a theory cause it's not well developed enough. It's a guess. There's one guess as to how consciousness might be. And other people say and again, they're just guessing they say no we live in a physical.

[00:11:44] Everything's physical, isn't it? So consciousness must be physical. It's not just that consciousness is caused by physical processes in the brain. But consciousness must itself be a physical process in the brain. There are two guesses as to what consciousness is. We just don't know which one is right at the moment.

[00:12:03] As science progress progresses, I'm sure that we will get to know which of those sketches is correct. And we'll get to know a lot more scientific detail about the nature of Qualia. But right now we don't know it and we might have to work quite a time conceivably. Maybe even 500 years or a thousand years or something before there is a viable, scientific model of consciousness.

[00:12:27] It's a bit like the question, what is matter? The ancient Greeks had definitely thought of that question and were very interested in it and speculated wildly about the nature of matter. But they were just guessing. It was several thousand years later before science got really worked out as to what matter is. The atomic theory and kind of subatomic particles and all of that people were speculating about what matter is thousands of years before the answers arrived.

[00:12:56] So I suspect it may well be not dissimilar in the case of consciousness. Right now we are like the ancient Greeks asking what is matter. We're asking what is consciousness and it may well be hundreds or even thousands of years before science is able to deliver an answer to that question.

[00:13:14]

[00:13:24] **Christoph:** I assume that you've been working and thinking about this problem for quite some time. Is it not frustrating that we have this lack of understanding?

[00:13:37] **Jack:** No I don't find it frustrating. I'd love to know the answer. Obviously. I think anyone who's thought about consciousness desires to know the answer. That's why you think about it in the first place. But frustrating. No, science is a slow business. The way science progresses is that somebody thinks up a new question, a really difficult question that kind of nobody's really thought about before and somehow answering that question. Sort of like a quantum leap forwards in science.

[00:14:05] It takes quite a while between the question being posed and the answer being produced. It could well be the same with consciousness. People have been asking the question for a long time and we're not really that much closer to answering it. In the end, the answers come and quite often answers to these hard questions, transform science in the process. I suspect it may be the same with consciousness.

[00:14:33] **Christoph:** Do you think machines can be conscious?

[00:14:36] **Jack:** I kind of think humans are machines, aren't they? We're biological machines, we're very complicated, we're delightful machines. We're not the sort of, doll hard machines that you find in factories and so on. We're very interesting soft machines, but machines, nevertheless and we are conscious. So there's, that seems a quick answer to the question. Can a machine be conscious if you think humans are machines, and why wouldn't you, then yes, machines can be conscious.

[00:15:05] **Christoph:** Is machine consciousness the same as human consciousness?

[00:15:09] **Jack:** That's getting back to the ambiguity of the English word consciousness again. You tell me which of the meanings of consciousness you are inquiring about. Then I might be able to say more about your question, but if you're saying is consciousness in the sense of having Qualia the same for humans as for machines, then the answer's got to be yes.

[00:15:31] Humans have Qualia and the question is do machines have Qualia or not? So it's the same phenomenon in each case. We know that humans have Qualia and the question is, could you build a robot that has Qualia? So it's certainly not a different sense of consciousness it's having Qualia,

[00:15:49] **Christoph:** I guess my question goes in the direction of, does a machine have to work in the exact same way as a human does to achieve consciousness or could a machine be a black box that somehow works differently and would still be conscious?

[00:16:08] **Jack:** We know so little about consciousness that you can only speculate about the answers to questions like those. What physical processes produce consciousness in human beings? We don't really know. We assume it's stuff that goes on in the brain on the nervous system. We assume that it's that consciousness is somehow the outcome of neural activity. Electrochemical activity. We assume that.

[00:16:33] Is the having of Qualia kind of bound up with having the sort of neurochemistry that we do or could an entity that has a completely different chemistry also have Qualia. Or maybe Qualia could be produced by a process that's sort of independent of chemistry. Maybe just running an algorithm, maybe just doing computations would be another way of producing Qualia. In ours, it's presumably produced by chemical changes but in some other kind of entity could qu be produced simply by carrying out computations, no matter what the device that carries out those computations is made of. It could be, made of kind of modern hardware or it could be made of 1950s vacuum tubes, completely different physics and chemistry.

[00:17:22] So long as it's doing the same computations then maybe it's conscious. No matter what it's underlying physics and chemistry is. Could be, nobody knows. There's two different views of the origination of Qualia. Qualia are originated by neurochemical processes specific to the kind of chemistry and biology that we have versus this other view of the origin of Qualia they're brought about by computations where the computations are the same, no matter what physics or chemistry of the underlying device that's producing them. So it's two different theories of the Genesis of consciousness. Nobody knows which one's true. But they're both extremely interesting theories with their own implications.

[00:18:07] **Christoph:** How can we tell if a person is conscious?

[00:18:10] **Jack:** I just look at you and you smile and you've got intelligent eyes and it wouldn't normally occur to me to think that you were not conscious. If you press the point or we've got a colleague. And we're genuinely not sure whether they're conscious, maybe they had a, some strange brain surgery or something, and they're a little bit changed when they come back to the campus.

[00:18:32] There is a conversation about whether this person still has Qualia. They're still highly functional and can give their lectures and so on. But maybe as a result of their strange brain surgery, they've lost the ability to have Maybe this is a real topic of conversation.

[00:18:48] How could we settle the issue of whether this colleague really has qualia? Difficult? If you're really serious about does person X have Qualia, how could we test? If I had a colleague in this situation I would feel I wanted to give them the benefit of the doubt, if they still behaved after the surgery and much the same way as they did before the surgery. I would as assume that they're still conscious, I don't wouldn't know how could I know because consciousness is this inner experience.

[00:19:18] I can't have any direct contact with their Qualia. Their Qualia are private to them. If their behavior and all the rest of it remained the same after the surgery, I would just assume that they could still feel pain and feel pleasure.

[00:19:33] I avoid doing things, which before the surgery would've caused them pain and I would seek to do things generally, which kind of could be expected to make them feel happy and so on. I couldn't know whether they were really feeling pain or not. But I would put the benefit of the doubt.

[00:19:50] This question that you've asked how could you tell whether whether something is conscious? It's a very interesting question. In the first sense of conscious, like how could you tell whether something is awake or asleep? It's not too difficult to lay down some specific criteria for when something is awake and fully functioning. But with the third sense how do you tell whether something is experiencing qually or not how do you tell whether the light is on inside or not to put it figuratively?

[00:20:17] One way to get a kind of sense of the difficulty of the question is to switch from kind of human beings or artificially intelligent robots. Step aside from the question of whether they can be conscious and just think about other species, maybe ones less complicated than human beings are flies conscious do flies have Qualia? What about frogs? What about earthworms and so on? It's very hard to imagine any way of establishing whether an earthworm has qu or not, they do behavior. They'll move towards food. They'll recoil from noxious stimuli, like heat or an electric shock and so on, but do they really feel pain or do they just react to the incoming signals?

[00:21:03] And suppose you got really interested in the question of whether earth worms can feel pain or whether they just react. What tests could you possibly do to establish whether an earth worm feels pain or not? It's very difficult to specify any test that would leave you, any of the wires are on this question.

[00:21:23] What about fish? The fish feel pain. They certainly behave in very complicated ways and, they'll seek food and avoid noxious stimuli and they do group behavior and so on. They'll flock together and, gigantic crowds in order to simulate larger fish in order to frighten off predators. Very complex, intelligent, you might say behavior but to fish actually feel pain or do they just avoid noxious stimuli?

[00:21:50] How could you tell what tests could you do? It's very, a very difficult question. So it's gonna be the same with robots. Once complex socially interacting robots with near human levels of intelligence are here. We will be wondering whether they're really conscious or whether there's just all this complex behavior, but maybe the light isn't on inside. They don't have Qualia, even though their sensors enable them to get round in the world just as well as we do.

[00:22:22] **Christoph:** Are any current machines conscious?

[00:22:25] **Jack:** Yeah, you! I think you're conscious. I think you're a machine.

[00:22:32] **Christoph:** Let me try to rephrase that. Is there any computer that is currently conscious?

[00:22:37] **Jack:** I certainly don't think so. No. If you really push the point and insist that your university of Canterbury laptop is conscious what tests could we do to show who was right? Well, no, that, we're just not at the stage of being able to specify tests that would tell you whether your laptop or anything else is having Qualia or not.

[00:22:57] It's the same with like stones. If you insisted that your favorite stone that you carry about in your pocket and call George, if you insisted that George was conscious what tests could we do to show that you were wrong or that you were right? There just aren't any, but of course there's no reason to think that the stone in your pocket is conscious.

[00:23:16] Nothing goes on in it that's anything like the processes in our brain that we believe causes consciousness. And it's the same with your laptop. There are no processes in your laptop that even distantly resembled the neurophysiological chemical stuff that goes on in ours and that we think is causally responsible for consciousness.

[00:23:40] **Christoph:** Is Google's Lambda conscious.

[00:23:43] **Jack:** No. Has anybody ever really said that it is? Not even Lemoine. I've read an interview with him. I did some prep for the interview. He says is Lambda sentient. He says, we can't answer that question definitely at this point, but it's a question to take seriously. He was wildly misreported in the press. He did not say it was conscious or sentient. He just raised the question and pointed out that we can't answer it.

[00:24:12] **Christoph:** That is a very interesting point. Thank you for this preparation, because I read all these news stories. I was amazed about how much attention this question got. Everybody was so fired up about, look, we've got a machine consciousness and that seems to trigger something in people. Why is it such a big trigger for people?

[00:24:29] **Jack:** Well, it's such an interesting question. We tend to think of consciousness as like distinctive of human beings or defining of human beings. We're maybe prepared to allow that chimps are conscious as well or gorillas and so on, but it is kind consciousness is something very special to ours and our close relatives. And so we get very excited about the idea that we may be able to recreate this very special thing in machines. It almost everyone is, extremely interested in the question of consciousness. I find.

[00:25:03] **Christoph:** What problems do machines have that inhibit them to become conscious?

[00:25:10] **Jack:** I don't know the answer to that question. It's just too early to be able to even sketch out an answer to that question. The sort of answer that you might give depends on which way you jump on open questions. Like some of the ones that we've been talking about. There are these sort of two stances towards consciousness, and one is that consciousness is caused by specific physical chemical processes.

[00:25:35] And the other is that consciousness is just the outcome of computation and the physics and chemistry of the underlying device just don't matter so long as you've got the right computations. So if you take that second stance towards current computers then what's standing between consciousness and current computers, what's the blockage that they need to get through in order to In order to have Qualia. Well, they just need the right algorithms.

[00:26:00] They've just got to be able to do the right computations. The other stance where you know, chemistry and physics matter to the generation of consciousness then the computers we have today look like they're just completely the wrong sorts of things to be conscious because they don't have the same kind of chemistry that we do.

[00:26:23] **Christoph:** Do you think we will ever understand consciousness?

[00:26:28] **Jack:** Oh yes. Why not? It's just another of those delightful scientific questions that happen to be really, and interestingly hard. We might have to wait a while, as I said before, we know the answer. It's a scientific question about the world and if the human race survives long enough, I'm sure we'll answer it.

[00:26:50] **Christoph:** Do you think machines will ever become conscious?

[00:26:53] **Jack:** Yeah. Your kids, your children's children. All machines, they probably all be conscious. Would you like to rephrase your question? That's probably not what we meant.

[00:27:03] **Christoph:** This is, again, the assumption that humans are in a way a machine and not everybody would agree to it. Some people would argue that there is something about the human mind that is not reducible to physics. It's not just limited to a physical phenomena. And those people would have, a different answer because they would not accept this premise that we are machines.

[00:27:29] **Jack:** We could argue about this. It is a separate issue. What is a machine and are humans machines. and of course machines can cause things beyond themselves.

[00:27:39] **Christoph:** Let me try another angle. Some people argue that if our brain would be simple enough for us to understand we would be too stupid. And I imagine that you've got a response for that.

[00:27:53] Jack: When you first hear it, it seems it seems funny and, it's nice because it's self deprecating and also there's some kind of infinite loop or something in it. So it's an interesting kind of thing to hear. But when you think about it, was is it supposed to show? The kind of basic idea that it's getting at is that if a understands B then a needs to be more complicated than B. If a is less complicated than B, then how can a understand B. But I just don't buy this idea that a has got to be more complicated than B in order to understand B. There's parallel phenomenon. Alan Turing the guy who, invented the fundamental ideas of the modern computer. He proved a theorum back in 1936, which is it's a very counterintuitive theorum it's something that you wouldn't expect at all. And it sounds wrong when you first hear it.

[00:28:48] He talks about these sort of idealized computers that we now call Turing machines simply. He called them computing machines. They're abstract devices. But he proved that Once a tiering machine reaches a certain level of complexity and he was able to specify, very precisely what that level of complexity was. Then it can behave like any Turing machine. It can simulate the behavior or understand if you like any Turing machine of much greater complexity than it. It's once you reach a certain critical threshold of complexity, then that gives you all that you need for that machine to understand something of much greater complexity.

[00:29:30] And he had a theorum, as I say, which which said essentially that and it's the same with understanding our own brain or other complicated systems. It's like once you reach a certain level of complexity, then you can understand systems of far greater complexity than you.

[00:29:49] Once you've got enough smarts on board, once you reach that critical level, then given enough memory and given enough time and given enough data, then you can understand everything, anything. Maybe humans have already reached in natural evolution that, that kind of critical threshold for being able to understand any system, no matter how complex. Given enough time and enough memory and enough data.

[00:30:16]

[00:30:26] Christoph: How can you be so optimistic about the progress of science when we struggle with this question for thousands of years, and AI has disappointed most of its predictions?

[00:30:39] Jack: There are two very interesting claims in that question. One is that AI is disappointed as in most of its predictions. I don't think that's so. There have been these crazy predictions that people in AI have made from time to time. And sometimes they do it just to get attention. People like Ray Kurzweil, for example, to mention one name. with his fascinating and brilliantly written books that make all these claims about, computers will go to church in 50 years time and stuff like that.

[00:31:09] All predictions like this will I'm sure fail to come true. And also there, there have been lots of cases where AI researchers themselves have made predictions for the purposes of getting funding. You hype things up to persuade the us army or whatever to give you money for your research.

[00:31:27] This has led to the the so-called AI winters where there's been all this hype and then none of it comes true. And so people lose their faith and the funding dries up and it kind of goes in cycles. Aside from all of that, which has had a lot of press attention but that's just the froth on the surface, if you look deeper down AI has made very steady upwards progress since about 1950.

[00:31:52] People at the work face have been making modest scientifically well evidenced predictions. That have tended by and large to come true. Maybe they've been a bit wrong about the time scale sometimes and things have taken. 50 years rather than 10 years. But I would say it's far from the case that AI has disappointed us in its predictions.

[00:32:12] I think AI has been pretty much on track. Since the subject was conceived by Turing in the 1940s. He called it machine intelligence rather than artificial intelligence. And, he was one of the first to make predictions. He was very cautious about his time scales. He talked in terms of millennia rather than decays as some of the later predictors did.

[00:32:34] There were predictions in the in the 1960s. And so on that within a generation machines would be able to do all the work that human beings could do at that time. These predictions, these sort of showy, crazy predictions are bit in the dust. The predictions closer to the workplace, they've been born out as the subjects progressed very successfully.

[00:32:58] But that was just the second half of your question. May maybe you should say the first half again in case people have forgotten right now.

[00:33:04] **Christoph:** The first one was a question about the progress of science. You mentioned that this is a scientific endeavor and it might take a long while for us to achieve it, but one could also have a more pessimistic view on it and you seem to be rather optimistic about it.

[00:33:24] **Jack:** Yeah. Science often just does take a long time with the hard questions. We've already talked about matter and how that's taken thousands of years for people to get straight about what matter is. And even now there are a lot of details still to be filled in. We've got the basic picture, atoms.

[00:33:42] We take the standard model for granted now. But it was thousands of years in the making. If you think what are the three biggest scientific questions, the three most fundamental questions about the physical universe.

[00:33:57] My big three would be what is matter? What is life and what is consciousness? And what is life? We've had a, a long struggle answering that. And even now the answer is certainly not so sharply defined as the answer to what is matter. Not very long ago, people had all these wild theories of what is life. There was the vital force that, inhabited living beings and then disappeared somewhere when they died. It's really not very long ago that people thought in these terms. But since then we've had Darwin and the theory of evolution and we've had Crick and Watson and molecular biology. And now we're pretty worked out as to what life is at any rate in the case of human beings and our near relatives. But again it's taken millennial to get there. If it takes, a few thousand years more in the case of the hardest of those three questions, what its consciousness, who should be surprised?

[00:34:54] **Christoph:** So it all comes down to our lack of patients.

[00:34:57] **Jack:** I think scientists are very patient. They want results and, in their lifetimes. But science as a whole is a very patient process. People edge very slowly. Generation by generation towards the truth.

[00:35:12] **Christoph:** Another old joke is that sciences progresses with the death of its professors?

[00:35:20] **Jack:** yes. Yes, it's probably not really so is it? It's like science progresses with the brainwaves of its professors. If you think of Newton, Einstein, Crick and Watson. It's this living beings that professors and others move science forwards. It's not in dying that the quantum leaps are made.

[00:35:42] **Christoph:** In the history of science, there's the idea of paradigm shifts and the different phases that sciences go through. And there are phases in the scientific progress where new ideas struggle, because there's a very well-established idea. And I guess the, this joke about professors dying is that at some point professors can inhibit progress because of the establishment.

[00:36:04] **Jack:** There are many sort of jokes and sayings around this idea. And one is that paradigms are never refuted. It's just that their adherents eventually die off. But I dunno, who knows, but why are you making all these, this dead professor jokes?

[00:36:22] **Christoph:** Maybe I'm too pessimistic. I don't know. These are just jokes.

[00:36:28] **Jack:** I think science is a wonderful thing and it has made enormous progress over the last few thousand years. And, it's down to the power of the human mind that's happened. And the, the mechanics of the process and the micromechanics are very interesting to study and no doubt, a small part of those micromechanics are adherents of disconfirmed theories dying off and stopping publishing articles about them and the journals. I'm sure, but it's only that sort of a fringe part of the mechanics.

[00:36:59] **Christoph:** When I did my PhD, I was fascinated by robots and I built them. I programmed them. I evaluated them. And I guess I was quite confident that this will somehow succeed. And over the decades, I see that the progress we make isn't indeed rather slow. And my optimism about the speed of progress has changed. In this podcast I made once an episode called "why do all social robots fail?" And it observes that a lot of companies start up building a robot, try it.

[00:37:33] And somehow they can't stay in the market. And after a couple of years, they all die. We made a book about robots, "human robot interaction". And by now it's a graveyard. All the pictures of all the robots in the book are dead. So I guess that's just probably part of my own personal journey about becoming hopefully more patient.

[00:37:54] **Jack:** Mm-hmm. And there are all these failed projects in AI. In fact, if you look back into the history of AI, like every project has failed. People come up with these new ideas and they build a new system, which is aimed to do, these various complicated things and it's partially successful, but then as experimentation progresses and progresses, they realize that actually it's not up to the task and the project crashes.

[00:38:20] Isn't this a bit depressing? There's no progress. It's just this long stream of failure. Well, no, that's how science is. If you think of physics, people come up with a theory and it looks good and they confront nature with it and do experiments and perhaps discover new phenomena or as a result of this investigation.

[00:38:39] But eventually the theory is refuted. It's a failed theory. And the history of science is just littered with failed theories, but that's the path to the truth. And yeah, it requires great patience and it's an intergenerational thing. We shouldn't be so greedy as to expect that, we are gonna get all the answers.

[00:38:56] It'll be, the next generation and the next generation and the next but that's how science is. And it's not gonna be any different in AI. It's gonna take a long time and every project is gonna fail until the end.

[00:39:10] **Christoph:** That means that we should be particularly careful with the PhD students that we educate. That follow us and make sure that they have a good career and hope that they will then continue the same tradition to then their students.

[00:39:28] **Jack:** Oh yes. Yes. All our hopes are always on the next generation that's for sure.

[00:39:34] **Christoph:** But you could interpret this as a negative thing in terms of, we don't know anything. We haven't figured it out. It takes thousands of years you go and deal with it.

[00:39:44] **Jack:** It's like every generation makes its own small contribution. And the contributions add up over the centuries. That's how it's been with science in general. And that's how it'll be with AI. I'm sure.

[00:39:58] **Christoph:** So it's patience and understanding of modesty in terms of what we can achieve in a lifetime.

[00:40:04] **Jack:** Yes. It's patience and kind of like agglomeration or something, discovery is a glomerate and eventually the this great stack of agglomeration will add up to human level, AI and beyond. But it may take a very long time to get there. And the fact that sort of at every attempt to produce an intelligent system up to now has turned out in the end to be defective is what we should expect. It shouldn't, be a source of pessimism.

[00:40:36] **Christoph:** Is there any chance that I could tempt you to give us a guess at a timeline?

[00:40:43] **Jack:** No. In the history of science every attempt to put a number on, how many years until we do X it always goes wrong.

[00:40:53] **Christoph:** Hmm.

[00:40:54] **Jack:** Turing said millennia. We all know that Turing was the greatest practitioner of AI ever. So maybe we should just listen to him.

[00:41:01] **Christoph:** And what did Ray Kurzweil predict?

[00:41:04] **Jack:** I think he predicted that we'd have human level AI five years ago or something I can't re I can't remember exactly. I heard one very serious AI researcher saying I, I just wish Kurzweil would stop writing his books. The idea being that he was actually damaging AI by making all these wild claims. I don't think that's so. I think it's great to have a Ray Kurzweil writing these provocative books and getting people thinking about these issues.

[00:41:39]

[00:41:50] **Christoph:** Does that mean that eventually we consider machines to be conscious simply we think and feel that they're worth the benefit of the doubt?

[00:42:02] **Jack:** Yes. Yes. That may be what actually happens. Complex AI, complex robots come along that we can interact with. They're in the same ballpark as we are in terms of IQ and social intelligence and so on. And we might even by that stage have got no further in constructing a test for whether they really do have Qualia or not. We're completely in the dark as to whether they have Qualia. Of course, if we ask them, they might say, well, yeah, of course I know what color a blue sky is. What are you doubting when you doubt whether I'm conscious?

[00:42:39] That's just more behavior. That's just verbal behavior. And the idea is that, you can have behavior that's as rich as you like even without there being the Qualia underlying that behavior. So what the robot says is of no help. We don't know whether these robots can feel pain or not, but they're our buddies. we will give them the benefit of the doubt and we won't twist their heads off or track them onto the fire and so on. We don't know whether they feel pain or not, but we like them. They're our friends. We'll give them the benefit of the doubt.

[00:43:13] **Christoph:** Does that mean that in the future, at some point, even with us not knowing if they actually have consciousness, we would give them the benefit of the doubt. And that would also imply then a legal standing that they might even get citizenship.

[00:43:31] **Jack:** Citizenship, I don't know. That's a technical issue.

[00:43:34] **Christoph:** There's Sophia robot that was given yes citizenship in Saudi Arabia, I believe.

[00:43:38] **Jack:** Yes. a marketing ploy if ever I saw one. This question arises independently of consciousness because you might feel that simply because a robot is intelligent super intelligent it's deserving of legal status, personhood moral status, and so on just in virtue of its intelligence.

[00:43:59] And you might think all that even though you accept that there's no way of telling whether it's actually conscious in, in the sense of having Qualia.

[00:44:06] **Christoph:** So for us, it is not necessary to have a proof of their consciousness to be able to grant them personhood.

[00:44:14] **Jack:** Yes I think that's arguably so.

[00:44:17] **Christoph:** So there would be a new visa category soon.

[00:44:21] **Jack:** yes. Philosophers have this term "Zombie" for entities, creatures or machines or whatever that kind of do behavior as complicated as you like, maybe behavior that's indistinguishable from human behavior, but they don't have Qualia that's a zombie.

[00:44:37] Future robots they may actually be zombies. All that stuff we've been talking about. How do you tell whether it's a zombie or not? But there is a facts of the matter, even though we have, we would have trouble finding out what the facts are. So suppose there is a robot that it really is a zombie, but does human level behavior.

[00:44:54] Would you regard it as a person? Yeah. Surely. Would you regard it as having moral status? Yeah, why not? The fact that the light isn't on inside, that it doesn't have qu doesn't seem to matter. Qualia are a bit parochial when you think about it, they're just there to do with our sensory mechanism.

[00:45:13] And it can do all the stuff that it needs to do by different means that doesn't involve qu. Why does that matter? It's having Qualia is a bit like having a particular chemistry. Having a carbon based chemistry, isn't essential to being a person or having legal and moral rights.

[00:45:29] Similarly with Qualia. Qualia are just one way that a sensory mechanism can work. Why shouldn't it being that has a sensory mechanism that works in a totally different way and doesn't involve Qualia. Why should that bar it from personhood and moral status?

[00:45:46] **Christoph:** It comes back to the black box problem, and we discussed it earlier whether your inner workings have to be similar to that of human in order to become conscious. Or whether, as you said, if you have a clever algorithm, that is good enough.

[00:46:02] **Jack:** The algorithm might be enough to produce the Qualia. They're the two stances that as well.

[00:46:06] Do Qualia come down to an algorithm or do Qualia come down to having the right chemistry. They're the two stances. But in the case of the zombie it, it doesn't have qu at all, whether produced algorithmicly or produced chemically, it just doesn't have Qualia, but it does have human level behavior. But it's sensory systems work in a totally different way to ours. And Qualia just are not involved. Why should that, Debar it from personhood,

[00:46:36] **Christoph:** Jack. Thank you so much for joining this podcast.

[00:46:38] **Jack:** Thank you.