The Future of Work podcast is a weekly show where Jacob has in-depth conversations with senior level executives, business leaders, and bestselling authors around the world on the future of work and the future in general. Topics cover everything from AI and automation to the gig economy to big data to the future of learning and everything in between. Each episode explores a new topic and features a special guest.

You can listen to past episodes at www.TheFutureOrganization.com/future-work-podcast/. To learn more about Jacob and the work he is doing please visit www.TheFutureOrganization.com. You can also subscribe to Jacob's YouTube channel, follow him on Twitter, or visit him on Facebook.

00:00 Jacob: Hello everyone, welcome to another episode of the Future of Work with Jacob Morgan. My guest today is Toby Ord, he's a philosopher and Senior Research Fellow at Oxford's Future of Humanity Institute and he has a brand new book out called The Precipice; Existential Risk, and the Future of Humanity. So Toby thank you for joining me.

00:21 Toby: Oh great to be here.

00:23 Jacob: So I gotta admit when I looked at the book and read the book, I got a little freaked out because the title itself is scary. And then so as I started to go through the book and read a lot of the content, you made some very interesting arguments and points so that I think a lot of people are thinking about now, but maybe we just don't talk about as much. But before we jump into some of those things, why don't we start with just high level, as a philosopher and as a research fellow at the Future of Humanity Institute, what do you spend a lot of your time studying and thinking about these days?

01:02 Toby: Yeah, I've generally focused on big picture questions facing humanity. So that's not what people who study ethics in philosophy are normally doing. Normally the question is looking at the questions of everyday life, of what should I do. So studying the ethics of when is it okay to kill? Is it okay to kill if you're a soldier during a war? Or when is it okay to lie? Things like that. But I'm interested in these questions on a much larger scale. And so, earlier on in my career, I spent a lot of time looking at the ethics of global poverty. So this is a huge issue facing billions of people in the world and what are the obligations that we in the rich countries have to do something about that and how can we best help? So that's what I was looking at when I started my career. And I've also been interested in this other theme, which is where the book comes in. Looking at these kind of grander questions about the history of humanity and the future of humanity. And are there any risks that could threaten our entire future? So that's what I'm... I've spent a lot of the last 10 years asking and thinking about.

02:12 Jacob: How do you even begin to understand that? The history aspect makes sense, right? We have documents, we have records, we have data, we have lots of things we can look at for the past. But when you think about the future and some of these existential risks which we'll talk about in just a few minutes, how do you even go about thinking about that, determining if they're actually risks?

02:36 Toby: Yeah, it's a good question. I should say with the past, we have written records dating back about 5000 years. But humanity if you take it to be Homo Sapiens, goes back 200,000 years. And that's sometimes fascinating in and of itself, to think that for almost all the time that people have existed and for lots of the greatest things that humans have ever done, you know, a lot of fierce

friendships and strong loves that people have had for each other, fighting against adversity, a lot of the adventures that people have had, finding all of the plants and understanding the animals and the ecosystems and first person to see the tiger or the rhinoceros or the first person to enter Australia, this whole new continent of different animals. That a huge amount of that happened before we even had the ability to write down and understand it. So thinking about that, about the past, it has really opened my eyes to just how epic the span of human history has been. And then that helps in thinking about the future as well, to think about the next 10 years, or the next 20 years, which is what a lot of people think about when they think about the future. But also know about the next 200,000 years or millions of years, how long could humanity last?

03:54 Toby: So that's something I'm very interested in, I have looked into a lot of the astrophysics of questions about the earth's lifespan and things like that. And when it comes to particularly the risks that we might face over the next 100 years. Yeah, I've had to read a lot about science and technology and really talk to a lot of experts. That's been a real focus with the book. It looks at a lot of issues in cutting edge science and I really... This is a real area where it's easy to screw it up when you're writing a book like this if you have a great idea about something closer to your own discipline, but then you have to say a lot of things about other disciplines for it to make sense. It's easy to just kind of make it up. So I wanted to really make sure I didn't do that. And I talked to really the cutting edge experts in all of these different risks and I also have them look over the book before it went to print to make sure that I hadn't made any errors and that I was faithfully conveying the cutting edge information about these things.

05:00 Jacob: How or what sort of a timeline, and right after this we'll jump into what some of those risks are. What sort of a timeline were you looking at for these things, or was it spread all over the map? Some very far out, some closer.

05:15 Toby: You mean, when would the risks strike?

05:19 Jacob: Yes.

05:20 TO: Yeah. So I generally set myself the next century as the timespan. With a lot of risks, such as the risk of asteroids and in general with the natural risks, these are things that basically could strike us at any time. And the chance it happens in 100 years is basically 100 times the chance that happens in one year. So it doesn't really matter what timeframe you look at things like that over. When it comes to risks from emerging technologies, then it really can matter. If I said the next decade, then some of these things such as the risks from advanced artificial intelligence probably wouldn't really register because they're unlikely to happen in that timeframe.

06:00 Jacob: Yeah.

06:00 Toby: So I wanted to pick something that was long enough, and felt good to really get your teeth into. And I thought 100 years worked pretty well.

06:10 Jacob: Okay, well let's jump into what some of those risks are. And I think some of the risks people will very much be able to relate to for example, some of the technology risks, climate change that you talk about in the book is something everybody is concerned... Well, a lot of people are concerned with. [chuckle] So maybe you can outline what are some of the risks that you talk about in the book?

06:32 Toby: Sure. So I divide them in the book into the natural risks, first. So things like asteroids, comets, the supernovae, so stars exploding, that's another possibility, super volcanoes, so these massive eruptions of volcanoes that are so big like the one in Yellowstone National Park, where instead of towering above the ground, they're kind of these sunken craters 'cause they're too massive to actually be able to stand up. And that there is indeed some risk from them. And these are the natural risks. Then I also look at the risks, the anthropogenic risks of today. And my focus there is on nuclear war, on climate change, and of all kinds of other environmental catastrophes. Ways in which by destroying some crucial part of the environment that might be the end for us. And then I look at these anthropogenic risks that aren't quite here yet, but are on the horizon. And my focus there is on engineered pandemics and also unaligned artificial intelligence.

07:49 Jacob: So quite a few things going on out there. So it sounds like there's a lot of stuff that could potentially wipe us off the map. And so for a lot of people listening to this, they might be thinking, "Alright, why should I care? Why... I can't control a comet coming over here or a volcano or anything like that." So for people listening to this, who maybe have that mentality in their mind, how can we make this a little bit realer for them? Why should we all care about these things?

08:18 Toby: Yeah. So I can see people not being that concerned about the asteroids, say. And I think that maybe they're right about that in some ways. Because, when we look at the current levels of risk, as we best understand the problem at the moment, I think that there's about a one in a million chance of us being wiped out by an asteroid in the next hundred years. So that is a small chance. And it's, at least in terms of your own life, is we often neglect chances that are smaller than one in a million. Maybe they're right about that. I think that if they read a newspaper report, which said that there's an asteroid, which is on an almost direct course to hit the Earth, and it's definitely gonna come in much closer than the orbit of the moon, and there's a one in 10 chance that it's gonna hit the Earth. I think that they would be really caring about that.

09:07 Toby: And that this would be... They wouldn't just kind of turn the page in the newspaper and kind of carry on about their day. This would be the biggest issue facing humanity. So I think that it does partly come down to the amount of probability of these things. Something like an asteroid has this nice feature that we can understand it scientifically, quite clearly, and we can get these fairly robust numbers out of it. Whereas with other things, such as the risk of an engineered pandemic, either making humanity go extinct or causing the permanent collapse of civilization. Sometimes we read some stories about how bad these things could get and we feel very alarmed but it's hard to put precise numbers on it. And so I think that's one of the reasons that people, I guess, either panic a lot or shrug it off, is that it's hard to just have some kind of mid-level number that goes.

10:05 Jacob: From the risks that you talk about in the book. Which ones do you think are the most, I don't wanna say real, but most immediate or closest in terms of time horizon?

10:16 Toby: Yeah, time horizon, I would say nuclear war is something that could happen soon. Although, luckily, what... The Cold War is over. And it's not just any nuclear war that could cause a threat of human extinction. It's only the largest scales of nuclear war that would really pose such a threat. And even then, we would probably get through. It would be devastating, but we probably would make it through. So maybe it would require another Cold War starting up with Russia or perhaps with China, if they increased their nuclear arsenal in such a war. So I guess that it's hard for that to happen in say, the next 10 years. With climate change, I should say, by the end of the century though, anything could happen. It's not actually all that long since the end of the Cold War and the century is a long time. So the political situation could get very different.

11:21 Toby: With climate change, it's actually quite hard to say whether climate change poses a real risk of human extinction or the permanent collapse of civilization, which are the types of levels that I think about in the book. So, and if it did, it wouldn't happen in the next hundred years. But perhaps the next hundred years would be a time where we pass the point of no return or something like that. So it could still be very critical on that level of timeframe. And then within the next hundred years, I guess, within the next... After a couple of decades, I think that the risks from these engineered pandemics and artificial intelligence get quite high too. So I guess it's a bit of a mix there.

12:03 Jacob: Yeah. Well, let's talk about one that I'm sure a lot of people hear about. Well, I suppose there are two main ones that a lot of people are familiar with. One is climate change. And the second one is artificial intelligence. Maybe we could just quickly touch on climate change because I know that a lot of businesses and organizations around the world are very much talking about this. And it's becoming a very central issue to the point where a lot of employees around the world want to be a part of organizations who are taking a stance on this and who are helping the environment, helping the world. So what are you seeing in terms of the climate change discussions, the risks? How would you say that we are currently doing, as far as that goes?

12:47 Toby: Oh, in terms of the discussions of the risk, I would say not well at all. I don't think that any normal person who doesn't spend their whole time thinking about this could be forgiven for having any opinion ranging from, "We're almost certain to go extinct.", through to, "That's completely impossible." based on what they read. Because you see all kinds of articles making all kinds of claims. And there's not many people really holding them to account and making sure that they don't go far beyond what the science suggests.

13:22 Toby: So when I was writing this section on climate change, I expected to be able to conclude that it doesn't pose an existential risk. So one of these risks of extinction or permanent collapse of civilization. But I found that as I went for the literature, it was harder to be sure than I'd thought. So, I'd thought originally that we could rule out extreme warming such as the kind of Venus like runaway greenhouse effect where it gets 50 degrees or 100 degrees hotter than it is today and the oceans boil off. I thought we could just completely rule it out. But it turns out that while there are good papers that suggest it's not possible, a good paper suggesting something is not actually 99.9% certainty that it can't happen. There's plenty of papers that get over-turned. It's not settled Science and I was as quite surprised by that. Still very unlikely to happen but it's hard to say how unlikely.

14:16 Toby: And then when it comes to... I thought that the history so called Paleo-climate data, just the fact that it's been harder in the past, than it's gonna get now. Surely, that should show us that the earth can recover from getting this hot. But it turns out that data is not that robust. I wouldn't be surprised if in the future people say, "Oh we got that wrong using our methods and new methods say that it was the temperatures in the past were a bit different, and there were many things that were very different about the world back then, such as there was a super continent. And that could change how that temperature affects things. And then perhaps more importantly than all of that, the rate of change, the amount of degrees of warming per year, looks like that may actually be unprecedented in the whole of the Earth's history and in which case you can't draw many conclusions. So in the end, while I wanted to be able to rule it out and couldn't, the main risks are really, it's not that we know of a particular effect of climate change that could definitely pose this existential threat, but rather that we're not yet at a position where we can completely rule it out. So I put this at something like one in a 1000 chance.

15:34 Jacob: Okay, well let's talk a little bit about everyone's favorite topic. And that is AI and technology, because we keep hearing that debate non-stop. The movies keep coming out there like Terminator. It's everywhere. It's become part of culture of these days. And I suppose the number one concern that a lot of people have first is the impact that it'll have on jobs or the economy or on what we're gonna be doing as individuals. And then maybe one level up from that is what happens if/when AI becomes human intelligence or surpasses us. So maybe we can start with the first one. Have you thought about the impact, or have you been studying the impact of technology on jobs and just what we're gonna be doing in the economy?

16:23 Toby: So that hasn't been a central focus of my work but I have been following the debate on this, where one side roughly speaking, one side thinks that we've had lots of technological change in the past, we have automated all kinds of jobs. It used to be that that most people did jobs in agriculture and now something like 99% of those jobs have been automated, and so only a couple of percent of people work in agriculture. And so we can kind of survive in almost all jobs we used to do being automated. At least so long as it happened slowly enough. There's this one kind of angle on that, where they say the employment rate in 2020 is not that different from the unemployment rate in 1820 or 1220 or something. It turns out that most people have jobs in all centuries. That's just one approach. And the other approach is the analogy with something like horses where it used to be that there were jobs for a lot of horses in the world and then if you look at a chart of how many horses there were in the US over time, you see that it radically comes down as horses just become something that people have for leisure rather than at doing useful jobs of taking people places, taking the mail places and so on, working in agriculture.

17:45 Toby: And so, the kind of question is, will it be like it's been in the past for humans or will it be like horses where ultimately, there maybe a few kind of oddball jobs for humans left but almost everything's gone and almost... And there're hardly any jobs for humans? And I think it's very difficult to know which of those will be true. And so the way I see it is that we can't really rule out... We can't be more than 90% confident in either of those views I think in this case. Which means that we're facing really at least a 10% chance in terms of how you think about it that we're in the world where all the jobs get automated and that it radically changes how work works. I just think that we have to have a reasonable belief that that could happen. And therefore, even a 10% chance of that is a huge thing that everyone has to plan around. I think you can't really avoid that.

18:46 Jacob: Yeah, so I mean a chance is a chance.

18:48 Toby: Yeah, exactly, and yeah, there's all this kind of... I don't know what people are hoping to do in academic debates. You hope to show that your side's right. But a policy-maker who's listening to two academics, say two economists, talking about this topic that they shouldn't end up thinking you know, one side's 100% right and they're definitely gonna happen. If it isn't totally... Well, one side is saying that this is an unprecedented change to automation because we've previously automated the physical work, making people increasingly move towards the work that requires a large amount of cognitive ability, whether that be dexterous hands or whether that be knowledge work.

19:33 Toby: And that if we then automate that stuff, there is nothing left. So they are saying this is unprecedented, and they seem to have a pretty good case and it's just very hard to then rule that out like to get more than... Yeah, as I said more than 90% confident that it's not true. And then you have really gotta take that seriously when you're planning.

19:50 Jacob: When it comes to technological progress. And I know you talk about this in the book, can you give listeners a sense of just how much progress we have experienced? Because sometimes it's hard for us to feel it because the day-to-day... We don't experience the changes, but if you look back throughout history of just how fast things have evolved, can you give us a sense of just how quickly technology is evolving?

20:13 Toby: Yeah, sure, so I think it's particularly useful to think about this with the last 200 years. So basically, since 1820, so think back to the time of the Industrial Revolution. So at that time it was 9 in 10 people lived on the equivalent of less than $2 a day. So only... Another way, think about that. Only one in 10 people had more than what we count as extreme poverty today. And that rate has gone down, that proportion until now, it's less than the other way around, less than one in 10 people is below that extreme poverty level now, and of course they're still living on far too little but back then, almost everyone lived on far too little. Another thing is literacy, in 1820 was about 10% of people, were literate in the world. And now it is more than 80% of people are literate. More than 40% of children died in the first year of life in 1820, and now it's less than 5%. And life expectancy, so how long people lived on average has more than doubled over that time, so we're living lives that are twice as long. So I think that this is a huge amount of change. Particularly in this period since the industrial revolution, in terms of the quality of our lives.

21:51 Jacob: Yeah, it's very, very exciting to see some of these positive numbers that we are seeing. And I suppose just in terms of technology, when I had somebody, for example, like Pamela McCordoc the podcast. Or whether it's a technology executive at a company they're always trying to quantify just how much technological progress, we've had in terms of what we used to launch spacecraft, several decades ago, to the power that we have in our iPhone, that we carry around in our pockets now, to where this is gonna be in 10 years, 50 years, 100 years. So if you were to maybe look at specifically the AI technology piece, what are you seeing there as far as progress goes? Are we making... Because obviously AI is not a new concept, it's been around for many, many decades, since I think the '60s, the '50s.

22:44 Toby: '50s.

22:44 Jacob: The '50s and we haven't achieved true AI yet, but it seems like now a lot of people are saying we're gonna get there. So has something changed over the last few decades? That's bringing us closer to that?

22:58 Toby: Yeah, I would say that probably... Let me see. So back in the early days we thought, somewhat intuitively, that there were certain tasks like playing chess that were the pinnacle of human intelligence, there are other ones as well, like advanced mathematics or logical reasoning, and we thought that these types of things were gonna be the hardest things and AI scientists worked on them and actually had a lot of success even quite early on. It took until the Battle of Deep Blue and Kasparov in the '90s to really become the best... For AI to beat the best humans, at chess. But it didn't take that long before they could beat me at chess.

23:48 Toby: But what was surprising and... Was that there were a whole lot of other tasks that we treat as actually really easy tasks, the type of thing that a two-year-old could do, such as picking up an egg, or identifying a cat, that we think of as really easy. But it turns out that the AI systems that we were developing, couldn't do those tasks. So the early researches were very optimistic because they thought that they were taken great progress on the hard tasks, but what was surprising was

really the ones that we think of as easy are hard for AI and vice versa. So that was a big thing to eventually to learn and take into account. But now we're finding that some of those things that were... That a two-year-old could do, that our AI systems couldn't do. We're actually working out some good ways to deal with them. So one of the big problems was that these early systems were quite symbolic.

24:46 Toby: You told them, you programmed in stuff about a chess board, and in terms of the symbols... That they have a king here, and that it can move here and it's a 8 by 8 grid and so on. That was all directly programmed in. It wasn't that it was looking at a chess board. And they were very good at manipulating symbols, but they weren't good at what we call the symbol grounding problem, of working out what those symbols stand for... So they could do things with the word cat, but they didn't understand what kind of thing in the world this kind of fuzzy ball of things with four legs running around, that actually is a cat. And the real breakthrough recently, has been with deep learning. So a way of using neural networks with the huge amounts of data and computation that we have available has really helped us do this simple grounding problem, and actually have systems that can take raw pixels of input and actually make progress with that.

25:46 Toby: So one of the best examples, I think, was what DeepMind did before the famous games of Go and Chess that they played, beating the world's best players. They had this as Atari playing system that would play these old Atari Games. And it could learn to play them just from the raw pixels. And that's something completely different to what any of the early AI systems could do. So if you put one of these Atari Games in front of it, and just tell it the score and let it see what's happening on the screen. Then, after a month of playing it, it could achieve human-level performance on about half of these different games it was shown. And so this was like a real breakthrough in doing something that we know... We think Atari is easier than Go but it's because of this kind of paradox, where the hard things are easy, and easy things are hard, that this was a real game-changer.

26:42 Jacob: What is AI? How would you explain it? Because a lot of companies say it. Every company says it, a lot of companies say they offer it. But it seems a lot of people have different definitions of what AI is. So how do we know what AI is? How would you explain or define it?

27:03 Toby: Yeah, so one of the definitions that I like best, although it's a bit narrow, is that intelligence is the ability to achieve your goal in a wide variety of circumstances. So you, therefore, have to be able to adapt to your circumstances, in order to deal with different kinds of circumstances, to find your way around obstacles and so forth. So it's a nice kind of short definition. But it is very focused on an agent. It's something that has goals, that's kind of trying to achieve them. But we might think that systems that... Take this GPT-2 system that OpenAI developed last year, that has read a huge corpus of text, a very large amount of texts from the internet. And then if you start it off with a kind of plausible-sounding couple of first sentences, it can write the rest of a few pages for you and sounds at least like a plausible person on the internet.

27:57 Toby: And that system isn't trying to achieve any kind of goal. That's not really how it works. It's not a goal-directed system. But it still seems to be exhibiting at least some kind of intelligence or doing some kind of cognitive work, we could say. So maybe you want a broader definition that can include things like that, as well. And at the kind of most broad, there's this question about whether things that are getting badged as AI at the moment are really just doing statistical methods, that wouldn't have been called AI 10 years ago.

28:29 Jacob: Yeah, that's the big challenge today, because a lot of people are saying everything is AI, and it's really just, computing or an algorithm doing something. So I think that's actually a very interesting distinction is something that has goals. So what could be an example of a goal that AI might have? And please don't say wipe out the human race?

[chuckle]

28:50 Toby: Yeah. So these game playing things very simple examples. The aim of AlphaGo was to win a game of the game of Go. And it's kind of interesting and somewhat instructive to think about exactly what that goal was. Because it was given, it could kind of a cheat. It wasn't looking at a board of Go either. It was told exactly kind of how a board is constructed in terms of the mathematics. And so it wouldn't notice, for example, that maybe you could win the game of Go by getting the other player drunk, or something like that. And that's totally out of how it's considering the game. It's not kind of noticing that there are physical objects that are kind of playing the game with or anything like that. It's just trying to work out what would it do to beat another copy of itself? So while it has a goal, it's not quite what we think the goal is. It's not really a goal about the real world. It's a goal about it's kind of abstracted version of the game of Go.

29:51 Jacob: And I actually followed that, I love chess, and anybody who's ever listened to any of my podcasts or watched any videos knows that I'm pretty obsessed with this, so I love that you brought this up. But I remember in the game, what year? Was it 2017 when this took place against Lee Sedol the Go champion. And I believe it was, was it game five or game...

30:17 Toby: Game four, I think.

30:18 Jacob: Where there was like that move that the computer played and everybody was like, "Oh my god, this is the sign of creativity in AI." And everybody like lost... I'm not gonna curse on the podcast, but everybody lost their marbles.

30:30 Toby: Yeah. That right.

30:31 Jacob: And all over the place, people were giving talks about it. It was in papers, "AI has achieved creativity, everybody runs for the hills." That seems to have died down a lot. Nobody's really talking about that anymore. So maybe first, can you explain what the significance of that was for people who are not familiar with it? And I'm just curious to hear your take on it.

30:51 Toby: Sure. So I'm no Go expert, but I have asked about this question by people who are and what's going on there is that in the game of Go, you take turns placing, one player plays white stones, one player plays black, and you choose an empty location on the board, and you can place a stone there. And in certain cases, you capture other people's stones, or you claim territory and the person who claims the most territory at the end wins. And one of the ways to claim territory is to cordon off areas around the corners or the edges of the board, and there's a concept of how close are you to the edge of the board and how far away from the edge of the boat can you play and be confident of being able to claim the territory from where you are to towards the edge? And what AlphaGo did with that move, is it played one level further away from the edge of the board, than people thought you could get away with. So if you can succeed in doing that, your stones can claim more territory, and also have more influence over the center of the board. So it's much more powerful for you if you can play in that kind of daring way.

32:00 Toby: And effectively, people, it seems like humans have been playing it too safe because they didn't realize that if they were playing... Maybe if they were playing better, they'd be able to prevent any attacks that tried to stop them claiming that territory. And so this was something that it was... We certainly would have thought it was creative if a human had done it, which is not quite the same as saying it's necessarily creative. And ultimately if you have a game like Go or again like chess, at some level the game is kind of closed. So there is... You could imagine a kind of a branching tree diagram for the game of Go or the game of chess where the first move of chess, players have, what is it? 20 moves I think that they could make, and then the opponent has 20 moves they could make and then... The number changes as you play, but ultimately you can only go on for finitely many moves, and so there's only finitely many different games of chess that you could play.

33:03 Toby: So in theory, if you had enough computation ability, even just with a brute force algorithm, it could find the best possible games of chess that anyone could play. The types of games that if we saw, we'd say were dazzlingly creative solutions to these problems, even if it was just a boring brute force system with enough compute. So that's a bit funny as to whether you should really call those things creative or not. But since the AlphaGo system did not have a astronomical amount of computation, the kind of thing that's more than the whole universe worth of computers that would be needed in the example I gave. So it did it with less computational ability than the human brain, so maybe we should call it creative.

33:48 Jacob: Yeah. That's what a lot of people have been saying and...

33:52 Toby: And... Have...

33:53 Jacob: Oh, go ahead.

33:54 Toby: Have you seen some of these chess games that AlphaZero has played?

34:00 Jacob: Oh yeah, the recent ones, actually, some at the end of, well, last year, and I think there might have even been some this year. Yeah, AlphaZero playing against itself... Actually no, it was AlphaZero playing against [34:13] ____...

34:16 Toby: Stockfish.

34:16 Jacob: Stockfish, but there was two versions of AlphaZero I thought they made.

34:20 Toby: Oh.

34:22 Jacob: One that had the rules and sort of some things embedded it in and it was looking at games from the past, and another version where it had to learn everything on its own. And the version were it learned on its own beat the other version that looked at human games and information 100 to zero.

34:42 Toby: Huh? Oh. Wow. Okay, I should find that and watch it. The games I've seen which were from the earlier one against Stockfish were still amazing, but they're so far above my level that I had to watch them with expert commentary but it was... Yeah, it was dazzling.

35:00 Jacob: Yeah, it was... So okay, now I have to ask. Are you a chess player Toby?

35:05 Toby: Not really.

35:07 Jacob: But you get the rules of the game, you play it a little bit?

35:09 Toby: I get rules of the game, exactly.

35:10 Jacob: Alright, alright. [chuckle] So looking then, bigger picture, when we look at the future of humanity with technology and AI there's of course a lot of conversations around something like a Skynet coming or something like from the film the Matrix where technology just takes over and we're slaves to it. Is that gonna happen? Be honest with us Toby.

35:35 Toby: I hope not. So. Yeah. Should we break down, what are the types of concerns here?

35:43 Jacob: Yes. Yes please.

35:46 Toby: So the way I see it, is that AI systems are getting increasingly more sophisticated, they're getting more able to solve a wider range of tasks and to do so better than humans. So they're becoming more general and then also better at each of the task that they can do. So one definition of something called Artificial General Intelligence, is an AI system that could accomplish every task better and more cheaply, than human workers. And recently, a few years back, 300, top researches in machine learning were surveyed on this question of "When could an AI system do that?" This obviously has big implications for the future of work as well. Right? If the system can accomplish every task better or more cheaply than human workers then it's not clear of what we're doing. And quite amazingly, on average, they estimated a 50% chance of this happening by the year 2061 and even a 10% chance of it happening as soon as 2025, which is five years from now it was nine years from the time when the survey was asked. So maybe they wouldn't still say that. But they're saying that it's not impossible for this to happen, in a decade, very soon. And I guess they'd be somewhat surprised if it didn't happen by the end of the century, but not all that surprised.

37:15 Jacob: Well, Ray Kurzweil, didn't he also... What was his year? Did he say 2040 or 2030 where they...

37:21 Toby: So. Yeah, I don't know exactly. I can't remember the Kurzweil year, but there's something weird about this question. If someone says, "When will something happen?" and you give some year as an answer, you say, "It'll happen in 2040." Like, "How?" It's weird that you're so confident that it'll happen at this one particular time. I think the better way to do it is this kind of question where you say, you imagine like a curve that you're drawing on a piece of paper, and you imagine every year over the next century and then you say, "What chance is there that it will happen before that time?" And then you kind of draw this curve. And I think that that's a better way of doing this. And so this kind of question of where does that curve reach 50%?

38:00 Toby: I think is a good way of kind of, if you just had to pick one number one date. And so these experts said that it hits 50%, so just as much chance of it happening as not happening in 2061. Now if 2061 comes and it hasn't happened that doesn't mean these people are wrong, 'cause they actually said there's as much chance that it wouldn't happen, as that it would happen. But it's... I think it's a better way of asking the question that doesn't somehow involve you saying there's a 90% chance that happens on this one year or something, and no one could ever know with that level. But if you take these people at their word, they're basically saying something like a 50% chance that AI

systems will be able to do all of this... Everything a human can do, at least in terms of intellectual tasks this century.

38:49 Toby: And if we did that, as well as there being kind of massive implications for work, there's massive implications for our species. So homo-sapiens, if we look at this 200,000 year history, and we ask how did we get to where we are? Why is it that it is humans that are in control of their destiny? Humans that in a way that say, chimpanzees or black birds are not in control of their destiny. If humans do something that happened to wipe out these species there's not much they could do about it. Hopefully we won't, but it's in our hands, not in their hands unfortunately for them. But what is it that put us in the position of power over the planet? And ultimately, it all comes down to our mental abilities. So intelligence, maybe things that we don't normally call intelligence, but are still mental, such as our ability to communicate with each other through language.

39:50 Toby: But it's definitely the mental, not the physical abilities of humans that are the reason we have got this position, this commanding position where we control our own destiny and could have an immense potential for the future. And so if we do create AI systems, these general AI systems that can do everything we can better and more cheaply than we can, why would it be us who we are in control of our future from that point onwards? It's a... That's... There could be an answer to that question. Perhaps the reason that we're still in control is 'cause we carefully kind of encode rules into these systems to make sure we're still in control. Maybe we manage to make them do what we want, or we manage to make them do what they want, but we cunningly make it so that what they want, when they build their ideal world they're also building our ideal world. So there could be some answers to that question, about why is it that things won't go very wrong, at that point.

40:50 Toby: But it turns out that those tasks of either making the AI systems listen to us and do what we want or to make them aligned with us in terms of their values, having the same values as us, are both extremely hard and that people in AI are looking at those problems, are really searching around for solutions. They've got a few ideas but they're saying that this looks extremely hard and we might not be able to get that done in time before we have systems which have that level of power.

41:18 Jacob: I know that you also spend a lot of time, you've advised governments and leaders at various organisations and governments around the world. What did they ask you about the most? What are they most either worried about or concerned with or wanting your advice and guidance on the most?

41:39 Toby: So some of this was on my earlier work about global poverty. So trying to understand how we can most effectively help people in poor countries. And some of it has been... Yeah on future trends and technologies and ideas for example, about interest in AI and work. I would like them to always be asking me these other questions about existential risks. These are risks to the entire future of humanity and what they could be doing to protect us. They don't tend to ask me about that. Hopefully, after this book comes out, they will... But my experience when talking to them about those existential questions is that... And they say, "Wow that's really interesting, but it's above my pay grade." And everyone seems to react like this at least up all way through the national level of government. That it's something where it just feels a bit too big for them to deal with. And they're used to thinking about the new cycle the next week or so or about the election cycle. But something that's, that you're talking about, what do we need to put in place such that we can be protected from engineered pandemics in 20 or 30 years time?

42:56 Toby: How do we need to start working now in order to avoid that? It's so far beyond their normal horizons and it's at such a level thinking about not just a country and not even just global level, but the entire future of humanity that they're not really used to thinking about those questions at all. And I'm hoping to make them better at thinking about these things.

43:16 Jacob: Are these questions we should all be thinking of, or just people in positions of power?

43:21 Toby: Yeah. I think they're questions that we should all be thinking of. I think that they're, they're outside of what we normally think of as the domain of morality. If you asked someone to say a few things about what does it mean to be moral, or you tell me about ethics, it's unlikely that they would be talking about this kind of stuff within the first thousand words that they say. But it is something where it's clearly linked to questions of good and bad and right and wrong. If someone takes some kind of risk that threatens the lives of everyone else on the planet, when they're building their new technology and threatens not just that, but threatens to break, to sever this thread of humanity, that's lasted for 200,000 years and could last for hundreds of thousands of years more, it does seem like they're doing something seriously wrong if they're taking these risks recklessly without appropriate reason.

44:24 Toby: So it is part of the domain of morality, but it's not something that we normally think of. But I think that that could change, just like it did for environmentalism where thinking about the environment wasn't really considered part of leading a moral life, up until about 1960 and then from '60 to '70, it radically changed. To the point where when I was growing up in the '80s, most of our moral education at school was about not littering and looking after the environment and so forth. Almost as much of it as there was about being nice to other people. So similarly animal welfare is something that wasn't really part of, it wasn't on anyone's radar 100 years back, but then that really changed. And I think that these things partly changed because the world changed, humans got powerful enough to really affect the environment and we started to notice some of the bad effects we were having. And farming practices got more industrialized such that we were having much worse effects than animals, and so our kind of public morality changed, it adapted to that.

45:25 Toby: And it adopted actually pretty quickly within decades let's say. And I think that something like that could happen here too. So the way I see it is that there have been risks that have faced humanity over hundreds of thousands of years. These natural risks. But it was only with nuclear weapons in the 20th century that we reached a point where humanity's escalating power over the natural world was so great that it could threaten our entire continued existence. And yet our wisdom and ability to actually behave responsibly had grown only falteringly if at all. And so it put us into this precarious position, where we still are, which I call the precipice hence the name of the book. And this is a time where we suffer these existential risks. And I think that this time can only go on for a few centuries either because if the risks stay at the current levels or increase, continue to increase, then I think we couldn't survive more than a few more centuries of this. But I also think it's possible that we'll survive it because we'll actually get our act together and we'll lower these risks and get them down to more sensible levels. That we'll grow up about this issue.

46:44 Toby: And I think that there's some kind of hope for that because if we look back to the Cold War, when people were thinking about nuclear weapons there was a lot of interest in this. The biggest ever protest in America's history in Central Park was against nuclear weapons on the grounds that they posed a threat to Humanity's continued existence. With the end of the cold war, a lot of the momentum disappeared behind those ideas. But it's come back with climate change, and

there's a lot of momentum. And again, huge protests and public action on this and public recognition. And what I'm saying is that nuclear war and climate change are both things of this wider category of existential risks, and that's kind of the real set of things. And we could continue on just waiting 'til one of them's got really bad before we kind of it rises to our attention, but it'll be even better if we can see them in advance and we can say, "Are there any other things that could be like this this century? What about if, biotechnology continues to get really advanced could that threaten us? And if so, is it possible to just head it off at the past and develop defensive technologies more quickly than the offensive technologies such that we don't end up fighting fires the whole time?"

47:56 Toby: So I think that this idea of existential risk, I think it could take off and as it has in the past of people really seeing the ethical issues about the continued survival of humanity.

48:08 Jacob: Do you think there's an overall trend in the world of focusing on these big picture issues more frequently? So for example, a lot of organisations are standing up for something like climate change, a lot of them are focusing on things like purpose and meaning and impact. Like these bigger picture issues whereas it seems like several decades ago, maybe some of this was out there but not to the extent that it is now especially with social media and all this stuff spreading all over the place. Do you think that we're just generally thinking more about these big picture issue ideas more?

48:45 Toby: Yeah, I think that generally we are. There are a few cases where the trends go the other way and you could say that we're spending a lot of time fixating on very small things. Our attention being fragmented and things like that, not giving us enough time to think about some of these deeper things. But one way that I look at it in terms of... Is in terms of this perspective of humanity, and you can think of this kind of increasing set of moral perspectives, where we've always had this idea of the individual perspective. What should I do when you're thinking about right and wrong? And occasionally, we'll ask these slightly bigger questions about what should my group or community do? What should we be doing? What are we doing right, what are we doing wrong? What could we be doing better? And then if you think of say, the 19th century, there was a lot of move to thinking about what should your nation be doing, and then in the 20th century, there was extra interest in taking that to the next level and thinking globally. What should everyone in the planet be doing about climate change or about the ozone layer and so forth?

50:00 Toby: And so I think that the next step in this kind of pattern is that we should be thinking occasionally about what should humanity be doing. Where we think not just all people at the moment over the whole globe, global issues. But we think about these issues over all time, and I think that that can be quite different. So if you think of humanity, we have been around for about 300,000 years. And if we're live as long as a typical species, we'd have about 800,000 years more. So they typically lost about a million years. And if we think of that in terms of a single life, then humanity would just now be in its adolescence and yet we find that we're kind of willing to risk it all for these variant small times, for the equivalent of just improving one hour in its life. It risks its whole future. So I find that this analogy is quite useful and that when you really are thinking on these kind of time spans, it helps to put it into a every day frame where you realize that it is like an insane teenager who is taking completely unreasonable risks for short-term reward.

51:15 Jacob: For most people listening to this who might be thinking, "You know what, this is interesting stuff but it's too big picture for me. I don't know what to do. This is just really overwhelming." What can we as individuals do to I don't know to think about this stuff better to

educate ourselves about these things. How does this impact us just on the ground level, if we're not in positions of power. I'm just worried about getting a pay cheque, waking up at the right time, taking my kids to school. This is so high level." What do you say to that?

51:48 Toby: So I think one of the key things is having a public discussion about this. So asking these questions. Is it true that humanity's entire future is at risk or how high is that risk? And what are the largest risks and what's anyone doing about it, and what should we be doing about it? And lots of questions like these, that we should be asking them in our families and so and with our friends and colleagues, the kind of thing that you can talk seriously about down at the pub and opening up these wider conversations about it.

52:24 Toby: So it's something where I think that... This is one of the reasons that I wrote the book was to try to show people open up their eyes to all of these issues that are facing us that potentially threaten our entire future and also to see how bright our future could be, if we can make it through this time. It was kind of my being so inspired by what humanity could do, and its potential that made me fight more fiercely to try to protect that potential and so I wrote this to try to let everyone really see that. Previously, discussions about this had been a bit dry and academic and harder for people to actually understand all of these issues. So wanted the people to be able to see it and to start these conversations, both personal level and then also as kind of like larger conversations in society. And then another thing that they can do is potentially to donate to organisations who are actually working on these things. I list a few in the book and for example, groups working to fight nuclear war, by avoiding proliferation, and promoting disarmament. This is...

53:34 Jacob: I was actually gonna ask you about yours, because you have one called "Giving What We Can" society, so can you share a little bit about that and what you have been doing, 'cause I think it's actually a very interesting and noble cause that you've started.

53:47 Toby: Oh, thanks. Yeah, so I started it in 2009 so about 10 years ago and it's a society... It's not itself like a charity that one would donate to it's instead more a society of people who have made giving a huge part of their lives. So there are 4,000 members who have each made a pledge to give at least a tenth of what they earn over the rest of their life to charities that can do the most to help others. So it's focused on when giving quite a lot, like a tenth of one's earnings and also on giving it effectively. Because we have found that some places you can give, can be shown to have 10 or 100 times as much impact as others. And we give some recommendations on that. And then if you are thinking about this and you make a choice to give 10 times as much as you were previously gonna give, say 10% instead of 1% and to give it somewhere 10 times as effective, you could have a hundred times the social impact over your life, with your giving, that you thought you might be able to have and you could with that, you could save many people's lives.

54:57 Toby: It's certainly possible to save, say 100 people's lives in your life and all the while living on 90% of what you would have earned anyway, and having a really still good life yourself, you don't have to radically change your career, you could do all of this. So that was the idea and I made this pledge, and I founded the society. Yeah, we have 4000 members, together we have pledged more than a billion dollars over our careers, to help others as much as possible, and so far the members have given more than $100 million and have already given enough to transform the lives of tens of thousands of people.

55:36 Jacob: So over $100 million so far.

55:38 Toby: Yeah, that's right.

55:39 Jacob: Geez, that's amazing, congratulations.

55:42 Toby: Well thanks.

55:44 Jacob: Well, I think we pretty much covered everything that I wanted to look at. Is there anything else that you want people to know or to think about before we wrap up? And then I'll ask you where people can find your book and connect with you and all that sort of stuff. But any parting words of wisdom for the listeners?

56:04 Toby: Yeah, here's one. Maybe words of foolishness rather than wisdom, but you might hope that these risks would be being dealt with at the highest level, but it's actually, it's pretty shocking how neglected that that they've been. So, as two examples, the Bio Weapons Convention, the BWC, which is meant to be the kind of the equal of the Chemical Weapons Convention and the Anti-Nuclear Convention, that the total funding of it is less than that, of a typical McDonald's restaurant.

56:39 Jacob: Oh my God.

56:40 Toby: And the existential risk on the whole, for all of these risks put together that humanity is currently spending less on safeguarding its future, than it does on ice cream. So this is not issues that are... That you think, "Well it's above my pay grade and I'm sure there's people handling it." These are issues where everyone is kind of passing the buck further up and we're not handling it very well. So I think we really... We have the potential to have a really great future. It's not a pessimistic book. And I think that we want to with clear eyes see the types of risks see how high they are and then act appropriately and defend our future, so that we can have a great future going forwards.

57:25 Jacob: I love that message of optimism and positivity. Well, where can people go to learn more about the book and connect with you. Anything that you wanna mention for people to check out, please feel free to do so.

57:38 Toby: Sure, yeah, people can find the book on lots of different bookstores, Amazon, other places and then when the book comes out you can find out a whole lot more at theprecipice.com.

57:53 Jacob: Perfect. Well, Toby, thank you so much for taking time out of your day to speak with me and share some of these really interesting insights and concepts. It's refreshing sometimes to think so big picture instead of focusing on the day-to-day that so many of us are used to, so thank you.

58:09 Toby: Oh, thank you.

58:10 Jacob: And thanks everyone for tuning in. My guest again has been Toby Ord make sure to check out a copy of his book. And again, it's called "The Precipice: Existential Risk and the Future of Humanity." And I will see all of you next week.