Rationally Speaking #215: Anders Sandberg on “Thinking about humanity’s long-term future”

Julia Galef: Welcome to Rationally Speaking, the podcast where we explore the borderlands between reason and nonsense. I'm your host, Julia Galef, and I'm here with today's guest, Anders Sandberg.

Anders is a researcher at Oxford, at the Future of Humanity Institute. His background is originally in computational neuroscience. That's what he did his Ph.D. in, but his research now focuses primarily on long term futures: What are the plausible, what are the likely and possible trajectories for humanity in the next centuries or millennia? And what, if anything, can we do to steer those trajectories?

Listeners may already be familiar with Anders' work in part because it came up in a recent episode of the podcast, the episode we did with Stephen Webb on the Fermi Paradox. The paper that we discussed at the end of that episode on Dissolving the Fermi Paradox — Anders was a co-author on that.

And then also, if you follow me on twitter, I recently shared a paper by Anders on the critical scientific question: what would happen of the earth was suddenly made of blueberries?

So, as you can see his interests are wide ranging. But today on the show we're going to focus on that central thematic cluster of his work, long term futures of humanity, and how to think about them. Which I'm very excited to talk about. Anders, welcome to the show.

Anders Sandberg: Thank you, Julia. It's great to be here.

Julia Galef: Yeah, well we'll leave blueberry earth for another time, but we'll link to the paper because it's delightful to read. And it's gotten a lot of well-deserved attention recently.

Anders Sandberg: It demonstrates that if you want to really get attention for your paper, you should publish it in July when nobody else is doing it, and certainly interest just explodes.

Julia Galef: I mean the fact that it was about blueberry jam, and blueberry granita layers of this hypothetical earth, in very explicit and rigorous scientific detail — that, I'm sure, had something to do with it, aside from the timing.

But yeah, so you've published a bunch of papers about long term futures. You're currently working on a book I understand, called Grand Futures, on this topic.

Anders Sandberg: This is correct.

Julia Galef: Excellent. So let's start off in true Rationally Speaking form with an objection. I'm sure you're very familiar with this objection, I'm sure it's come up a lot:

What do you say to the people who object that it's basically impossible to predict the future, and any attempts to do so are just ... I mean, you can speculate. You can speculate about what could happen. But we should have pretty low credence in any particular speculation. And that's low enough that it's not very actionable, to do such speculation.

That's sort of a fundamental objection to the central thrust of your research. How do you think about that?

Anders Sandberg: So my main objection to that objection is that we actually do predictions every day. And indeed that is what we've got brains for. A brain after all is an organ intended to make sure that we can eat better without getting eaten, and typically in higher animals it does that, by making various forms of predictions about the near future.

The important difference of course between the near future and the far future is, well the far future is sometimes much less predictable. And the critic would probably say it's nearly always very unpredictable.

I disagree. Very, actually a fair bit of our future is predictable enough that we can say interesting and true things about it, that might be useful actually in the present.

Julia Galef: Yeah, I mean I imagine that's sort of the crux of disagreement. Where people making this objection would say, "Sure, there are things we can have confidence in. We can have confidence in things that involve extrapolations of the laws of physics, like we can predict that entropy will increase, and we can predict that humanity can't literally live forever, because eventually the universe will expand and ... etc, etc."

But in terms of useful or actionable things... I think the belief or the assumption among most people, including most scientists, is that there's just nothing in that intersection.

So maybe it would be helpful for you to give a couple examples of things that you think we can say with some confidence about the future that are non-obvious or non-trivial and interesting.

Anders Sandberg: So I think that it's useful to recognize that the limits that are set by the laws of physics — and we might of course quibble about, "Do we know all the laws of physics?" and pretty obviously we don't, so we might have to update this. But there is a fair amount of limits that we have extremely good reasons to believe in. The reason to believe in thermodynamics is not just good theoretical arguments, even though they're very strong, but a lot of empirical arguments. It would be exceedingly weird if we found a way of overthrowing that, even though we can't rule it out.

So that means that we can to some degree lean on the known laws of physics. And not just laws of physics in the elegant scientific form, but also the things we have demonstrated to be possible using actual engineering.

So there's a very nice conceptual diagram Eric Drexler came up with, where he was outlining a space of possible technologies, and it has a boundary set by limits set by the laws of physics. And somewhere in the middle of the region that is allowed, we have this small spot of technology we have achieved.

But between that spot and the limits, there is this unknown area with likely possible technologies. And he argued that quite often we can explore that, because we can demonstrate that, given technology and physics we actually know works — if we were to make that particular machine, we can demonstrate that it would have this particular properties.

So even though we might not have... a Dyson sphere surrounding the sun, we can actually prove things using very standard physics, and tell them what they would be doing. And that gives us a bit of knowledge about what's possible.

It doesn't tell us what will be done. We can't tell whether we will eventually get to mature and modern technology, or build an Dyson sphere around the sun. But it can show the properties they must have. And at least show some of the upper or lower bounds of what they could achieve.

So this is one of the ways I'm using the book, to look at the possibilities of the future. Trying to see both what looks like is ruled out by well-understood laws of physics, but also things that it would be exceedingly weird if they were not possible. Because we are already doing smaller versions of it, and it's a matter of scaling it up if we really, really wanted to.

Julia Galef: Interesting. So does this seem ... I mean at least to my naïve eye, this seems non-trivial... or, it seems non-obvious and interesting. But is it also actionable? Are there things that we would do differently now, because we have deduced that certain technologies are physically possible?

Anders Sandberg: The most obvious thing is the question of should we be spreading out into space, and how quickly do we need to do that? So there are two aspects. The first one is of course, could you actually settle space?

And the second part is, well, how quickly do you need to do that? Because the remote galaxies are becoming more remote every day. The expansion of the universe means that there are parts of the universe that we can never reach. And if we wait too long, we will never be able to reach many remote parts of the universe.

If there is some value in getting there, we might need to start very early. So in this case you can apply what we know of astrophysics and relativity theory, and that tells us a bit about the speeds of spacecraft, and we can evaluate for example how long we can afford to wait.

And it turns out that if we wait more than a few hundred billion years, essentially all galaxy clusters are totally separated. We will never get outside our own cluster. That still means that, well, if you started within a few billion years, we can get quite a bit of the universe, even though we lose about 17 galaxies per year.

Now whether that is really good reason to start early or later, depends very much both on your value theory — How much do you think you lose, by losing these potential colonizable galaxies?

... but also, on how much you believe that in the future we can get faster spacecraft. Because it's quite often much better to wait a long while if you think that your technology is going to give you a much faster spacecraft, and then go fast, than step out in a fairly crummier spaceship, and then get overtaken by everybody else.

Julia Galef: I see. So maybe a way to characterize the general usefulness of this kind of theorizing, is that it gives us a better sense of what the payoff structure would be, for different courses of action. Or the possible payoff structure. Like, the possible costs, or upsides and downsides, of colonization, at a certain point or a later point. Possible upsides and downsides of humanity dying out now, versus not.

And then in terms of what we do — we use that deduced potential payoff structure, plus our value system, and then make better informed decisions about what to do.

Anders Sandberg: Exactly. Now it's very useful to understand this structure, because then you can start looking at what things are sensitive to changes in assumptions.

For example, what if it turns out that [expansion] of the universe is slower than expected. That might actually change what we'll be doing. What if we find out some physics that suggest that spacecraft might actually be slower than we expect? In that case, maybe we should cut down on our ambitions, and so on.

Julia Galef: By the way: I think it probably isn't feasible, or worth attempting, to go into a discussion of the technical details of potential space colonization right now in this episode...

But it is probably worth pointing out, because I think this will not be obvious to many listeners, that when you talk about the feasibility of space colonization, you're — I believe — assuming humans being digital. That human consciousness will be, at that point in the future, uploaded onto computers. And that's why the incredibly long distances in space won't be fatal to this idea.

Is that correct?

Anders Sandberg: That is correct. So in my paper with Stuart Armstrong, where I looked at intergalactic colonization, we assume that you would be using digital consciousness as encoded in rather small spacecraft. Of course it doesn't have to be humans. It could be artificial intelligence.

It seems to me after reviewing the literature that you can probably get the solar system with biological humans. It's tough in some of the places, but you can do it. But going to the stars as a biological human, that's going to be extremely tough. So although it might be allowed by the laws of physics, it's somewhat likely that unless the future values of the society doing it really, really demands that its biological people, it's probably not going to be done by bio-humans.

Julia Galef: Yeah. It's funny how much I think this one detail results in people talking past each other. Like, I think people like you who discuss space colonization, often feel that it's not necessary to explicitly specify that you're talking about digital consciousnesses — but that is not obvious to the listeners. And they just think it's completely just a non-starter, that we could colonize the stars as humans. So I always try to call explicit attention to that assumption.

Anders Sandberg: Which is a very good practice. It's a little bit like when people exclaim, "Oh, that is impossible" — to immediately ask them, "In what sense of impossible? Impossible as bound by the laws of physics? Impossible in the sense that, oh, that requires unknown technologies or unknown science? Or we can't do it over the next technology generation?"

Quite often, people move very smoothly between them without thinking, and that produces again a lot of people talking past each other.

Julia Galef: Absolutely. Yeah. And interestingly, an additional in-practice meaning of "impossible" — when people use the word, what they often mean when you push them, is, "I assign less than 20% probability to that."

... Well, sorry, maybe that's not true when they literally use the word "impossible." But when they say confidently that such and such "won't happen," they often mean "less than 20% chance." Which is often in fact what the person saying "this thing could happen" also believes. That it's less than 20%, but above 1%, or something.

So they in fact don't have any real disagreement, but they're using language differently. This is a surprisingly common state of affairs.

Anders Sandberg: Yeah, it's a very good point.

Julia Galef: And yeah, again, I'm sure many listeners will still be skeptical of components of this model. Including whether it's reasonable to think that we can have digital consciousnesses. But I'm just gonna ask you to take that as a premise for the sake of this episode, and maybe we can discuss it on another episode.

So, zooming out a little bit more... There's some people — including, I think one of your co-authors on your most recent paper on long term trajectories, Robin Hanson — who argue that we should be skeptical about our ability to steer the long-term future intentionally.

And a central part of the argument is: past humans mostly have not been able to steer the future intentionally. So if we think that we can, that's a little bit suspicious. What makes us think that we're in such an unusual position? What makes us different from the reference class?

What do you think of that?

Anders Sandberg: I think that Robin is right about the general set of humans. Because indeed most humans probably don't affect the future in the large-scale sense.

But it's not clear it's true for "all humans in the world" situations. Because in particular we see a lot of path dependencies in history. A fair bit about our society has been shaped by surprisingly small groups and individuals. Sometimes by accidentally or deliberately doing the right or wrong thing at the right moment. Sometimes by having deliberate agendas.

Julia Galef: What would be an example of a small group of humans intentionally shaping the future? Like, in the way that they intended to. As opposed to just doing a thing that had unintended consequences.

Anders Sandberg: So for example, two good examples of groups that pushed society in particular directions, are the Fabian Society and the Mont Pelerin Society. Both of them were fairly successful in pushing for what we would call a democratic socialist agenda, and a Libertarian agenda, globally. And they had deliberate aims at doing this, they had a strategy.

They were probably also somewhat lucky. Because I would imagine that there were probably 10 groups we never heard of, but had similar ideas, and never succeeded. In this case, they actually had the right [strategies] and managed to make the right choices, and got a big effect.

Julia Galef: Yeah, I basically agree with that. And I have other examples in mind that I think qualify.

I think the Founding Fathers qualify as a small group of people that affected the long term, or at least the medium term future — the multiple century long future — in spreading democracy around the world.

And I think there's a few examples of foundations, philanthropic foundations or individual scientists funded by those foundations, affecting the long-term future in really positive ways intentionally. Such as for example, the Roosevelt Foundation, which funded scientists to come up with agricultural improvements that would save lives in the developing world. And one of those people they funded was Norman Borlaug, who created the green revolution, by creating crops that were much hardier and could feed more people.

Anyway, I've heard people like Robin make this argument before, but I find it a little confusing. Just because... I'm all for looking at track records to reason about what's realistic for us now to expect, but there just seem to be some really important and obviously true differences that make our generation special.

Like, past generations didn't have existing, or near term likely, technology that could dramatically wipe out civilization. And by “wipe out” I mean literally render extinct. Or even more plausible than that, just decimate.

That's not a thing that was true in the past. So it doesn't seem unreasonable for us to think that we're in a special position where we can do things that would make those technologies more likely to impact humanity, or less likely. Does that make sense?

Anders Sandberg: Oh yeah. Of course being able to wipe out an entire future in some sense is one way of making a very big mark in history.

Julia Galef: Yes, absolutely. And then reducing that chance counts also ... like, if you think there are things you can do to make that more likely, you should also think there are things you can do to make that less likely. Which counts as positively impacting the future of humanity.

Anders Sandberg: Oh yeah, and I think in general the reason our world might be different might have to do with the causal structure of our current situation.

So when you think about how to affect the long term future, if you're in a very noisy and chaotic environment, you might do something, but other small factors are also going to mess up the processes. So the end result means that you cannot actually push the future in a desired direction because it's just moving along due to all the other influences.

This is a bit like trying to control the weather by clapping your hands, when there are lots of butterflies and other weather patterns going on. You don't have much of a chance doing anything.

In other domains of course, things are very regular. If you move a rock on the moon, it's going to remain in that spot until it gets hit by a meteorite, probably in hundreds of millions of years. So depending on the environment, you have very different chances.

David Christian, a historian who coined the term "big history,” explained that using a metaphor — for once, he used quantum mechanics as an analogy right. Usually when people use a quantum analogy or metaphor, they're messing things up.

But David really made the point well: He said that individually, history's very quantum mechanical. There is a lot of randomness in the interaction, which means that it's very hard to predict anything local. But in the large [picture], many parts of history are actually fairly regular. The growth in wealth, for example, has been exponential for thousands of years.

... So just like quantum mechanics turns into classical mechanics as you scale things up, a lot of local interactions turn into classical history when you scale them up.

Unfortunately that means most of our interactions average out. This is why most of our choices won't change the direction of the future.

However, sometimes we can deliberately scale up the quantum interactions. That's why transistors work... Because we deliberately set up the conditions. So the small causal influences, that are on the quantum scale or individual scale now, can be scaled up.

The interesting thing is, it's not just that we have weapons of mass destruction that might destroy the future. We also set up a lot of new ways of having causal impact on each other, and the future. Some of them are probably just increasing the noise level, but I do think we are also getting better tools for coordination and mass coordination, that are likely to have strong effects on the future.

Julia Galef: Are you talking about the internet, for example? Or something else?

Anders Sandberg: The internet is the most obvious. And again, the internet is not one tool. It's a platform that allows us to construct various tools. Social media are complicated because we have very different styles of uses of social media, ranging from everything from fake news and a lot of noise in popular culture, to various ways of rapidly coordinating people for an emergency, or for solving scientific problems.

The interesting part here is it's so early days. Social media has existed for 20 years. That's inconceivable. It's almost less than a human generation. And it probably takes us a few generations to figure out how to use any tool well. So we should expect actually that the power of social media, to coordinate people into doing various things, is going to grow quite a bit over the coming decades.

And that is interesting, but also very hard for us to predict. But we can probably predict it's going to help groups to coordinate. Some of these coordination activities are going to be adversarial, which might lead to a lot of bad [outcomes], but you're also going to see that most coordination is intended to reach mutually beneficial goals. And we're going to get better at that.

So I think there's good reason to believe that some of these tools are likely to help us actually control parts of the future better. We might also learn more about which parts of the future can be controlled, and which cannot be.

Again, going back to that chaotic versus ordered situation: we can make very accurate predictions about lunar and solar eclipses thousands of years in the future, because of the behavior of classical mechanics and solar systems. But next year's fashion, or next year's stock market — well, that's because there's a lot of very densely causally interconnected

humans trying to outwit each other. It's not going to work out that well to make a prediction.

But we can recognize this difference, and put our money in somewhat safer investments, by realizing that we shouldn't be trusting people making strong predictions about the stock market.

And we might use other data to figure out how to send our space probes, and be fairly aware that this is going to work out well, because the law of gravity is not changing anytime soon.

Julia Galef: Hmmm. Switching tacks a little bit, what do you think about the argument that if we want to affect the future in a positive way, our best bet is not to do anything intentional, but just to cast a wide net, and fund a lot of different scientific research, a lot of different technological development?

On the logic that in the past, looking back at humanity's track record, that is what has caused the world to get better. Not, for the most part, intentional attempts to steer the future. Just people investigating and discovering things. People just trying to create value for the near future. And those things accumulated to increase humanity's capabilities, and quality of life, and so on over the long run.

Why not just keep doing what has worked in the past?

Anders Sandberg: I think there's a lot of truth to that. Except that most of the things that are really beneficial were not short term good. It was not the people making sure that their own garden was well watered or inventing a solution to just their own personal problem. But looking a bit ahead. Actually figuring out more general tools, figuring out scientific solutions to problems that were peculiar, but maybe not that applicable at the time.

So you really want to cast your net much wider than most people normally would do. Because I think most people would solve problems that are close to them. Because they want to get rewards relatively quickly.

The reason you want to cast your net very widely is that generally we are pretty stupid, and the universe is way more complicated than we can get into our brains. So in general we need to develop experimentation in order to gain the information we need to see where we should be headed. So I'm very much in favor of having people cast this wide net, try to invent various things, and try to make things better in general.

But that doesn't mean it's useless to try to think long term. You can start recognizing that some actions do have long term effects. By understanding ecology, we realize that extinction is forever, or at least until we can start de-extincting species which still requires us to store the genetic materials somewhere. Maybe we should get started on that now.

What you really want to actually have is broad understanding, both across time and space. And that's what leads to bigger planning. Because for

example we can look back and say: what information have we been missing the most? What would historians and archeologists really, really wish they had? And it turns out that a lot of every day information from the classical world is just gone, and just remains mysterious to us, because we only have the texts written by the people for other purposes.

So actually saving your bills and receipts and everyday email might actually be a quite important thing we want to store for the long term future. Even though we can't even foresee what the purpose of that is.

But on the other aspect, you also want to know where to strive for, and that requires thinking about fundamental values... Where would we like to end up?

Because that tells you a little bit about where to prioritize your net-casting. You might not know what kind of physics the far future really benefits from. But you might notice that maybe we should develop more efforts of making physics that helps us survive, rather than creating new weapons of mass destruction.

And I think also we need the hope of thinking of the long term future. If we knew that the future would be just like the present, that there was no way of actually getting it better, I think most of us would say, "Yeah, in that case we might want to save it," but we're not going to feel that strongly for it.

But if we have the hope that it could become amazingly much better... that's actually a very good reason, not just to try to reduce existential risk, but also go on to try to cast those nets into the murky waters of knowledge — can I find something shiny here that leads us towards the future?

Julia Galef: I want to pull on that thread, the motivating force of learning or realizing that the future could be very vast, and potentially very positive — how that should affect our actions today?

What would you say to someone who says: "Look, you know, if future generations exist, then of course I care about their welfare, I want them to flourish, I don't want them to suffer, I want to minimize suffering. But I don't particularly care about ensuring that future generations exist. I care about individuals and the welfare of individuals. I don't particularly care about the species."

How do you respond to that?

Anders Sandberg: Sam Scheffler had an interesting thought experiment where he suggested: suppose you know that the month after you died, the world will disappear. In that case how would it change your life?

And he argues in his book Death and the Afterlife, this actually would have a strong effect. A lot of our activities don't make sense unless we assume that we care about the future after us. And not just that there's a little bit of future, but that there's actually quite a lot.

A lot of our human activities are very, very long-term centric. It's not just building cathedrals in the middle of the village, knowing that it's going to take a century to finish, but it's also setting up societies. Again, not just for our children, but children's children. Because we think that matters.

I think it's relatively rare when people actually don't care about future generations coming into being. There's some people that argue that, or they're even [saying] that we should prevent future generations from coming into being because it's bad for them. But most of us tend to assume it is a good thing that there are future generations.

Now whether we want a lot of them, or what kind of life it's supposed to be, depends very much on your value theory.

Julia Galef: So I used to be in the category of people that I was just referring to — the people that feel like "I want individuals to have high welfare, but I don't particularly care about the continuation of the human species."

One crux for me was: do other humans have strong preferences that humanity continues to exist?

Because I do feel the strong moral intuition that I care about whether people's preferences are satisfied. Even if those preferences are about things that happen after their death. And so even if I personally don't feel that I care about the continuation of the human species, it matters to me if millions of other people want the human species to continue.

And it's interesting, it seemed to me just in conversations with people about the long term future, that a lot of people just didn't seem to care. Like, if you ask people, if you talk about extinction risks, a lot of people will sort of shrug and say, "Does humanity really deserve to continue?" Or "Does it really matter once I'm dead?" Or things like that.

But you're suggesting that people's revealed preferences say something different. That they would be spending their time in different ways if they really didn't care about the continuation of the species.

Anders Sandberg: Exactly. And quite often people have very odd claims. On one hand they say they don't care about humanity going extinct at some point, and yet they are very keen on recycling. And again I guess Robin Hanson would say, "Yeah, but the recycling's all about signaling." So maybe that's more about socially than whether the environment survives. Which might be true.

But I think the interesting part here is that people tend to have this weird "far" mode. Psychologists would say, it's a construal theory: When you think about stuff that's outside your normal life, you think about it in a very different way from the everyday things.

Again, getting back to my initial point about us making predictions... When we make predictions about our everyday life, they work out in a

very different way than when making predictions about the future. Especially the abstract future of the species, where it's nobody around.

Even when you start bring it over to something concrete, like talking about, "So, what about my great grandchild? Let's imagine her life in the year 2100." You suddenly make it concrete. A lot of framing and cognitive bias and modes of thinking that we use in everyday life come into play. ...But it's very clear that people, when talking in general about the long term future, unless they're careful, they get a way with a lot of very sloppy thinking. Because most of the time it doesn't matter to be very accurate about the long term.

Julia Galef: Yeah. Can I share with you one other thing that shifted my thinking about how much it matters whether humanity continues to exist over the long haul, and see what you think of it?

Anders Sandberg: Yes. Please.

Julia Galef: It kind of has the ring of fallacious thinking — but it feels intuitively right to me, so I'm quite curious what you'll say.

So basically, when I look back at humanity's past, it seems to me that a large percentage of humans who've lived in the past had pretty rough lives. We didn't have anaesthesia, we didn't have ... we couldn't really treat infections or a lot of diseases. People didn't have a lot of autonomy. There was a pretty narrow set of things you could do with your life. There was a lot of cruelty and violence.

But all of that toil and tedium and suffering and thwarted preferences was kind of, in a sense, necessary to create modern civilization. And the only way that I can not feel really depressed about the suffering that humans went through in the past, is to kind of retroactively "make it worth it" — by causing there to be lots of future generations of happy humans. Not just happy, but flourishing, thriving humans.

Or to put it in a different way, it just seems like such a shame if humans went through all these generations of living rough lives and got to the point where they had almost made it possible to bring into existence many more generations of flourishing humans, but then they just stopped.

To give an analogy to maybe make this a little more clear: let's say you're an adult and you're pretty unhappy with your life and you're considering committing suicide. And then you think back to your parents, and maybe your grandparents, and you remember, "Gee, they sacrificed a lot and scrimped and saved and gave up a lot of their own dreams, in the hopes of giving me a life where I had the possibility to do great things and be happy."

And that's already a sunk cost. They've already spent that time, that sacrifice. But isn't it just kind of a shame if I don't try to make the most of that sacrifice going forward? Instead of just giving up. And so that's sort of how I feel about humanity at large.

And the reason I say it has the ring of fallacious thinking is that it's kind of a sunk cost fallacy — but I think I might endorse it, in this case. What do you think?

Anders Sandberg: Yeah, it's a bit like the hope that the future will redeem the past.

Julia Galef: Yeah, exactly.

Anders Sandberg: It's actually quite interesting to think about flipping things around. So we can imagine a history that gets better and better and better until it eventually ends. And then another history that starts out in a golden age and then gets worse and worse and worse.

Both of them of course have the same amount of goodness in them. But I think most of us would still say the first one's better than the other one. And I wonder why. It might be that actually it's always enjoyable if the future is better than the past, because we're built like that. We don't like disappointment.

If I get an ice cream and it's worse than expected, I'm actually feeling much more worse, rather than just the sheer utility of that slightly crummy ice cream. And the same thing might be about the future. We might to some extent, need the future to increase because that actually give us something extra. The expectation that the future is a good and better place is super important for us.

Julia Galef: So zooming out again, I just want to try to ... I want to see if I can get us to summarize the ways in which it's valuable to theorize about the future.

We talked about the how there are useful things we can deduce, just from the laws of physics, about what technologies will be possible, and what that implies about the potential gains to reap from the colonization of space, depending on when we start.

We also talked about theorizing about possible really good futures, and how that might motivate us. Why that might give us more motivation to try to ensure the continuation of humanity.

I'm sure I'm missing at least one or two, but what would you add to this general list of ways that it's valuable to think about the future?

Anders Sandberg: I think it's also useful for giving some of the planning that might actually have a long term effects.

So there is this concept of a long reflection, which I think is really important. It's Will McCaskill that suggested this. Maybe we need to get our act together before settling the universe, to figure out some ground rules, some overall goals. Because there might be only a particular window of time where we're all causally connected. Once we go off to distant stars, we're no longer going to be able to coordinate. So if there are things that we need to coordinate, we need to do that before we get out of reach.

So I mention earlier that we might go so far away that in a few hundred billion years it's absolutely impossible to send out the signals. But probably even long before that it's going to be hard to get everybody together.

Right now we're a third of a second away from each other on this planet as the electron flies. And it might be that we need to solve certain problems, figure out some philosophical ground rules, or even set some legal or technical ground rules, in order to ensure that we get a good future.

So that's a choice point we need to be aware of, that it's approaching. Not approaching super fast, of course, but... at some point it's going to be too late to coordinate. So before that we should have at least made a serious attempt at doing that coordination.

What does that coordination require? We don't know yet. We need to figure that one out even before we start coordinating.

So sometimes understanding the structure of the future, like in this case the causal structure of expanding space time, tells us something about what agenda we need to set.

Julia Galef: Interesting. Okay, well that's probably as good a place as any to stop — but Anders, before I let you go, I wanted to ask you my favorite question to ask my guests at the end of an episode, which is: Is there any book or article or blog, or even just a person or thinker, that you have disagreements with, but you nevertheless think is valuable to read or engage with? Anything come to mind?

Anders Sandberg: Well of course, with the default answer will always be Robin Hanson to that question, no matter who you ask it to!

But I also think of Gustav Arrhenius, who is also sometimes my boss at the Swedish Institute for Future Studies. He's written a very nice paper about life extension, where he has argued that basically there's no point in doing life extension from a population ethics perspective, because you get other people.

Julia Galef: Sorry, when you say you “get other people,” you mean... his point is that there's nothing particularly good or important about extending the lives of existing people? It's just as good to have those people die and then create new people?

Anders Sandberg: More or less. He's got a few subtleties, and analyzes ... he's a proper philosopher unlike me. And I disagree with him, and I think it's important to engage with that as a consequentialist pro-life existential person like me. He's also of course justly famous for his Impossibility Theorem on population ethics, which is a real headache inducer.

Julia Galef: I agree. I think we've discussed that paper on the show before. It's one of my favorite pieces of philosophy, although yes I agree, it's a real headache

inducer. So we'll link to that and we'll link to ... sorry, was it a paper that he published with the argument against life extension?

Anders Sandberg: He's got a book, but the paper I'm thinking of is Life Extension Versus Replacement. Which was in the Journal of Applied Philosophy, in 2008.

Julia Galef: We'll link to that as well, and to your most recent paper on long term trajectories. And of course to your blueberry earth paper as well.

Anders Sandberg: Of course.

Julia Galef: Anders, thank you so much for being on the show. There are a bunch of dangling threads that we should talk more about in future episodes, but I'm so glad we got you on for an inaugural Rationally Speaking conversation.

Anders Sandberg: After all, I think we're going to have a very long future. It's going to room for so many interesting questions.

Julia Galef: So many episodes.

Anders Sandberg: So many different chapters, so many different casts of blueberries.

Julia Galef: Excellent. I can't wait. Well: this concludes another episode of Rationally Speaking. Join us next time for more exploration on the borderlands between reason and nonsense.