#253: Intellectual honesty, cryptocurrency, & more (Vitalik Buterin)

Julia Galef: Welcome to Rationally Speaking, the podcast where we explore the borderlands between reason and nonsense. I'm your host, Julia Galef, and today's guest is Vitalik Buterin, the creator of Ethereum, an open source blockchain platform, and its corresponding currency Ether, which is the second biggest cryptocurrency in the world after Bitcoin. Vitalik came up with Ethereum eight years ago, when he was 19, and the year before that he co-founded Bitcoin Magazine, the oldest publication devoted to cryptocurrencies.

But the reason I started following Vitalik a few years ago, and reading his blog is because he's also a really sharp and insightful thinker about politics, economics, rationality, how to improve the world. And even though I'm not that into crypto myself, I came to really enjoy reading Vitalik's public communications as a leader of Ethereum because – as we talk about in our conversation – I find his leadership style refreshingly nuanced and intellectually honest.

So that is one of a wide range of things we talk about in this episode and I hope you enjoy it! Here is my conversation with Vitalik Buterin.

[*musical interlude*]

Julia Galef: Well, Vitalik, let's start by talking about your most recent blog post, which is about something I've been personally very interested in recently. That is: Why prediction markets seemed kind of disappointingly irrational in predicting the results of the last election.

Could you just summarize what the irrational behavior was? What is the mystery in need of an explanation here?

Vitalik Buterin: Sure. This last election, I've been following the prediction markets pretty closely, since around the start of September or so. And even before that, on and off, but from September really intently.

And the thing that immediately struck me is just this divergence between the percent chance that Trump will win, according to basically all of the smart people that I follow on the Internet and on Twitter, versus the number that the prediction markets gave me.

And I had some different kind of ideas in my head about what that difference could be. It could be the prediction markets being wrong. It could be the experts being wrong. It could be people just underestimating a 20% probability that Trump kind of wins the

election, but does so using some completely unfair trick, involving the Supreme Court or whatever.

So this was a big mystery for me at the beginning, and I had all these different theories.

But then my surprise really increased drastically after the actual election. After the election, within a couple of days the mainstream media outlets declared Biden to be the winner. Some foreign governments started congratulating him. But on the markets, the price that Biden would win was still only about 85 cents. So about an 85% chance that Biden would win -- and the 15% was not "other." The 15% was explicitly Trump.

Julia Galef: Right, right, and it also didn't budge very much in response to things.

Vitalik Buterin: Yep, exactly. I think the 15% chance did seem maybe kind of reasonable for about the first week or so. But then Trump made a challenge; the challenge got rejected. Trump made another challenge; the challenge got rejected. And there were just all of these rejections, all of these smoking guns that people on the side optimistic about Trump kept predicting, that just never ended up actually happening. And eventually got rejected by the Supreme Court.

And just after weeks of this, the price just, it stayed at 15 cents. And you know, what the heck is going on here?

So yeah, this was of course when I started really taking the plunge, and basically just looking at this as an opportunity to kind of test out the prediction markets and to try kind of betting against Trump myself.

Julia Galef: And could you actually take a step back for a moment and explain why it would be surprising if prediction markets were just really bad at giving the right answer? Like, really systematically biased in one direction. Why would that be surprising?

Vitalik Buterin: The usual argument here is basically like: If the price is wrong, then anyone who thinks the price is wrong could come in and profitably participate.

If you think the chance that Trump is going to win is, say, like 40%, but the market says it's 60%, then you can go in and you can basically buy tokens that, from your point of view, give you a 60%

chance of getting a dollar at a price of 0.4 -- which is a very good deal.

So the surprise just comes in the form that you have people who are clearly certain that the chance that Biden is going to win is really high, so why aren't they taking this offer?

Julia Galef: Right. And what was your conclusion, after placing some bets yourself and looking into it?

Vitalik Buterin: Placing the bets on the markets definitely turned out to be tricky in a bunch of subtle ways. So, people have already talked about this in response to some of the more traditional kind of non-crypto prediction markets, in terms of PredictIt and things like that.

But there were very specific explanations that had to do with very specific details of PredictIt. For example, people criticize the $850 per person limit a lot. Because the theory would be if there is someone who is very smart, is an expert at predicting, and has the right idea of the probability, and they want to make some money on the markets -- well, they could, but they'd only be able to put in 850 dollars. And so they would have a very hard time kind of counteracting all of these overoptimistic voices that were betting in his favor.

But crypto markets don't --

Julia Galef: About PredictIt in particular, was there also a limit on how many people could bet on a particular question, or was it just a limit on how much money each particular person could bet?

Vitalik Buterin: I don't recall seeing a limit on how many people could bet. I remember there was a per-person limit, and I also remember some of the markets had very high fees. If there's a withdrawal fee of 5%, there's no way at all to profitably push a price above 95 cents for anything.

Julia Galef: Right, so all of these artificial constraints -- which I think are legal in nature, right? Like the US has laws saying you have to limit... yeah, so all these artificial constraints limit the extent to which, I guess, the smart money can come in and correct the dumb money. To put it plainly, yeah.

Vitalik Buterin: Yeah.

Julia Galef: You were about to say, the crypto markets, though, don't have these legal constraints.

Vitalik Buterin: Right. The crypto markets, the fees are low. Anyone can come in, anyone can participate, anyone can bet as much as they want.

So the fact that the prices even on the crypto markets were so far off from what the actual probability seems to be, that was surprising.

Julia Galef: Yeah.

Vitalik Buterin: Yeah, like if there was some market inefficiency that was specific to PredictIt, and some of these more legal constraints, they should not appear on Augur and Omen and so forth. But they did, and the prices were the same.

This is where some of my own adventures came into play and really helped to sort of flesh out my own picture of what was going on. I ended up actually putting about $300,000 into buying No Trump tokens, and I bought them at an average price of around 85 cents --

Julia Galef: They pay out for a dollar if Trump lost, and you paid 85 cents to buy them. So it's a return of 15 cents, if --

Vitalik Buterin: Exactly. And it seemed like a very good deal, right?

But then in the process of buying, and in the process of holding these tokens for a couple of months, I did come to realize what were some of the reasons why not many other people were following me through the same process.

One of them for example is just capital constraints, right? In order for me to win 50,000 dollars, I had to put in 300,000 dollars. And in order for someone on the Trump side to cancel me out, they only had to put in 50,000 dollars.

Just putting money into things is not free, because putting money into things has opportunity costs.

Julia Galef: It's kind of an asymmetry against the people who are betting on the common wisdom, essentially. Because to make a profit -- because the odds are already in your favor, if you think they're not in your favor *enough*, you have to put in a lot of money to get a return on that. Compared to the people who are betting against the common wisdom.

Vitalik Buterin: Yep.

Julia Galef: Do you think that that is... Maybe I'm getting ahead of ourselves a bit, but do you think that is the main explanation for why prediction

markets have seemed to be so reliably tilted in Trump's favor? Or is it something about, like... I could imagine other explanations. Like, maybe the Trump fans are just especially motivated to bet on him as a way to show support, or something like that.

Vitalik Buterin: I feel like it's a couple of different effects coming together at the same time. First of all, the Trump fans were definitely motivated. But second, there's definitely this kind of bias away from certainty. Like, you know, the market's biased away from zero; it's biased away from one.

We saw this, I think, even a bit earlier, with Andrew Yang getting up to about a 5% chance of winning at one point. And I love Andrew Yang, but there's no chance he had a one in 20 chance of becoming the president.

Julia Galef: When was this? Was this during the primaries or was it after he was already eliminated?

Vitalik Buterin: No, fortunately, it was not after he was eliminated. It was still during the primaries.

Julia Galef: So I can cling to some shreds of hope in prediction markets, then!

Yeah, so your conclusion then, about the crypto markets, was that in fact, things like Augur and Omen do have not the same kinds of artificial constraints as PredictIt -- but they have their own kinds of constraint in terms of making it technically difficult and expensive to participate.

Vitalik Buterin: Right, exactly.

Julia Galef: Is that a sufficient explanation for what we saw?

Vitalik Buterin: I think so, yeah. It's a combination of capital constraints and just technical difficulty. Because, the other thing I talked about in my post is how within the crypto ecosystem I didn't have any dollars, I only had ETH. And I wanted to keep my ETH because I didn't want to miss out on the possibility of the price of ETH going up. And so I had to go into another contraption that locked up my ETH into a contract and let me borrow some DAI, which are crypto dollars, and then convert them, and then convert them back.

And so there was this long kind of chain of gadgets that I had to walk through. And the ability to walk through those gadgets and do it all correctly is definitely just technically not in a lot of people's reach.

So I feel like, yeah, the capital issues and just the technical difficulties kind of compounded each other.

Julia Galef: I guess I'm wondering what you think is the right solution going forward, given that there are these sort of fundamental frictions involved in prediction markets that hinder their accuracy.

Vitalik Buterin: I feel like we should just continue with all the experiments. I think Metaculus is great. I think the crypto prediction markets are great. I have a feeling that some of these systems are just naturally going to get better on their own over time.

One other issue I think is just that people have this aversion to doing weird, new things. And especially weird, new things that require them to stick a whole bunch of money in.

And so, as prediction markets become less weird and new, more people from the outside will be willing to participate. Existing participants who have a history of being accurate would have a larger share of the pool in the future rounds, and the system would kind of warm and get better over time.

Julia Galef: I guess it seemed in your post like you were maybe conflating these kinds of psychological features of the human brain -- where we just don't do things that are knowably good for our goals... like, we procrastinate, or we don't do even basic research on the average salary in the career we're going into, before we embark on that career path. Things like that. There's plenty of examples, in my opinion, of humans just not being very strategic...

So it seemed like maybe you were kind of conflating that with the “epistemic modesty” explanation, where smart people who think the market is being irrational look at it, and they assume, "Well, but the market is efficient, and so if I think that I can beat it, I'm probably wrong, and therefore I won't try to beat it."

Do you think those are two separate explanations, and is one of them the one you meant to point at more than the other?

Vitalik Buterin: No, that's a really good point, actually.

There is definitely a difference between just seeing that the way the mainstream world does things, and the way that your own direct, explicit thinking leads you are different, and then just admitting that, yeah, your thinking is probably wrong... versus thinking that your thinking is probably right, but still not doing anything about it.

Julia Galef: Right, exactly.

Vitalik Buterin: Yeah, no, that's ... yeah, I agree. I definitely did not make that distinction and I probably should have. If I had to guess now which one of those that would be, I would probably say some of both.

But the other thing of course is that human brains don't have an explicit distinction between epistemic subroutines and goal satisfying subroutines. The two things do kind of float together into each other.

Julia Galef: You know, this question -- of the efficient market hypothesis and how much to rely on it -- what it reminds me of is a debate in the rationalist community a couple years ago about whether we as a community should have been better at recognizing early on that bitcoin was worth investing in.

And to be clear, there were a lot of rationalists who did get rich on cryptocurrencies. A lot more than the base rate in the population, or even than in Silicon Valley, I think. But there were also a lot of people who looked at bitcoin and were like, "Nah, it seems really unlikely that I could outsmart the experts here, so I'm not going to bother investing even a little."

I'm curious, actually, do you think that was a reasonable inference for someone to draw at the time? Or do you think there was something about bitcoin early on, that a rational person should've been able to tell, "Here's a case where I can beat the market," or "where the efficient market hypothesis doesn't hold"?

Vitalik Buterin: It's actually kind of funny that you bring up rationalism, because I remember even on 2014 there was this thread on, I think, LessWrong, where I just made a post... I think if you dig hard enough you can find this, where I just said, "Hey guys, bitcoin has at least a five percent chance of taking over a significant portion of gold market share, and so therefore its expected value is somewhere in the tens of thousands of dollars, and right now it's a few hundred, so this is why ... "

Julia Galef: I did not know you posted that! I knew Gwern did, and maybe someone else. I didn't know that you were on there.

Vitalik Buterin: I'm not going to derail our time and look for that post right now, but I'll look for it after.

Julia Galef: I'll find it and link to it, yeah.

So you do think, I guess then, that this was a knowably… that this was a case which a reasonable outsider should've been able to tell was an exception to the efficient market hypothesis?

And, what's the rule?

Vitalik Buterin: I would say so. Bitcoin and cryptocurrency are very interesting, because buying bitcoin in the 2010s is like buying Google shares in 1998, except all along throughout the entire decade you have people telling you that “Buying bitcoin right now is like buying Google shares in 1998.”

Julia Galef: But like, you still have to be able to say why this is actually a case where buying bitcoin is like buying Google shares in 1998…

Vitalik Buterin: Right, exactly.

Julia Galef: … as opposed to all the other cases where someone is claiming, "Oh, you should buy this penny stock or whatever, it's like buying Google in 1998." How is an outsider supposed to know that this is a case where it's actually true?

Or not know, but have enough confidence to be worth spending the time and effort.

Vitalik Buterin: Right. I think the argument that I made, either in my post or just around that time, was just like: If you just examine the class of new and interesting things that are arising, then there's a fairly small number of members, and cryptocurrency was one of them. And in terms of just market share of interestingness, or kind of potential future importance, it felt like cryptocurrency was already at a couple of percent, but its market share in terms of monetary value was like 0.00-something percent.

Julia Galef: Wait, what's the unit here, of interestingness?

Vitalik Buterin: I don't know how to define this…

Julia Galef: Okay, you're just saying the comparison of scale – it’s an order of magnitude more interesting than it is valued, yeah.

Vitalik Buterin: Yeah.

Julia Galef: Okay, and you don't think… Yeah. I guess I feel like “interesting” is doing a lot of work here.

Vitalik Buterin: I agree. I remember when I made this argument, I think it was on the LessWrong thread, there were people accusing me of playing reference class tennis.

Julia Galef: Yeah, so do you think that interestingness is... Let's say, take someone who doesn't have your particular taste for interestingness; is there an outside-view kind of way they could get to the same endpoint? Like, by saying, "If such-and-such kinds of people find this thing interesting, then that should be a really good sign?" That it's not necessarily guaranteed, or anywhere close to it, but it's risen above the threshold of worth investing a little bit of money in, or something?

Vitalik Buterin: I think I was thinking of the class of things that tech people are excited about, in general.

Julia Galef: Yeah, I mean, are there any other things in that class? I'm just trying to think of other examples to test this out.

Vitalik Buterin: Right. What was there? There's artificial intelligence, virtual reality... what other things would people get excited about? Self-driving cars, that's a part of artificial intelligence. Yeah, it felt like, I guess-

Julia Galef: So it is a mixed bag in terms of how... well, it seems like a mixed bag anyway, in terms of how well they pay out.

Vitalik Buterin: I agree.

Julia Galef: But which is fine. Like if you choose over the course of your life 100 things to invest a little money in, and two [edit: ten] of those things become big, then as a whole that set of investment decisions seems pretty great. So maybe it's fine that it's as mixed as it is.

Vitalik Buterin: Hmm.

Julia Galef: Also, to be clear, I should say there are two separate criticisms you could make, or that were made, of people in the rationalist community who had heard of bitcoin and failed to invest in it. There's the epistemic criticism and there's the instrumental criticism. We've been talking about the epistemic criticism, of people who thought it wasn't worth investing in and therefore didn't invest.

And then the instrumental criticism is: There were a bunch of people, I know some personally, who thought it was worth investing in but just never got around to it. That's more of a failure of

instrumental rationality where, you know... there were some trivial inconveniences. It was a little bit of a pain to figure out how to invest. But if you actually thought it was a really good opportunity, then it should have been worth it to you to spend a couple hours figuring it out, or ask someone. And so that's a failure of instrumental rationality. Or so the argument goes.

Vitalik Buterin: Yeah, no, I think that's fair. And yeah, I mean, all of these events definitely convinced me that the world loses a lot just because there's so many people that aren't willing to just take that final leap from, “The math says that it makes sense” to actually internalizing that and doing something out of it.

Julia Galef: Let's talk about another one of your recent blog posts. This was on concave versus convex dispositions, which... I'm a sucker for classification systems that point out some underappreciated axis along which people vary. And even more so if that axis is about differences in people's thinking style, or world view, so I'm excited to talk about this.

Could you just briefly explain what is the concave versus convex distinction?

Vitalik Buterin: Sure. The distinction is basically that if there's a trade off between option A and option B -- where option A has some advantages and option B has other advantages, basically -- how do you view the idea of going to either one of those extremes, versus the idea of doing something in the middle?

So, the reason I use the word convex and concave is that, if you make a graph and the horizontal axis is what strategy you choose... and you assume that there is some way to kind of interpolate between two strategies; so one example would be if A is being a vegan and B is being a carnivore, then 50/50 would be having a steak and a salad as a meal.

Horizontal axis is your strategy between A and B, and then the vertical axis is basically how well off you are as a result.

Julia Galef: And that could be defined in a bunch of different ways, right? Like the nutrition value of your diet, or the animal welfare benefits of your diet.

Vitalik Buterin: Or how much you enjoy it. All of it, you know.

Yeah, and so the two shapes that that graph could have -- one is this kind of U-shape, where it's kind of better on the edges and worse in

the middle. Now, it could still be much better on one edge versus the other edge, but you know, the middle is just not a good path to take regardless. And in math we call those convex functions, right? For example, y = x^2 is a convex function. It's sloping upwards.

And then the other shape the graph could take is the lowercase-N shape, right? And this is what we call concave. The square root would be a concave function. And so the general mathematical way of thinking about this is if you have A and B, and you had to choose between a compromise between A and B or a coin flip between A and B, would you take the compromise or the coin flip?

Julia Galef: Oh, I actually didn't realize -- I think I thought you were saying that the convex person has a strong preference for A over B, and so they wouldn't necessarily go for the coin flip, they would... ok, go on.

Vitalik Buterin: Yeah. Like, basically, the convex world... it is a very kind of second order thing. Actually, it reminds me a bit of that hilariously funny post from Slate Star Codex, In the balance, where he just goes through that kind of series of gods... where you start with order versus chaos, and then do you pick the extremes or do you pick the middle? And then you kind of go down the hierarchy from there.

This is very second level. It's, "Do you take the edges or do you take the middle?" And yeah, I go through some examples of things that people with a more concave mindset would say, things that people with a more convex mindset would say, even kind of institutional arrangements that are more concave or more convex. Some situations where being more concave or more convex makes sense.

Julia Galef: Could you give an example of a time when the convex approach seems to be the best? That might be... Well, maybe it's my concave bias talking, but I was going to say that that might be less intuitive to everyone.

Vitalik Buterin: Yeah, no, it's a good point, and I did give a couple of examples in the post. One example I gave is just going to war, because when you're in a war, generally, you either want to concentrate your soldiers and invade on location A, or you want to concentrate your soldiers and invade at location B, right? And if you just do both at the same time, then, well, your opponent is going to go off on one side, pick off your soldiers there, then go off on the other side, pick off your soldiers there, and you lose.

Julia Galef: Got it, so it's sort of situations where you fully commit to one thing rather than halfheartedly committing to a bunch of things.

Vitalik Buterin: Right. And even the fact that you are fully committed to one thing is perhaps even more important than which thing you commit to. That's extreme kind of convex, right?

And one of the comments I made is that more centralized styles of decision making can make sense in contexts where your world is very convex. And militaries do actually tend to be centralized.

Another more relevant example to very modern times I gave is travel lockdown, with the coronavirus, right? If you don't restrict travel at all, then you have the virus but things are convenient for people. If you restrict travel by 50%, then things are less convenient for people, but really you still have basically just as much coronavirus. But then if you restrict travel 100%, then things are less convenient, but then you actually have no coronavirus, right? And that last 5% of travel restriction is as useful as your first 95%.

Julia Galef: As someone who took only two thirds of her course of antibiotics for strep throat, at one point in college, I can attest to the value of the 100% or nothing approach.

Vitalik Buterin: Yes, indeed. Yeah, no, so the other kind of, like the one piece of political commentary I gave there is that a lot of the political regimes that did not fare too well with the coronavirus, at least during the first year, were regimes that were kind of designed around compromise and designed around making decisions that don't kind of go against the wishes of either side too badly.

And you know, I also definitely have the kind of concave bias and so I definitely think that that kind of thing is generally a really good thing to have -- but in the case of the virus, it wasn't.

Julia Galef: You know, one thing I was going to ask you about the convex versus concave distinction -- which I now think is probably the wrong question, but -- I was going to ask whether this is just kind of a subset of nuanced thinkers versus black and white thinkers. But that was the question that formed in my mind when I still thought that convex thinkers were the people who were like, “100% this view is the right view,” or “100% this policy is the right policy,” and no compromise is acceptable. That's my sort of archetype in my mind of a black and white thinker.

But actually, the thing you're describing, of someone who says, "Whichever one we do, let's fully commit to it, and I'm less particular about which one it is than that we fully commit to one thing"... I guess I'm less confident that that is an archetype. Is that a type of person you are familiar with?

Vitalik Buterin: You're definitely right that that's not as much of an archetype --

Julia Galef: It can still be a way of thinking, though.

Vitalik Buterin: Right, exactly. I think that in practice, especially the convex people that I'm criticizing, they tend to be people who do also have an idea about whether option A or option B is better, and they have a strong idea of whether option A or option B is better, and so in their mind it's basically pure option B and everything except for that is irredeemable. Whereas if you take a concave person, even if they have a very, very strong view about, say, option B being better than option A, they'll be willing to consider 90% or 95% solutions. Yeah.

Julia Galef: You mentioned in your post that you see the Ethereum community as being more concave, and the bitcoin community as thinking in a more convex way. Why do you think that is? Is it something intrinsic to Ethereum versus bitcoin? Or is it more about the personalities of the founding leaders, or what?

Vitalik Buterin: There's definitely psychological things that are attractive about, say, bitcoin, intrinsically, to certain kinds of people. I think bitcoin is institutionally convex in some ways. On the block size debate, it took the very, very left side of, "We should minimize block size in order to maximize the ability to read the chain. And if it takes like $500 to write through the chain as a consequence, we just have to live with that, and everyone can just use some kind of second layer system."

And then I think there's also the fact that bitcoin just has less functionality, and so in order to be a bitcoin focused person, you have to really believe in the specific types of functionality that it does provide. And store value is a really big piece, right?

And then the third thing, of course, is that these ecosystems don't exist in a vacuum. They tie into these kind of mainstream political narratives, and bitcoin definitely has a lot of the especially more extreme libertarians. It had a lot of the kind of end-the-Fed people. And it was very much a reaction to statism and central banking and all of those things.

Whereas Ethereum is kind of weird, right? Because Ethereum, there's elements that are a reaction to those things, but there's also elements that are a reaction to bitcoin.

Julia Galef: Oh, interesting.

Vitalik Buterin: Yeah, like even in 2014, the aspects of bitcoin culture that were kind of extreme in that direction, they already existed. And I do feel like it could just have been in part my personality, and in part there were a lot of people who were kind of floating around crypto at the time that were ready to hear the message, but that wanted something that's more moderate.

Julia Galef: More nuanced?

Vitalik Buterin: Yeah.

Julia Galef: I know I've said this to you before, but I really do admire how as a leader you hold yourself to an unusually high standard of nuance, and just intellectual honesty in general. You talk about your different levels of uncertainty in things. You don't pretend that your chosen policies have no trade-offs. You talk about things you think you were wrong about, and so on. And this is very different from what most leaders do, not just in the crypto space, but in general.

Could you talk a little bit about why you chose that path? Was it a goal-oriented thing, like, you think intellectual honesty produces better results when you're leading a group, or a movement, or a company? Or was it more just a disposition, where you just prefer being intellectually honest regardless of the consequences?

Vitalik Buterin: I think part of it is definitely that, in addition to a crypto person, I've always been a rationalist. And so the good rationalist values of not being overconfident, of not being overly dismissive of tribes that you disagree with, of trying to see the best arguments from both sides, like steel-manning people instead of straw manning people, and all of these ideas that we love -- those are definitely things that have had an effect on me from the beginning.

There's definitely a kind of moral aspect to it. Like, I'm not just into cryptocurrency to smash fiat and replace it with something that's not fiat. I actually want to try to make a serious effort to actually make sure that we create a movement that's better. And that's a difficult thing to do. I think even within the Ethereum community, we've definitely not always been successful.

Julia Galef: And how does the intellectual honesty tie into that goal, the kind of higher vision that isn't just about profit?

Vitalik Buterin: I think part of it is a matter of what kinds of people you attract. Like, I definitely, even at the beginning, had an explicit goal of attracting certain kinds of people into the community, and even repelling certain kinds of people from the community.

Part of it was that I think if you do have a more intellectually honest community then you can more easily and more productively interface with outside communities. And interfacing with outside communities was also, I feel like, one of my strong values all the way through. I feel like I've tried to reach out to the rationalist community, different kind of economics and game theory communities, the academic community, the Marginal Revolution type people, Glen Weyl and Radical Exchange type people... And yeah, just trying to make something that had broad appeal to people, I guess.

Julia Galef: And have there been downsides as well to the intellectually honest approach? I know I've seen some media outlets seize onto times you've said, "Well, I'm less than 100% confident on such and such," and the headline will be like, "Vitalik admits he has no confidence in Ethereum," or whatever.

Vitalik Buterin: This definitely happens.

Julia Galef: Well, so, A, how big of a deal is that? Is it just an annoyance that you roll your eyes at, or is it actually a serious downside that just doesn't fully outweigh all the benefits and the personal conviction reasons for sticking to intellectual honesty?

Yeah, I guess -- question A is how big of a deal are downsides like that, and question B is are there any other downsides? Do you feel like it makes you a less effective leader in any ways?

Vitalik Buterin: No, I think those downsides definitely exist. For someone who wants to write a hit piece on the Ethereum community, there's plenty of things to quote mine. But on the other hand, despite the easy ability to write a hit piece on the Ethereum community, not many people have done it.

Julia Galef: It's true, yeah.

Vitalik Buterin: Yeah, no, I feel like there's subtle things about having a kind of friendly approach that do rub off on people.

Julia Galef: Yeah. I've also trawled through some of the comment threads on Reddit or Twitter when you've said something, like, cautioned people at the peak of crypto mania in -- what was it, 2016? 2017?

Vitalik Buterin: 2017.

Julia Galef: 2017, yeah, like when you cautioned people, "Let's not get ahead of ourselves. It's still very volatile. It's still overvalued. Here's why we

shouldn't consider ourselves to have won..." There have been a number of threads like that, where you've said something quite nuanced, or sort of cautioning against hype or against cheerleading.

And it's not that there aren't comments from people saying like, "Hey, Vitalik, that's no way to be a leader" -- but those comments do tend, as far I've seen, they tend to be dramatically outweighed by the comments expressing gratitude to you for your honesty.

Is that also your impression, or do I just have rose-colored glasses?

Vitalik Buterin: I think that's definitely true in a lot of communities. One thing that is hard sometimes is cultural translation. Things that I say get translated into Chinese, for example, and when things get translated into Chinese they can easily lose nuance.

Julia Galef: Oh. Any examples come to mind?

Vitalik Buterin: Trying to think... There was definitely that one time that I said something like, "It is possible that the 2017 bubble was the last really huge one," and this got into the various Chinese chat groups, I think even more than the English ones, and a lot of people did end up interpreting it in a way that caused them to kind of lose some hope. So I definitely feel like I could've phrased that much better.

Julia Galef: Yeah, live and learn.

Vitalik Buterin: Yeah, exactly.

Julia Galef: One thing I've been curious about, as you know, because I've mentioned this to you before, is this disagreement over whether being intellectually honest -- expressing uncertainty, being nuanced, and so on -- whether that has downsides in terms of whether people look up to you, and trust you, and consider you a strong leader.

And there are people who will explicitly disagree with your approach to leadership. Like, people who say things like, "No, no, no, you can never express less than 100% confidence in your claims, or you'll demoralize, or demotivate your followers or your employees."

I'm just curious, what's your model of why they disagree with you? Do you think they have a mistaken picture of how the world works, and they're just wrong about the consequences of expressing uncertainty? Or do you think, for example, they're right about their

audience and it's just that you guys have very different audiences? Or something else?

Vitalik Buterin: There's definitely people who want to kind of feel hope and who react negatively to messages that go against that, even in the short term. I think a big part of it actually is this kind of short-term versus long-term thing. It's very easy to make a message that riles people up, and makes people happy for short periods of time, but then you end up having to pay for it over the next year or so.

And I think I even have made mistakes in that direction. I have made mistakes where I gave overly optimistic predictions about when Ethereum 2.0 was coming out, or when the next hard fork is coming out, or when roll ups are going to be ready -- and for the next month people get excited, but then a year comes and people get disappointed. Two years come and some people leave. And like, a more balanced approach would've lead to a lot of immediate anger and resentment initially, but once the shock is over it's better, because people are on a more realistic page.

Julia Galef: Yeah. That absolutely rings true to me and matches a bunch of other data I've seen. And also I have another theory about what's going on here with this disagreement, and I want to run it by you and see what you think.

I used to kind of just believe the common wisdom that expressing uncertainty makes you a less effective leader, and I was like, "Well, that's an unfortunate trade-off in the world." And then I did a bunch of research and talking to people about this for my book, and now I actually think this is wrong. And that a more accurate summary of what's going on is that what really matters is social confidence. You know, poise, self-assurance, charisma. Willingness to put yourself out there, take risks, try things. That is what determines whether people see you as a leader, see you as competent, want to follow you.

Whereas the explicit credence you assign to your claims actually doesn't make that much difference in how much people trust you or look up to you. And I think a large part of what's going on here is just that people are conflating these two kinds of confidence, social confidence and epistemic confidence.

What do you think about that?

Vitalik Buterin: I could see something like that being true...

Julia Galef: Like, yeah, I was just going to say -- for example, Jeff Bezos and Elon Musk both said multiple times that they're only 10% confident that they're going to succeed, or only 30% confident. And yet no one thinks of them as unconfident men, because they go out there, they take charge, they do things. They're clearly self-assured. They're confident that what they're doing is worth trying, they're confident that they're worth listening to. And the fact that they express less than 50% confidence in the success of a particular venture, it just kind of goes in one ear and out the other, you know?

Vitalik Buterin: Yeah, no, well, Elon actually even went even further. At one point, he literally said he thought the Tesla stock price is too high.

Julia Galef: Actually, Elon is a great example, because someone -- I forget who it was, it was someone who had worked with Elon -- who said about Elon, this was a quote that got spread around a lot, it was something like, "What makes Elon such a success is that he literally cannot conceive of failing."

And Elon has explicitly said the opposite of this many times! Like, he said that he expected the most likely outcome for both Tesla and SpaceX was failure. He just thought that it was high expected value, and so it was worth a shot. He says this explicitly, but it doesn't stick in people's mind, because he's just socially a very confident person. Or, "socially" is not quite the right word for it, but in his actions and the risks that he takes, and what he goes out and does, and just his self-assurance that he knows what he's doing. It's very clear that's what sticks.

Vitalik Buterin: He gives the impression of just being a leader that it feels good to get behind.

Julia Galef: Right, right, and so then they parse that as, "He is 100% confident that he's going to succeed," which is not actually true.

Vitalik Buterin: Right. Like, I think saying that you're 100% confident is definitely one way of getting that effect, but there's definitely other ways of getting that effect. There's aspects of your personality that you can kind of show. There's even countersignaling. This is one of the kind of fun things that I do on Twitter sometimes.

One recent example is some of the bitcoin maximalists, they made this kind of chart where the X axis is your understanding of money, and the Y axis is how smart Vitalik seems, and of course in their mind the chart is Y equals 1/X. Right? Like, if you don't understand money, Vitalik seems smart, if you do understand money, then Vitalik's a complete idiot.

And one of the bitcoin maximalists did this chart, but for some reason they made it be a gif. They made it be like an animated thing where the lines kind of appear in and out of existence, and I just replied, "Hey, guys. I think a gif is the wrong format of this. Here you go, I took a screenshot and helped you make it into a png." And you know, I just put in an image that just has that exact same chart.

And you know, there were people who interpreted that as an alpha move, as the cool people say, so --

Julia Galef: I think the phrase, "an alpha move, as the cool people say," is another good example of countersignaling.

Vitalik Buterin: Yes.

Julia Galef: You know, relatedly, I think it's worth pointing out that expressing uncertainty does not always mean you express low confidence in everything, which I think is a common misconception. You can be perfectly calibrated and still sometimes express very high confidence. You just want to be, really, pretty sure you're right in those cases.

And actually, a good example of you doing this is that time you were at a conference a couple of years ago, and Craig Wright was on stage. Craig Wright being the guy who claims to be Satoshi Nakamoto, the mysterious creator of bitcoin.

And you stood up in the audience and said in front of everyone, basically, "Craig Wright is obviously wrong about X, Y, and Z, and he's a fraud. Why is he allowed at this conference? He's a fraud." It was very striking, it was quite a moment. And I love that in part because there is this common criticism of rationalist types who espouse epistemic humility, which is, "Well, you're just going to be wishy washy and you're never going to take decisive action." I just thought that was a great example of how being well calibrated does not mean being uncertain about everything and never taking action.

Vitalik Buterin: Yeah, Eliezer talks about this a lot, right? He says if the ... what was the statement? It was something like, "If the risk is small, the way opposes your fear. If the risk is great, the way opposes your calm." Something like that. Like, you know, there's nothing inherently virtuous about being unconfident and there's nothing inherently virtuous about being confident. Calibration, is all about being the right one in the right circumstance.

Julia Galef: Right. By the way, do you understand why it wasn't as obvious to some other people that Craig Wright was lying about being Satoshi? Like, Gavin Andresen, was it?

Vitalik Buterin: Mm-hmm.

Julia Galef: Yeah, Gavin Andresen, the guy who vouched for Craig Wright. Do you understand why he disagreed with you, and why the things that were obvious to you about the flaws in Craig Wright's case, why were those not obvious to him?

Vitalik Buterin: My understanding of the backstory is that Craig actually reached out to Gavin in person and did some trick in person where he showed that he could verify the signature. And of course in reality that's an incredibly insecure way of convincing yourself of having some identity. Because there's just so many ways that you can hack into a laptop when you have physical control over it, basically, and so --

Julia Galef: But like, presumably Gavin knew that?

Vitalik Buterin: Right, he knew that. So I think part of it is still just a big question mark.

Part of it, in terms of just the broader community, is that I interpret Craig Wright as being a bit of a Donald Trump figure. To understand Craig Wright's success, you have to understand kind of the cultural context of the bitcoin block size war, where basically you have the small blockers and the big blockers.

And this is of course my point of view, which totally disagrees with other people's points of view. But to me, the big block side, which emphasizes the ease of being able to use the blockchain, is just obviously closer to what people actually wanted from bitcoin all along. You know, it was peer to peer cash. You're supposed to be able to actually send it and actually use it as a currency.

And it felt like for the first few years, people just broadly agreed that this is the thing that bitcoin is supposed to be, and there was a block size but obviously that block size would be increased over time. But then around 2014 and 2015, suddenly, you had these kind of core dev, technocrat expert type people come along and say, "No, no, no, actually, we can't do that."

And they did use these fairly sneaky tactics, like they first agreed to a compromise of increasing it to two megabytes, and then four, and then eight. And then after that they had another meeting where

they ended up explicitly disavowing a hard fork. A lot of people on the big block sites just ended up feeling very upset and hurt as a result of this; but at the same time, the people on that side, they just did not have as strong an intellectual and technical capacity.

Then Craig Wright came along, and Craig Wright started basically saying the words that they wanted to hear. He said things about how, "Yes, of course, the small block people are evil, and yes, block size should be increased, not just to 32 megabytes, no, 128 megabytes, no, 512 megabytes."

Julia Galef: Was this like his, "Build a wall"?

Vitalik Buterin: Yes, exactly. It's like the equivalent of Hillary Clinton is Satan, and Barack Obama, clearly something's wrong with his birth certificate, and so on and so forth. Except the equivalence of that in this segment of cyber land.

Unfortunately, I think a lot of the big block people did end up just lapping it up. And I think that's a big part of why Craig just managed to ride the wave and become so successful within the big block community. And it did end up, it takes such a long time for the Bitcoin Cash community to get rid of him.

Julia Galef: So your take is that there was some amount of wishful thinking or motivated reasoning in... Like “He has to be trustworthy, because if he wasn't, then these things that he said that I really want to be true would then be undermined.”

Then once you've concluded he's trustworthy for those reasons -- I mean not consciously, but on some level -- then you already have this prior that he's a very trustworthy and brilliant thinker, and that makes his claim to be Satoshi Nakamoto a lot more credible, given your prior.

Is that the right model?

Vitalik Buterin: I think here, I might have to be the one to distinguish between the epistemic side and the irrational side. Part of it might be that Craig Wright seems more trustworthy, and I think part of it is more like, “I don't really care whether or not Craig Wright is Satoshi, I just care that he is saying the things I like and that he's hurting the people I hate.”

Julia Galef: Ah, I see, I like that parallel. Interesting.

Switching gears, maybe now is a good time to segue into talking about effective altruism. I'm curious about your take on some of the main objections... Well, let me back up:

Your approach to improving the world seems to overlap quite a bit with the effective altruist approach. You've talked about donating to life extension research. I happen to know you've also donated to global poverty charities recommended by Give Well. And you donate to AI safety organizations, right?

Vitalik Buterin: I have, yeah, I gave over $1 million to MIRI.

Julia Galef: Oh yeah. Anything else on your list of main cause areas to which you donate, that I'm missing?

Vitalik Buterin: I think those are the top three. Global poverty and anti-disease stuff is one, life extension is the second, and existential risk is the third.

Julia Galef: Great. I just want to throw a few objections at you, and see what you think.

One is called the cluelessness objection -- not just to effective altruism, it's an objection to utilitarianism in general. And it's basically that everything you do has these ripple effects through time. If you give poor people money, that could make them dependent on you, or it could undermine the local economy, or it could lead to overpopulation, or tons of other things that we can't anticipate. So even if the intervention you want to do seems like... you're a good effective altruist, you've done your Fermi calculations and you looked at the evidence, and it seems to have clearly positive direct impacts. But you don't know what the indirect impacts are going to be, and they could easily be negative and outweigh the positive.

So the cluelessness objection is just, you should have huge error bars around anything that seems like it's a good way to impact the world. Do you ever think about that, and what do you think of it?

Vitalik Buterin: I think my critique of those kinds of arguments is that once you get to things that are very causally distant, like if you have no reason to expect the second or third order effects to be negative instead of positive, then it's probably better to just act like they don't exist. It is possible that if you give someone money, then it'll reduce the motivation that they have, and it'll drive them away from doing some other thing that could've really helped their community. But it's also possible that if you give someone money, then they'll be

empowered and they'll get an education and do something else even better.

Once you zoom out even further, then you get into butterfly effect arguments -- and at the very extreme, like if you clap your hands, then there's a 50% chance you caused World War IV, there's a 50% chance you [prevented] World War IV.

The closer you get to that, the closer you realize that just about all of life is that way. And so the correct thing to do seems to be to just not think about those things in your calculation, and focus on the things that you can measure.

Julia Galef: I guess the cluelessness objection has more force if you were never that enthusiastic about effective altruism or trying to help the world in the first place, and it was sort of... you were choosing between trying to be an effective altruist, or just doing whatever else you wanted to do anyway, that benefits you or other things in the short term.

If the cluelessness objection shows that the effects of EA interventions are less certain than they had seemed to you, then that could justify pushing you towards just doing other stuff that you want to do anyway. But if you really do want to help the world, then it can still be pretty clearly worth it, even though the long term effects are uncertain.

Vitalik Buterin: I think that's a good way to summarize it.

Julia Galef: Cool. You mentioned Glen Weyl earlier; he's been pretty critical of the whole effective altruist, rationalist cluster, both the people and also the ideas. And I'm curious for your take on his critiques.

For example, on the 80,000 Hours podcast a couple years ago, he was criticizing the effective altruist attitude, and he gave kind of a caricature of it, he said, "We're going to deduce exactly what reason tells us is the right thing to do, and then we're going to use our money and power to make the world the right way that it should be, and that's largely going to be based on science and reason, and we're going to impose it. We're not going to pay that much attention to getting feedback from the people whose lives that it affects or being in conversation with them."

I'm not confident I fully understand his critique, but I think it kind of ties in with his critique of technocracy and the idea of thinking that your reasoning is better than other people's, and that it's kind of presumptuous to impose your ideas of how to improve the world

on other people. Does that sound like a fair characterization, and what do you think of it?

Vitalik Buterin: Yeah. And I think first of all, I think there's a lot of ways in which it overshoots the mark, even pretty much immediately.

One example of this is that one of the interventions that effective altruism recommends is Give Directly, which is literally just a charity that gives money to people in poor African villages and lets them do whatever the hell they want with it. That's literally as anti-technocratic as a charity could possibly be. You're giving people raw economic power and just saying, "Go, you guys know what you and your families need better than we do."

That's one thing. The other thing is that I think one of the weaknesses in Glen's perspectives on what the world's problems are, at least at that time -- and this is a critique that I gave even in my review of Radical Markets, I think it was way back in 2016 -- is that he talks about monopoly power, and he talks about tragedies of the commons and public goods. But one type of market failure, or thing that is under-incentivized that he fails to properly account for, is the category that I call entrepreneurial public goods.

What I call entrepreneurial public goods is kind of the intersection of entrepreneurship and public goods. So the core idea of entrepreneurship in this model is... like, you know the famous slogan, "If Steve Jobs just went around asking people in the 19th century what they wanted, they would've just said, 'A faster chariot'"? And the thing that should've been created is to just go and make something completely different, which is a car. Public goods are of course projects that benefit large groups of people, but where each individual beneficiary's share of the benefit is too small for any of the individual recipients to want to finance it themselves.

The intersection of entrepreneurship and public goods is really challenging, because the institutions that both we have, and that Glen provides, for dealing with public goods -- so like, democracy is one example and even quadratic funding is another example -- in some ways they're anti-entrepreneurial. In the specific sense that they favor things that are already popular. They don't have a built-in vehicle for favoring things that are not popular today, but where a few dedicated definite optimists are really convinced that people are going to be very thankful 50 years from now.

I thought that this was this big market failure. Like, when you have something that is a public good and that does basically need entrepreneurship at the same time, then venture capital by itself is

not going to solve the problem. Democracy is definitely not going to solve the problem. And even kind of smarter Glen Weylian forms of democracy, like quadratic funding, are also not going to solve the problem. You need some kind of combination of these ideas to actually get that.

To me, a big part of effective altruism is that it is an effort to identify a specific set of entrepreneurial public goods – so, existential risk mitigation, life extension research are probably two of the important ones. Just even reasoning backwards from what kind of institution would successfully fund things like existential risk mitigation and life extension research, the answer is definitely not just pure democracy. The answer is definitely not going out to diverse stakeholders and asking them what they want.

The answer has to be some form of people being confident and being willing to take a risk on their confidence and doing it, and their just being recognized as having been obviously correct decades in the future. And we don't have that yet.

Now, what kind of institution could come up with that? I think that's a really fascinating question, and we should definitely think more about that, but in the absence of that institution, we do have a few definite ideas of what some important entrepreneurial public goods are, and so we should just start building them.

Julia Galef: I think that's really well-put, and I hadn't framed the ideas in my head that way yet.

But I even feel just more fundamentally confused about the idea of "imposing our vision on the world." The critique of technocracy, or of high modernism, or other terms that get thrown around this discourse -- they make more sense to me when we're talking about policymakers imposing a policy on society that people can't just decide to opt out of. Like, you can't decide to opt out of the education policy that your state or your country has decided is correct.

There, I feel like the worries about trusting your own reasoning too much, and not leaving it up to sort of grassroots democracy, those are very justified worries. But when we're talking about a bunch of private philanthropists who are funding things, or offering people things, while still being constrained by the law, and by people's decisions to opt-in or not -- that seems like much less of a worry about the major failure modes of high modernism or technocracy there, I think.

Vitalik Buterin: I think one counter-argument might be something like Facebook. It's fundamentally a voluntary system that emerged from over two billion people opting into it. But over time it got network effects, there's clearly negative consequences that result from it, and a lot of those negative consequences are of the very diffuse and public variety. And so even if you personally are not part of Facebook, you're still suffering from some of the effects. At the same time, Facebook is trying new technology, and it did at least during its growth manage to not really be hampered by a lot of the regulation.

I do think if you want to steel man it, basically it is a worry about effective altruism creating another Facebook. I think my counter-argument to that steel man would be that just there's no a priori reason for why an altruist-motivated project would be more likely to turn into the next Facebook, in the pejorative sense, than just a corporation. And corporations feel like --

Julia Galef: And we allow corporations. Yeah.

Vitalik Buterin: Exactly, like there's lots of corporations. Like, basically at worst, effective altruism could increase the number of dangerous things in that class from, what, 40 to 41? But the best case is if we can reduce the possibility humanity blows up entirely in the next century from 10% to 8%. And that just massively outweighs it.

Julia Galef: You did a great job giving each argument and the counter-argument to that argument by yourself.

I don't know, I'm probably not as good at steel manning that argument as I should be, just because... I don't know. I get frustrated sometimes in these conversations, when it feels almost... not disingenuous, but it feels like... if people had any particular examples to point to, of effective altruism causing harm, or the EA kind of approach to improving the world causing harm, then I'd be much more sympathetic to the concerns. But it often feels just a bit like something pulled out of the air to object to us.

But I would say that, wouldn't I? I don't know.

Vitalik Buterin: I think one criticism of EA that I have heard -- and I will just say right off the bat that even when I mentioned this to Robert Wiblin, he replied, "Well, actually there's lots of EAs that are taking this seriously and trying to improve the EA movement" -- is that EA tends to focus on improving the world by throwing money at things. But a lot of the most effective ways to improve the world don't necessarily come from philanthropists throwing money at things, they come from institutional change.

One of example of this is, you could have myself and hundreds of people like myself be convinced to throw half our money into life extension research, or we could just talk to a couple of people in DARPA and the Senate and Congress and possibly some EU bodies, possibly some Chinese bodies, and just get $50 billion of government funding at the problem. Wouldn't the second approach be much more efficient than the first?

On the other hand though, I think just to go back to this, to make the argument --

Julia Galef: You don't even need me here at all, you can just do both sides of the conversation!

No, it's great, go on.

Vitalik Buterin: The way that I would counter-argue is that I think if rationalists *were* trying to dip their toes in and change policy, and have DARPA direct $30 billion to things that they consider important, the amount of hate that they get would just be ten times bigger.

Julia Galef: That is another reason I kind of grit my teeth at this objection, is that it feels like we're being put in a bind sometimes, with objections that fully fill the space of possible things we could do.

Vitalik Buterin: Yeah, it's like “Rationalists aren't trying to think about the importance of changing the world through better social structure enough,” and then “Wait, there are rationalists and they are in institutions that are important to social structure, and they're socially adjacent to some terrible people!”

Julia Galef: I feel validated.

I wanted to ask you about life extension in particular. So, there's two common objections I hear to radical life extension, which I'm sure you've heard as well, and I'm just curious how you react to them. The first is that if people start living to be hundreds of years old, then the Earth will be overpopulated, and that's bad.

And then the second is related, it's that if people start living to be hundreds of years old, the pace of change in society will slow way down. Because new innovations, new social mores, they only take hold when older generations die off and younger generations replace them. And so lengthening everyone's lifespan will slow down that very valuable process.

What do you think about those two objections?

Vitalik Buterin: I think there are different levels at which you can respond to this.

I think first of all, one really important thing to keep in mind is that there is often this bias where people just immediately go to worrying about subtle second order things, without even realizing the fact that the first order consequence is just really, really good, and it's so massive in importance that even if the all the second order stuff was as bad as it could possibly be, it would still not be bad enough.

Julia Galef: Do you think that that bias is caused by just reflexive contrarianism? Is it caused by a desire to be clever?

Vitalik Buterin: I think it's some of both. Also, I think humans have this bias where they care more about negative consequences that result from social structure than negative consequences that result in the world. Like, one example of a moral frame that just makes this really obvious is: Imagine a world where genocides happen every century, and someone comes along with the idea of, "Hey, we should stop doing that." And you get your wise philosophers to come along and say, "Well, doesn't a healthy ecosystem need to have both the lions and the deer?"

Obviously, they would in 2021 get canceled out of the room within microseconds, if you say that. But then if you talk about harm that comes from natural consequences, we haven't gotten to that.

That's the first thing, which is just, billions of people not dying is just morally so incredibly huge. Like I think it's important to just mentally stare at that moral fact for a few minutes. And then --

Julia Galef: Well, I'll just add that I think for a lot of people making this kind of argument, they wouldn't actually agree that all these people dying is a terrible thing.

Like, in their model -- even if you think they've acquired this model through some kind of motivated reasoning -- in their model, people getting to live 80 years or 85 years or whatever is important. And if people are not living up to their 85 years because they're dying young due to cancer, or war or whatever, then that's bad. But once you hit 85 years, then you've gotten what is due to you, and it's not a tragedy if that's where things end for you.

Vitalik Buterin: The fascinating thing is how inconsistent our beliefs are about that. In hospitals, people often spend huge amounts of money and resources prolonging their own lives or their parents' lives even by a couple of weeks. Even in contexts where really that's not worth it,

and all you're doing is just spending an amount of money that's worth years worth of vacations for the child, and instead you're giving it to just give the parent an extra four weeks of existence that's basically stuck to a hospital bed, and is not even very pleasant.

You can't really take people's opinions at face value, because just depending on how you ask the question, you do get these wildly different answers. I think ultimately, people do love life, and once you move the problem from far mode to near mode, that just becomes really obvious. That's the first thing.

Julia Galef: I derailed you, yeah, go on.

Vitalik Buterin: I did want to talk about social consequences, because I think it's an excellent example of, actually, going back to the conversation we had about effective altruism, and "Might there be bad consequences if you give people more money?" – like, sure, if you have a lawyer and you give him an hour with the task of coming up with as many bad things as possible that could happen if you give money to poor villagers in Kenya, they'll come up with a lot of reasons. And their case might even look compelling to some people.

But then if you come up with a lawyer and give them the task of, "Let's generate positive second order consequences from giving money to people in Kenya," then their document is going to be even more impressive.

I actually think that with aging, it's very similar. To give some examples: First of all, social transitions are the most risky when they happen very quickly. Like, the internet happened fairly quickly, and a lot of people blame that for being destabilizing. Early 20th century industrialization happened quickly, and a lot of people blame that for leading to communism and fascism and so forth.

But anti-aging is not going to come quickly. It's like, people live for 90 years, and then 10 years later, people would live for a maximum of 100 years. Ten years later people would live for a maximum of 110 years. So the change creeps in slowly, we'll have lots of time to get used to it.

Another thing is that, yes, there are risks from, say, cultural stagnation. But on the other hand, there are benefits from a culture that has 800-year-olds. One is long-term orientation. So like, in the rationalist community we talk a lot about how we don't value our distant relatives enough, and in reality there's these trillions of potential future humans whose lives we might be saving if we can

help the world not blow itself up. And a society with 800-year-olds might be more willing to take that kind of long-term view.

A society with 800-year-olds is also going to have more built-in illegible expertise about previous eras, like where --

Julia Galef: Illegible in the sense that they can't just write down what they know, and pass it on before they die?

Vitalik Buterin: Exactly, yeah. There's things that you can get by actually talking to someone who, say, lived through World War Two, for example, that you can't get just from reading a book about it.

I think another example would be: social systems would be much easier without the need to worry about aging populations, with a smaller portion of the population just needing active support from the government. The political system would have fewer things that it needs to focus on, and that could allow it to a better job at the remaining things it does focus on.

You could make a long list of these things. And I actually think, on the whole, a society with 800-year-olds to balance out the eight-year-olds and the 80-year-olds is going to be a good one.

Julia Galef: And do you worry at all about the overpopulation aspect, or is that part of what you were --

Vitalik Buterin: Yes -- by yes, I mean sorry, I forgot the question. I don't mean I'm worried.

Julia Galef: Right, yeah.

Vitalik Buterin: I don't think so, because I think humanity is great at improving the efficiency of its food production when it needs to.

First of all, you have to remember that for the next century, humanity's main crisis is not overpopulation, it's underpopulation. The charts show that basically everyone except for Africa is going to see their populations go down by 2100. So if, thanks to life extension technology, instead of nine billion people, we have like 15 billion people in 2100, we can handle that.

Then can we go from 15 billion people to even more billions of people? I think we can, and I feel like if we could solve a problem as difficult as fixing aging, then we can definitely find more efficient ways to feed and house everyone. If eventually we have to spread

out to multiple planets, then we could do that, and that's great, that's lots of people's dream all along anyway.

Julia Galef: Cool. On that note, maybe I will finally let you go, since we've been talking for almost two hours. Vitalik, it's been such a pleasure talking to you! I'm so glad we finally got to connect and I got you to come on the show.

Vitalik Buterin: Yeah, it's been great to finally chat, Julia.

[musical interlude]

Julia Galef: That was Vitalik Buterin, and I encourage you to check out his blog at vitalik.ca. I'll link on the podcast website to some of the articles and other interviews we touched on conversation today. That concludes another episode of Rationally Speaking – I hope you'll join me next time for more explorations on the borderlands between reason and nonsense.