Rationally Speaking #214: Anthony Aguirre on "Predicting the future of science and tech, with Metaculus"

Julia: Welcome to Rationally Speaking, the show where we explore the borderlands between reason and nonsense. I'm your host, Julia Galef, and I'm here with Anthony Aguirre, a theoretical cosmologist at UC Santa Cruz. He's the associate director of the Foundational Questions Institutes and co-founder of the Future of Life Institute.

                But today we're going to be talking about Anthony's most recent project, Metaculus. It's a site that aggregates predictions about science and technology. So, questions like, will SpaceX land on Mars by the year 2030? Or will someone in the US be killed by a drone by the end of 2018? Things like that.

                So Anthony and I are going to talk about, why is it useful to aggregate predictions about questions like these? What some of the challenges are of running a prediction engine like this — and some potential solutions, hopefully.

                Anthony, welcome to rationally speaking.

Anthony: Thanks, it's very nice to be here.

Julia: So first off, you know, there are other sites that aggregate predictions in a kind of a similar way as Metaculus, like PredictIt or InTrade. Probably listeners will be most familiar with prediction markets that have markets on, say, the outcome of political elections or a vote like Brexit.

                So how is Metaculus different? What's its sort of unique reason for existence?

Anthony: I would say it's different, both in sort of the mode that it uses — it's not really a prediction market per se — and also the sorts of questions that it addresses. Which are questions of technological and scientific interest, and ones that just are more relevant to the things I've always been interested in. Those aren't generally well treated by betting markets in prediction markets, which tend to focus on sports and politics, and sometimes geopolitics and world events and things. But not kind of on the nitty gritty, and not generally on the scientific or technical side.

Julia: And why wasn't the ideal solution from your perspective to just start seeding scientific and technological questions on existing, well-established sites like PredictIt?

Anthony: Well, I think there are advantages and disadvantages that prediction markets themselves have. So, I mean, there was — part of the motivation was just to have a system that I was actually integrated into and sort of somewhat in charge of. So there's that aspect of wanting something that you can do exactly what you want with.

                But there are also things that I just don't quite like as much about prediction markets, for the sort of thing that we wanted to do.

I think prediction markets are really great in certain senses, and especially for these sort of big ticket items like elections, where everybody knows what's going on in the election at some level. There are lots of people who are even willing to put money down to bet on it — and if they're not, they're willing to be interested in and comment on it. Things that have a significant impact for a large number of people, and... a financial impact for a large number of people.

So, I think markets are particularly well suited for those.

Whereas, when you get to some sort of nitty gritty little question like, is SpaceX going to land on Mars by such and such a date... It's just going to be very hard to find lots of people who are willing to put money down to make predictions about that. And that have the expertise, have been following the news enough.

That one's even more popular than some. But in general it's going to be hard to get, you know, the liquidity and I got a small market like that.

Julia: I see. So you wanted a kind of Schelling point for all the people who are interested in these kind of narrow, specialized scientific questions. As opposed to having them kind of get lost in the sea of more popular questions on PredictIt.

Can you talk a little bit about how Metaculus is not a prediction market like PredictIt?

Anthony: Yeah. So, a prediction market is essentially something where you effectively buy and sell shares, like in a stock market or some other market. And if you think, for example, that there's a 20% chance of something happening, then if the thing on sale right now is for sale, like say less than twenty cents, if you're going to get a dollar back, then you'll want to buy it. If you think that the probability is higher than, you know, the selling price and so on, or the buying price and so on.

So, it's a mechanism where you're buying and selling shares, and effectively you're just betting. You know, a prediction market and betting market, in some sense is exactly the same thing. And what's really nice about this is that it's a simple mechanism — if you believe that the market is very efficient, then it should converge on the right price that gives the right probability.

And another nice thing about it is that you can use it for hedging — if, say some particularly loathsome piece of legislation is gonna pass, or person to be elected to office, or something, and you'd really like to hedge your pain by getting some money. Then you can make a bet like that. Or if you're a company that has some actual financial interest, then you can hedge, you know, the danger of some event happening in the real world, by going into a prediction or betting market-

Julia: So doesn't ... sorry, brief tangent: Doesn't the hedging issue kind of distort the prediction market's ability to converge on the most accurate estimate? Like, you know, I kind of hedged, before the most recent presidential election. I was like, "Well, at least if Trump wins, I'll make some money off of it."

But, if a lot of people are thinking that way, then we're not getting people betting their true belief. We're getting them betting some mixture of their true belief, plus like, you know, emotional damage control or something.

Anthony: I think that's right. So, there are various systematic effects like that, which I think any prediction mechanism, like a market, can have.

Another other one is for example, if you have lots of money riding on something, you might actually try to affect it. That's less of a problem with something like a presidential race where lots of people are trying to affect it — and good luck, you know, using your $1,000 to try to affect it.

But if it's a smaller issue, or a really small issue that you're trying to get a prediction about... and you had a lot of money riding on it, then it might well be worthwhile to try to affect it in some way. To turn the bet in your favor.

So you know, games have betting — like sports games and things like that are very carefully controlled when people are betting on them, so that you can't cheat. And if you just have a straightforward betting market or prediction market, on relatively small effectable events, then that's another big issue that you have to worry about. So yes, I agree with you.

Julia: Or potentially, assassination markets — like betting on whether someone will be assassinated.

Anthony: Things like that can be particularly tricky. I mean not only for the ... there's an actual danger, I suppose, that someone might go and assassinate someone if they had enough money riding on it.

But there's just a sort of general distaste of, like, this person dies in some gruesome way and you get money for it. And I think this was an actual problem, with some previous prediction markets — maybe it was unfair, but unfair or not, there's a danger to it in that sense. That if you get paid for some particular outcome and that outcome is loathsome to lots of people involved, that can be unpleasant.

Julia: Right. Well, so we've talked a little bit about what makes a prediction market a prediction market, but how does Metaculus run if people aren't getting money for their bets?

Anthony: Right, so the idea of Metaculus is, rather than asking you, "Do you want to buy or sell a fund at such and such a price," it's just directly ask people, "What is your prediction as to the probability that some events will occur," or "When do you think something is gonna happen," or "What is your prediction about some number?"

In the case of the timelines or the numbers, you're actually giving it a probability distribution. So you'll say, I think the most likely number is, say 1,000 Teslas sold this month, or whatever. And I'll say that the probability of there being less than 500 is 25%, or more than 2000, is 25%. So you'll give a few numbers that will define a sort of probability distribution. It's effectively sort of training people to think about making probabilistic predictions, and then just getting those probabilistic predictions from them.

Then when all those predictions are gathered, then some algorithm can — and it might be a very simple one, like just averaging them all or something and taking

the median — but some algorithm can combine all those predictions into sort of one single, headline prediction for that thing.

So that's the basic mechanism, it's very simple in some sense. It's just polling people for their predictions. But there can be a lot done with that data once it's all gathered, in the way that it's aggregated. And in the way that it's weighted, for different predictors putting predictions into it.

Julia: Now the nice thing about a prediction market is that people are incentivized to try to be accurate, because they get rewarded if they are. If there's no financial reward for making accurate predictions, what makes us think that the result of the algorithms on Metaculus are trustworthy?

Anthony: Right. So, there's a few ways to look at that. One is at the incentive side and one is just that the results side.

On the results side, there's now an accrued track record of a couple of hundred predictions that have been resolved, and you can just look at the numbers. So, that shows that it does work quite well.

Julia: Oh, how do you measure how well it works?

Anthony: There's a few ways — going from the bad but easy to explain, to the better but harder to explain...

Julia: That's a good progression.

Anthony: And there's the worst way, which I won't even use — which is just to give you some examples of great predictions that it made. This I hate, so I won't even do it.

Julia: Good for you for shunning that.

Anthony: So looking over sort of the last half year or so, since December 1st, for example... If you ask for how many predictions was Metaculus on the right side of 50% — above 50% if it happened or below 50% if it didn't happen — that happens 77 out of 81 times the question resolved, so that's quite good.

And some of the aficionados will know about Brier scores. That's sort of the fairly easy to understand way to do it, which is that you assign a zero if something doesn't happen, and a one if something does happen. Then you take the difference between the predicted probability and that number. So if you predict at 20% and it didn't happen, you'd take that as a .2, or if it's 80% and it does happen and that's also a .2, because it's a difference between the 80% and a one, and then you square that number.

So Brier scores can run from basically zero to one, where low numbers are good. And if you calculate that for that same set of 80 questions, it's .072, which is a pretty good score.

Julia: Yeah, I mean I don't really have a frame of reference — like, I don't know what, an average college educated person guessing about these questions would be expected to guess, or to get for their Brier score. But it certainly seems low.

Anthony: So that's something that you can sort of … So, some references are a .25 is what you would get if you guessed 50% for everything. So that's sort of totally uncertain. A .33 is what you would get if you just randomly assigned a probability between zero and one for everything. It's slightly different things.

Julia: Really interesting, yeah.

Anthony: And a typical … so if you took sort of a middle of the road Metaculus user for example, they would be in the sort of .18 to .22 to the very best people are around .1, and sort of middle of the road people are in the high .1s to low .2s.

Julia: Oh, I see. So this is really proof that aggregation is better than the sum of the parts.

Anthony: Aggregation totally works, yeah. So you can see that. So it's only very, very seldom that somebody, an individual does better than the aggregate score. That can happen because they'll just get lucky, but it's very seldom.

             And this .072 applies to the predictions that Metaculus made that are sort of more carefully aggregated, so it's not just a simple average or median or something. But there's kind of a careful process that goes on with recalibration and… things that we've talked about. But it works as one entity.

Julia: Yeah. So then I guess the other half of the question is, why does it work? Why is it that accurate even though people don't have financial reward?

Anthony: The reward that people have right now is just fun. And what people on the site call "Meaningless Internet Points." They compete surprisingly vigorously for that. So, it's fun and interesting. We've polled the users, "Why is it that you use the site?" The main things that they come to the site for are:

             One, to get better at predicting. So they're incentivized just to be accurate, because they want to be accurate to see where they go wrong and get better.

             Two, for the competitive aspect, to just play it as a game — that this is fun and I want to see if I can outcompete other people.

             And to just think about sort of thought provoking issues about the future, in a context where lots of people are thinking similarly, and that the level of discourse is high.

             Now, there certainly is a model for adding an economic incentive. That model isn't happening now on the site. It may sometime in the future. But the model that we have in mind, rather than a market where you're sort of betting against each other, is that… if you think about what a prediction market is providing, it's providing predictions, right? So if those predictions are of value to someone, they should be willing to pay for those predictions.

             So, in a regular prediction market, you have both the people who are making the predictions, and the people who are consuming their predictions, all kind of in one big pot. Or you can just sit outside and consume predictions from a prediction market for free, without providing anything to the market. People

subsidizing the market are really the people who are making the bets, and kind of generating the predictions.

And it's always seemed to me that, if the predictions are of value to someone, they should at some level be subsidizing those predictions. And the people who are making the predictions, and working hard to generate them, should be subsidized and compensated for doing that.

So the mechanism that might come into play, if Metaculus continues to grow and wants to do this, is a method where people who are interested in having a prediction about some particular thing would say, "Here's my question that I'm interested in. I will stake some money on that question, come and make predictions about it."

And Metaculus would then essentially offer that money as a bounty to people who come and make predictions about it, based on how good predictors they are, what is their reputation as predictors. And so they would then get that money, whether they were right or not, on that particular question.

Julia: Oh, interesting.

Anthony: But in order to get future bounties, they would have to keep a good track record. So everyone is incentivized to have a good reputation, so that they get the bounties offered to them, but nobody is actually paid individually on the basis of the outcome of any particular prediction.

Julia: Why do it that way, instead of dividing up the bounty in some fashion according to the relative accuracy of the people who predicted?

Anthony: You could, but the danger there is that ... One of the main reasons we actually don't have prediction markets, in the U.S. at least, that are operating, is because there are regulations against them. So prediction markets, unfortunately, fall afoul of both online gambling and unregulated trading of securities. So they're doubly not allowed in the U.S.

Julia: What about PredictIt? Is that not based in the U.S.?

Anthony: PredictIt has some special exemption, because it's a semi-academic effort or something, and the money amounts are small. So they got some particular carve out, and the Iowa prediction markets have done the same thing.

Julia: Well, you guys seem pretty academic to me.

Anthony: Yeah. So we could probably get some sort of an exemption, similarly.

But I think unfortunately, one limitation that prediction markets have because they have that exemption, is that the total amount of money you can bet is quite small. So, they can never actually be used by, let's say, a company to hedge against some outcome. Because no company's going to worry about a couple hundred dollars.

But I also think that it avoids the other issues, the sort of weird incentives of the hedging that we talked about. The trying to affect the outcome.

And it helps overcome just a barrier to entry. So if you, in order to make a prediction about something, have to put in your credit card number, stake some money, worry about — you know, people are very loss averse, worry about losing money. It's a lot of obstacles to having people actually do it.

Julia: There's also having your money tied up for a while, if it's a longer term bet.

Anthony: That's true. If it's an amount that's significant to you, you have it tied up.

So, we've aimed for keeping all those barriers as low as possible, and the way it's set up, there's no risk to anyone. In the sense that, if you're making predictions and you just get things wrong, then yes, you'll lose reputation points and you won't get as many offers. But you're never going to actually lose money. So that barrier is much lower.

That's sort of the thinking that came behind it.

Anthony: And I think it's also, another sort of side effect of prediction markets is that there's really no incentive ever, you know, unless you're co-investing with someone, to share information about the prediction that you're making.

Julia: You mean to share your reasoning?

Anthony: To share your reasoning, yeah. You might try to persuade someone about the wrong thing. But ultimately you just end up competing against everybody. So you might try to convince people, to sort of manipulate them one direction or the other, to help you out.

But there's no sort of honest level of sharing of information — which, coming from a scientific background, that feels very frustrating to me. You know, science is like a very collective endeavor, where part of the reward system is convincing other people correctly, and coming to the truth in a consensus way. And working together and cooperating as well as competing, you know, certainly scientists compete with each other.

So I like having a mechanism that isn't strictly competitive. And we're sort of working on ways that we can bake in some more, like direct cooperation or teaming elements, so that people are even more incentivized to kind of share their information and reasoning, and try to persuade each other in an honest way.

Julia: What is the incentive on Metaculus for sharing? It seems like, in terms of meaningless internet points, currently, you're still ... There's a leaderboard, and you still kinda wanna be high up. To the extent you care. And then if you add in the business model you were describing a few minutes ago, you kind of want to be one of the people who gets to participate in the bounties.

Anthony: Right.

Julia: So if you share your reasoning, you're just helping other people.

            ... Although I guess you are, in theory, making the site more effective overall, which increases the number of bounties that are likely to be offered. That's something.

Anthony: Right. So that's pretty interactive.

Julia: Yeah.

Anthony: So I think right at the moment, the main incentive is just for fun and to have a nice discussion. I think there just isn't as strong a disincentive as there might be in a market. But I think with some additions ... I think the way that it is likely to be incentivizable, if that's a word, is to have sort of teams — either across questions, or even on a particular question, where you kind of form an impromptu team.

            Or just try to have a predictor who gains followers of some sort. So they might want to persuade people that their prediction is really good, so they will adhere followers and then the followers will get something, and the predictor will get something out of that relationship.

            So that's the kind of mechanism we're working on right now, to try to encourage that.

Julia: Nice. Backing up a step or so, what's an example of why someone — a person or a company, or even a government maybe — would want to have predictions on science and technology?

Anthony: Yeah. Interestingly, the actual motivation for Metaculus came just when we were founding the Future of Life institute. And I was thinking, "Wow, I'd really like to think about what we could do now to sort of affect the future in some positive way, in the big picture."

            And realized that, how are we gonna make decisions whatsoever, if we can't make good predictions about it? I think at some level, everybody should be asking that same question when they're making decisions. Because every good decision really is sort of implicitly a prediction. And a sort of value function on how different choices are gonna affect what you value.

            I had never thought of it in that much detail until we started the Future of Life Institute, and was really thinking about, "Okay, where are those points of greatest leverage?" Well, we've gotta have to have good predictions. And that kind of led, through various different steps, to Metaculus.

            So I think at some level, we're all running around making continuous predictions all the time, that's a lot of what our mind is doing.

Julia: Right.

Anthony: But in making that more explicit, you make better predictions. And I've gotten to the point where I'm now very frustrated when I feel like I don't have enough

information to make a quantitative prediction about something. If we're fixing something in the house, what is the probability that this thing is gonna fail in the next year if I don't fix it right now? Is it 70%? Is it 80%? I really wanna know. So I've gotten to that mindset, just through obsessively being on Metaculus.

But it's a mindset that businesses and other entities, governments and so on, really ought to have. Because when you think about just ... When you're deploying billions of dollars on something, whether it's a stock evaluation that's of billions of dollars, or you're allocating resources that are billions of dollars, even a relatively small error in what the correct assessment of the probability of something happening, is hundreds of millions or billions of dollars right there, that you're just throwing away, in expected value.

So I think any organization that's making quantitative decisions about how to deploy their resources, which is to say any organization, should be thinking along these lines. Even if they're not using a mechanism like this, they should be thinking, "What are the possibilities, and what kind of probability should we be attaching to those?"

Another example is in grantmaking. So I'm involved with grantmaking through the Future of Life Institute and the Foundational Questions Institute. And I've taken to asking the people who are reviewing the potential grants to think about: what are the three major possible outcomes of this research if it was funded? For each one, what's the probability that you think you would assess that they will be successful in achieving that outcome? And then, on a logarithmic scale, how important do you think that outcome would be, from world-shattering to completely trivial?

And you can then work out the numbers, if you want. Adding up the exponentiated impact times the probability, and adding them all up and stuff. We don't actually make grants using that, but it's an interesting number to see.

And most important, I think, is that just asking the people to do that makes them think about things a little bit differently. If you're choosing to fund this thing, but if you look at the numbers, and it's really low impact, by the numbers that you assigned to it — versus this other thing which is really high-impact, even though it's low probability or something — what's going on there? Why is it really that you're so attracted to this project? Is it because you just happen to know this person really well and like them, or it's cool and sexy or something?

So it's another interesting piece of data to kind of look at, what are your actual ... If your actual criteria are, "We want to have a big impact," what other things are creeping in there? Maybe those things are okay. But let's be conscious of what they are.

Julia: So one of the biggest challenges that I run into with making predictions is just specifying them clearly enough that it's clear whether they came true or not. I'm sure you must have this problem on Metaculus. How do you deal with it?

Anthony: Yeah, it's a problem that you really don't appreciate until you start actually doing it. How difficult it is to really nail down some question so that it's unambiguous.

Some are ready-made unambiguous for the most part, like who's gonna win some sporting event? But even … We put up a question, finally, about whether Trump would get elected in 2020 — and there's 20 comments sorting out the details of, are we talking about "gets enough votes," or are we talking about actually sworn in? Or what happens between the state votes and the meeting of the electoral college? So there're all kinds of pretty low probability, but when you have enough of them, they add up.

Julia: Yeah.

Anthony: Lots of questions go wrong if you don't get those things nailed down.

The way we deal with it, effectively, is: we come up with a question, then somebody else reads through it with a particular eye toward, "Let's make a very clear, unambiguous resolution condition." And then, generally, we post a question for a few days before it actually goes live, and people give feedback and say, "What about this? What about that?" Does the rocket just have to clear the launchpad, or make it to space? Or what if it blows up on the …

And it's interesting, because there're a couple of things I think you learn from that. One is, how vague most of the predictions we make are. And that's even when we're trying to be fairly precise. … If you look at most predictions that people make in the media, like "20 predictions about what will happen in 2018," almost none of those can you actually operationalize into something that you can check later, whether it happened or not. I've tried.

So I think there's two ways. One way to deal with that is what I've said. To just work really hard to sort of morph the question into a more operational one. Often you can do that, I find, without doing too much violence to what you really want to get at.

Sometimes you can't. So we just put up one, for example, about whether Ray Kurzweil will turn out to be basically right.

Julia: Oh man. That's —

Anthony: It's a super vague question. It was sort of deliberately made as super vague. But it's the kind of thing that you kind of want to know. Like, "Yeah, is there exponential progress and a singularity in 20 years, or is that kind of just basically wrong?" That's sort of what you wanna know, but it's really hard to figure out how to make that into a precise question.

Another option that we've been exploring are things like … We're not gonna operationalize this, but we are gonna have some body — or somebody, one of the two — that is charged with making a decision on this. So part of your prediction is gonna have to be to model that person, or that group of people who [judge] the decision.

And then, you can let the questions be a little bit more vague. So you can have a group that [judges], "Is cold fusion a real thing?" I'm actually on the hook as one of those people, that'll answer that question.

So you'd name the people. You understand who they are and what their reputations are, and so on, and you have to fold that into your prediction. So I think there're different ways of dealing with that.

Julia: That's the typical way that I've tried to solve it, just when making bets with friends. We'll pick a third party who we both sort of trust to be basically reasonable.

Anthony: And yeah, I think it's a good method. But you do lose something, too, in that the making precise of the question, I think is a very valuable exercise to go through, as well.

Julia: Something I saw mentioned on one of the question on Metaculous. It was consensus forming. What is that?

Anthony: So that's something we've been thinking about a little bit recently. Actually, the Kurzweil question is of this nature, where... if you think about what a prediction is, it's sort of making an assessment of the probability that something is the case, before, temporally, that happens.

So say there's something in the world right now, like "What's the gross domestic product of Russia?" I don't know. I could estimate it, I could give it a probability distribution. If I spent a minute on Google, I could narrow my probability distribution a lot, but there might still be some distribution of different sources and different methods and so on.

Anything we're talking about in the real world... becomes certain at some level, and there is a probability distribution that we want to assign to it. And events in the future are just a particular class of facts that happen to not be very accessible right now. But as time goes on, we can generally narrow the probability distribution about them.

So, the idea in some sense is: If you think of what people are so-called "predicting" is really just sort of assessing the probability of a given fact being true, and that fact happens to be, say, 20 years in the future. As time goes on, hopefully, more information will become available and that probability distribution will narrow more and more to a particular outcome. And then, as the time gets really close, or is even present, then that probability distribution will hopefully narrow to, like, 99.99% yes and .01% no, or this value plus or minus some tiny value.

So, there's a sense in which I think that can be considered a continuum — where, as time goes by, you just gain more and more pertinent information that allows you to collapse your probabilities. And a huge piece of information will sometimes come when, like, the sporting event happens and the transmission goes over the television, and you see that one team won.

You're still not sure — like, in a philosophical sense. It could be that somebody's fooling you and you're hallucinating, or all ... You know, so there's some tiny little width of probability that that's not true, what you think is true. But there's a huge piece of information that collapses it a lot.

And so, the idea is to just think of that continuum, and to let questions just stay open, and watch that probability distribution narrow, and narrow, and narrow, and at some point say, yes, it's narrowed enough that we say this thing happened. And until that happens, it's still an open question.

So, that's a sort of point of view that we're experimenting with, I think, whether it can ... Metaculus can not just be thinking about future events only, but even some current events, in terms of how they will resolve in the future, things that are currently contentious you can still make predictions about.

Julia: Yeah, I feel like ... And this is just my rough impression, but I feel like 10 years ago it was much more of a fringe view that artificial general intelligence was likely to happen in the next couple centuries, and that it was also plausible that there could be serious, significant consequences of the invention of AGI.

And I feel like it's much less fringe now. It's more, like, the normal view that, yes this is — maybe not likely in the next 10, 20 years, but likely in the next few centuries, and that yes, serious consequences are plausible and worth serious consideration. That's maybe the standard view. Without getting too specific about exact probabilities of that outcome, or of AGI in a specific timeframe.

And we couldn't really have had a prediction on AGI being invented, or, you know, what the consequences would be that would resolve in time to be useful to people. But if we had set up a market on, How would consensus evolve?, then that actually could have paid off by now, just 10 years later. So that's interesting.

... Although, I wonder — doesn't that create kind of weird incentives for people to ... Like, if the thing you're betting on is what will the bets end up being on the site, then-

Anthony: Yeah.

Julia: Yeah.

Anthony: I think it's at least possible that you could have that sort of meta quality run away in the wrong direction, right? So, you could imagine people saying, what's the probability that, sort of, this question is going to resolve true? And we've has a weird series of meta questions on Metaculus like that to sort of play around with it, because you can imagine that weird things will happen because it's a self-referential system.

So, I think we still have to resolve what is the right sort of way to think about that. How do you actually ... It's kind of an unexpectedly profound question when you start to think about it. How do we actually agree on anything, or find consensus about anything, in some kind of mathematically rigorous way? And so, I think we're still experimenting a bit with the right ways to think about that.

Julia: Okay, related question: The accuracy and calibration of the Metaculus community that you referenced earlier... kind of implicit in your description was, that's like accuracy and calibration on short term forecasting. Because, of course, you haven't been around for decades. And then you're also, separately on

Metaculus, asking people to predict very long term things — like, you know, will SpaceX land on Mars by 2030?

And the sort of premise there is that the skill of the community at making short term forecasts carries over to its ability to make accurate long term forecasts. Or else the former thing wouldn't be relevant in judging the long term forecasts. I feel pretty torn about how much I would expect short term forecasting to be the same skill as long term forecasting, or similar enough to be useful. What's your view?

Anthony: Yeah, I mean, certainly we don't really know, because we don't have both data sets to compare. I think my sense is that where things will go is that when you think of some long term question, it really should be considered ... and what we'll hopefully do in the future is consider it to be a kind of overarching question that we would like to break down into other questions that we can assess earlier.

And that in some sense, if you really want a long term prediction, what you should have is some sort of model for that, that is based on some subsidiary questions that you then combine their probabilities to get a probability for the long term one.

Like, for example, the will there be human-machine intelligence parity by 2040? You know, that's a big question that everybody should want an answer to, and just making a prediction straight about that question — you can do it, but I'm not sure how good of a reason we have to believe the numbers that are coming into that.

What would and will be much more interesting is to have a model for that, that says: here is some structure of what it would take to have a human-machine intelligence parity. Like, a tree of events that would have to happen in order for that to happen or not happen, and break them down into sub events that would have to happen or not happen. So, there are different structures, like a fault tree, or Bayes network and things that can be constructed like this.

And then try to make predictions on the little pieces, which hopefully will not be as long term as the big piece. And then see how, in your model, predictions about the long term thing change, as you change those predictions about the short term thing. And, as time goes on, you can sort of see that long term prediction changing as you gather more and more information.

I think it's likely to take something like that. Because we're not going to have the data for decades, to know how the short term ones translate into the long term ones. And if we want something that we can trust at some level, or get insights out of, I think it's going to have to be more of a networked set of questions in a structure like that. Because I agree that, at the moment, if you go to someone and say look, Metaculus has been super right so far in one-year questions and here's what they say about 2070 — you know, nobody's going to believe that, and they shouldn't.

So, I just agree. I think my point of view is that like other ... you know, like many endeavors, including scientific ones, forecasting is going to have ... there's not going to be a one-size-fits-all answer to it.

And there's going to be a whole bunch more machinery that should be added in on top of other ways we have of forecasting, like strict scientific models, where you have some particular set of equations you can evolve, or computer simulation that you do. Metaculus is another thing in the spectrum of doing sort of aggregated probability estimates from people. But there are lots of other structures, in between and amongst those, that we'll have to work out, and sort of add different ingredients into, to get the best possible forecasting system.

Julia: I just remembered a question that I'd wanted to ask you earlier in the episode. I was comparing Metaculus to prediction markets, like PredictIt, or Intrade — but a different comparison you could make is just to, like, investors in the real world, making decisions about where to put their money based on their implicit predictions. So, this is sort of more relevant to science and technology forecasting than it is to maybe, predicting the results of elections or ... I don't know, other things that you can't sort of naturally bet on in the real world.

But I guess my question is, why someone who wants to get a sense of, you know, what good forecasters think is likely to happen — why should they look at Metaculus? Instead of, say, looking at how much companies are spending on AGI, and the stock price of SpaceX, or other things that are representative of how people in the real world are actually betting, you know, based on their implicit predictions of science and tech?

Anthony: Yeah, and you should, certainly — but I think the tricky thing is that, you know, if you take, say, the stock price of some ... of SpaceX, or Tesla, I guess... [SpaceX isn't public] that's folding in lots, and lots, and lots of different things into its price. And you know, in the most rational and perfect of all possible worlds, it would be some kind of weighted sum of probabilities of different outcomes. And the price of Tesla if this outcome, versus if that outcome, would be like a superposition of those two as the value, and so on.

And you can model it that way. But if you want to know, like, what is the probability of the Model 3 being produced in sufficient numbers in the next six months, trying to disentangle that from everything else that goes into Tesla stock is very, very difficult. In principle it's in there, if everybody's doing their job right — but you're going to have a hard time going that direction. It's much, much easier to say, "Let's make an assessment of that probability of production numbers," and then think about what influence that's going to have on Tesla stock, than to go the other way, because there's lots more pieces of information.

And I think that's actually one of the financial utilities. Undoubtedly hedge funds and so on do this, right? They're thinking, "There's a Supreme Court announcement coming out soon that's going to affect various drug companies, say, and let's think about what the effect on drug companies A, B, and C is if, you know, outcome one or two comes out of the Supreme Court, and what are the probabilities of those..." and, you know, invest accordingly. So, they're certainly implicitly, or explicitly, doing that, and so they might be interested in the numbers that Metaculus has.

But other people should be doing that also. And it's not clear that they are going to have access to the sort of dedicated team of highly quantitative people who can assess probabilities that a hedge fund will have, right? So, in some sense I

Julia: envision it as a resource for non hedge funds — because they've got their own team doing it — for non hedge funds who would like to do something similar, and really have some way of getting those probabilities.

Julia: Cool. Well, Anthony, before I let you go, I wanted to ask you for a book, or blog, or some other resource that you substantially disagree with — like, you disagree with the worldview of the writer, or, you know, the argument they're making, or something like that — but that you, nevertheless, think is valuable to engage with. Like, you respect their methods, or you think the questions they're posing are worth considering, or something like that. Do you have anything in that category?

Anthony: Well, I thought about this question a fair amount, and realized that, you know, it's a difficult question. Because when I think about agreeing or disagreeing with a book, you know, I almost never do. Like, a book is going to have lots of true things in it, and lots of things that I'm skeptical of, or a person, is also a complicated thing.

Julia: Couldn't you put a probability on it, you know, or a number from one to 100...?

Anthony: Maybe a fraction of things I agree with in a book. But I found that the sort of longest thing that I can imagine fully disagreeing was, like, an article or a blog post or something. But that told me something ... And I'm dodging the question a little bit, but that told me something that I think is important: which is that, you know, when we're exchanging individual sound bites or, like, one paragraph arguments, it's really easy to find ways to disagree with people and have lots of contention.

           Whereas, if you take the time to sort of engage in a little more full way with a book, or with a person, or with a, you know, multi paragraphs — the common ground is there, if you look for it at all. And I think looking for that common ground and then working your way back from there to see, okay, now that we have that common ground, where are the places that we disagree, and can we identify why?

           It's just a so much more pleasant way of interacting with different ideas and different people than sort of saying, "Okay, my premise is that they think this, and I think that, and so we're in disagreement and now how am I going to defend my point of view?"

           So, I think that there are ... I can give examples of a couple of articles that I disagreed with, but I think what this told me in the bigger point was: read more books, and talk to more people at length. Because it's much, much harder to sort of overall disagree and fail to get something out of reading something or interacting with something that big.

Julia: Well, that's all a lovely sentiment, but it hasn't made me forget that you didn't answer the question!

Anthony: Okay, okay, so, I would say the very most recent thing — because I decided the only way that I could go is to pick something very, very recent... I was reading a paper about the argument that going out and colonizing the galaxy, which is sort of seen as, like, this great way to get lots of positive utility, was a net negative

thing. That if we stay here on earth, you know, we might go about our happy lives and might kill ourselves off or not. But if we colonize the galaxy we'll kind of divide up into all kinds of diverse species, and societies, and kind of factions, that will be constantly fighting with each other, using terrible weapons of mass destruction and so on, and this will be just a terrible place.

And so, I totally disagree with this. Like, my vision of going out in the galaxy and having lots of different species, and factions, and civilizations, and so on is, like —

Julia: It's pretty great! Yeah.

Anthony: — the prize. And here it was sort of being put forth as sort of the negative outcome.

So, I totally disagreed with that view of it, but thinking through, okay, what is most likely to happen? What does happen if we can't, you know, communicate faster than light [between] all these isolated species? What happens if, you know, the first thing you learn of a new civilization is the world ending weapon that appears from them? Because that might be the sort of game theoretic right thing to do in certain types of dynamics, and galaxies, and so on.

So, I'm sort of congenitally a little bit optimistic. So to read something that was fairly pessimistic and just ... you know, but well put together and thought through, I thought was very interesting and useful for me.

Julia: Did it —

Anthony: That was a paper by Torres, actually, Phil Torres.

Julia: Oh, great. Yeah, we'll link to that on the podcast website. So, did it basically force you to be more specific or rigorous with yourself? About either the conditions that would have to be met for this to be a good future... Or, like, the potential, you know, "cons" in the pro-con calculation, even if the overall net effect is still pro?

Anthony: Yeah, I think it made me realize that I had a sort of hidden running assumption that the outcome was generally going to be positive, and that that wasn't a given. And if the outcome could be positive or negative, then let's think about why would it become positive or negative. And are there dynamics that would lead to one or the other, and can we do anything about those?

So, it ... You know, uncovering that helped see, oh, maybe there's an opportunity for some input here. Maybe not by me, but by our descendants some day. Rather than just, "let's go along and see, hopefully everything will work out fine, because I think that's what will happen."

Julia: Yeah, nice. That's a good answer. Well, thank you so much for coming on the show, Anthony. It's been an enlightening hour.

Anthony: Well, thanks for having me. It's been a pleasure.

Julia: This concludes another episode of Rationally Speaking. Join us next time for more explorations on the borderlands between reason and nonsense.