Rationally Speaking #226: Rob Wiblin on “An updated view of the best ways to help humanity”

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

Rob is the director of research at 80,000 Hours, which is a non-profit that focuses on helping people figure out how to maximize the positive impact that they can have with their careers. Before that, Rob was the executive director of the Center for Effective Altruism. His background before that is in economics.

He's also the host of the excellent 80,000 Hours podcast, which if you aren't already a fan of, I think you should really check out, if you like Rationally Speaking.

So there are a lot of things Rob and I have to talk about, but the thing that really piqued my interest to focus on today is -- a few years ago, I think back in 2014, I had Ben Todd who's the president of 80,000 Hours and one of the founders, on the show. And we talked about some of the logic behind 80,000 Hours and effective altruism, and some of the basics around how to pick a career to maximize your positive impact.

It's been over four years since then. And some of the thinking of 80,000 Hours and the surrounding effective altruist movement has evolved. Some views have shifted, other views have maybe clarified. There have been some misunderstandings or misconceptions about what 80,000 Hours and EA in general actually believes.

So I thought it would be great to sit down with Rob and review the evolution of 80,000 Hours, and his thinking on how to affect the world positively. So Rob, that's what we're going to do today. Great to have you on the show.

Rob: Yeah, thanks so much for inviting me on. I've listened to the show for many years.

Julia: Oh, excellent.

Rob: I guess it's good to be speaking rather than listening for once.

Julia: I want to say, “Rob Wiblin and Julia Galef together -- What is this, a crossover episode?” but only fans of Bojack Horseman will get that.

Rob: The greatest crossover of all time.

Julia: I'm sure Rob has more productive things to do with his time than watch TV, so that's for the Bojack fans out there.

Rob, can you just give me the basics about 80,000 Hours -- What do you guys do? How big are you? How long have you been around?

Rob: So 80,000 Hours is all about helping people have a larger social impact with their career. We do research to try to figure out how people can do more good with their work. We publish that on our website, and we have our podcast. Also, we provide one-on-one advising to people who bring us their particular situation and we give them ideas for how they can potentially help people in a bigger way with their work.

Julia: Roughly how many people have you advised at this point?

Rob: How many people have we advised?

Julia: I mean, order of magnitude.

Rob: Oh, well, we've had about four or five million people on the site over the last seven years...

Julia: Nice.

Rob: I guess we have about 1-1/2 million visitors a year now.

I think in terms of coaching, I think it would be around a thousand that we've coached so far. We didn't do a lot of coaching for a couple of years, we were mostly just doing research. But now we're growing the in-person team who's providing advice to a whole lot more people. Hopefully, that number will go up quite a lot over time.

Julia: Nice. For listeners out there who haven't already heard an explanation of your name, what does the 80,000 Hours name refer to?

Rob: Right, so 80,000 Hours is roughly the number of hours that someone would work in a full-time career. I think it's 40 hours a week for 50 weeks a year for 40 years.

We chose that because it gives you an indication of just how important your career decisions can be, that there's a lot of time you're potentially going to spend. So you should think about how you can spend it well.

On the other hand, you could just think, well, 80,000 hours is actually not that much time relative to the scale of the problems in the world. So you're not going to be able to solve all of them with that amount of time, so you're going to have to prioritize pretty hard. Those are the two angles on it.

Julia: I think the first time I heard the explanation of your name, I made what I thought was a very clever joke... about how if you devote your career to researching life extension, then you can increase your number of hours, and you guys would have to change your name to 50 million hours, or something like that.

You were very nice about it, but it was pretty clear you'd heard that same joke a million times.

Rob: Yeah, there's a lot of jokes of that kind that I've heard many times. It's unfortunate that my laughter wasn't able to be sufficiently sincere. I should practice that.

Julia: You were really trying though. I appreciated that.

I alluded to an evolution in the thinking, or at least the structure of the public version, of 80,000 Hours arguments. What would you say are two or three of the main things that are different about the advice 80,000 Hours gives now than it did five years ago?

Rob: I think a lot of people, when they hear about 80,000 Hours, or effective altruism, or when I bring it up with them... they think that that basically means going out and making a lot of money, "earning to give," and then donating it to charities that have been proven to have a really large impact.

Julia: That is the media portrayal of 80,000 Hours. There used to be articles about you guys focused on that.

Rob: That's right, and I think that's really unfortunate. It's quite frustrating because that's actually not what I think most people in the effective altruism community are doing, or at least not what they're aiming to do in the long-term. And it's not what our advice is.

Basically, that perception came about because that's one of the relatively easier options that people could take to explain. Early on in the life of 80,000 Hours and effective altruism, if you wanted to offer something that would be interesting to people that they hadn't really thought of, that would indicate to them how they might be able to do more good...

Julia: Like in a counterintuitive way.

Rob: It's a little bit counterintuitive, yeah. Most people hadn't heard of this, "earning to give," back in 2011 when we were launching. It's an interesting piece of media bait that a lot of journalists picked up, and still pick up on to this day.

There's, also -- GiveWell is this charity evaluator. I think a sponsor of the show, in fact --

Julia: Yeah, a sponsor.

Rob: They've done research to try to find charities that have very strong evidence behind them, and where they have some idea of actual impact that you would get, the bang you get per dollar. That was relatively well-advanced at that time. That's an interesting approach that you could try to take to have more impact, would be to do things that have demonstrated to have a really large impact. But all of these things, I would say, are just minority views within this broader community, or this broader intellectual movement.

Julia: When you say they're minority views, do you mean that there are some people who think that most people should do earning to give, but the people who think that are just not very common?

Rob: Basically, if you break it down there's a substantial number of people who are involved in effective altruism who think that we should focus on global development and health. I think it's something around 40 or 50%. Although, perhaps if you went to people who are working full-time it goes down to more like 10 or 20%.

Then in earning to give, there's quite a lot of people who are doing earning to give. I would say it's maybe, again, about half of the community is trying to do good that way. Although a substantial fraction of those people, I think, are doing that with the intention of eventually going and doing something else, once they've developed their skills and found a really good fit for them.

Both of those things are substantial parts of effective altruism as a whole. But they're by no means dominant, and by no means the thing that it comprises.

In terms of doing evidence-backed interventions, using that as a strategy to have more impact -- again, maybe a third of the community thinks that that's one of the key ways that they're going to try to do more good, is to use really strong social science evidence to find things that really work.

I'd say maybe another third are, “yeah, that's an interesting approach that I would use among different methods.”

I think, probably the third that I'm in is being actually somewhat skeptical, because I think that interventions are [in] problems where there's very strong evidence for what you can do. Actually, it might be negatively correlated with things where you can have a very large impact, because those are likely to be fields that are very well-developed.

Julia: What's an example of a-

Rob: Well, I'd say global health is probably one, where there is a lot more evidence in that area than there are in most of the problems that we're more focused on now. That's because so much effort, so much intellectual firepower has gone into those areas.

Of course, it's good, all else equal, that you have more evidence about what works and what doesn't. But it's a sign that the area is not so neglected. It's not a problem that you can go out and pioneer. There's already millions of people working on it.

Julia: Yeah. I think about that sometimes in terms of the risk-reward trade-off.

Rob: Right.

Julia: Like when you're deciding where to invest, for example. It's not quite the same thing, but there's a parallel structure there.

Rob: Right. So giving to bednets, you might be viewed as investing in Costco or Walmart, or something like that. A very reliable company that's going to return dividends pretty reliably.

Julia: “Bednets” is antimalarial bed nets, which -- there's a lot of evidence around the number of life years you can save by purchasing X number of bednets.

Rob: Yeah, but I think if you're actually trying to maximize your impact it probably makes more sense to be more of a venture capitalist. To go out and look for things that are riskier, maybe harder to find, have a high chance of failing -- but where you can go and do something that other people haven't done. And where you would have just a larger bang out of your career, in expectation, even though there's a high probability that it won't work out.

Julia: So it sounds like you're talking about two different changes or misunderstandings at once. Where one of them is, “Do you do earning to give -- where you're working in finance or whatever, and donating your money to a charity? Versus, do you do direct work, where you yourself are doing research, or working at a charity, or directly working on a problem?”

Then the other question is, “Do you work on or give to something where there’s very rigorous evidence, [because] it's already pretty well-developed, like global health? Versus, do you work on or give to a cause that's riskier, but potentially more impactful for humanity or the future?”

Rob: Yeah, so you can do any combination. I guess there's four combinations you get out of those, and you can combine them however you like.

We're now more focused on things other than earning to give, most of the time. More speculative, higher impact, research and innovation on policy. That kind of career approach. Rather than earning to give, for things that are, I guess, boring would be the negative way of putting it. Or reliable. Predictable.

Julia: “Sensible.”

Rob: Exactly, “sensible,” yeah.

Julia: Okay, so earning to give being less emphasized or less important than maybe the popular conception, that's one thing. Are there other things that are different about 80,000 Hours’ advice now than people think?

Rob: Yeah. The corollary of focusing less on earning to give is trying to find leverage elsewhere by doing scientific research, or doing policy work, where you hope to move a lot of money, or legislative power through government.

I think another common misunderstanding that people have about 80,000 Hours' advice was that they thought that we were recommending that people go into typical prestigious corporate jobs early on in their career.

Julia: To do earning to give, or for some other reason?

Rob: One reason would be to do earning to give. Another would be just that you wanted to build up a lot of career capital.

One part that we did suggest early on was going into consulting, say for the first few years out of university, with the hope that you would build up a network, lots of skills, lots of connections, potentially money in the bank. That you could use to take risks in your career.

Some people took that path and it worked out for some, and didn't work out for others. I think it is a reasonable path, but these days because we've become more confident about the priority part, that we really ultimately want to see people get into as their careers mature.

It doesn't seem like going into typical corporate jobs is really anywhere close to the best way to getting into those positions. Basically, it seems like the career capital that people were getting, the kind of skills that they were building up in consulting, or other corporate jobs, just don't transfer over so well into natural sciences, or into policy careers, or international relations, things like that. The career capital transferability isn't so good.

We've become more confident in recommending somewhat unusual paths, or working on problems that most people weren't focused on before, just because we've had more time to think about it, had more time to hear people's potential objections...

Julia: I'm sure we'll get more into this later in the conversation, but what's one example of an unusual career path or field that you would now recommend people go into?

Rob: The stuff that we talk the most about these days is trying to improve the long-term future of humanity, [principally] by preventing global catastrophic risks.

So trying to prevent, for example, a war between China and the United States in the 21st century. Trying to prevent nuclear weapons from ever being used. Trying to prevent new technologies from really taking civilization off track, by being either misused, or used in an accidental way that causes a really large global catastrophe.

That's not what most people think of when they think about charitable work, or trying to improve the world with their career.

Julia: Right.

Rob: We were aware of those ideas in 2011, but were initially cautious because at least to most people it seems [not] to be common sense that that would be the way to have the largest impact. As we've gone on we've thought about it a lot. We've sharpened those arguments, seeing exactly what does the argument rest on, what kind of objections do people put forward? And decided that, no, actually, we really think that this is likely a very compelling way to do good. That's ultimately what you want to end up doing then.

Corporate jobs is just not really the sensible first best path out of university. You would want to just go and try to do something that would get you into one of the relevant roles directly.

Julia: So do you think that the change was more about you guys becoming more confident that these weird catastrophic risk reduction career paths were the way to go? Or is it about becoming more confident that you could make that case publicly in a way that wouldn't put people off?

Rob: I think it's a bit of both. I was personally already quite confident early on, perhaps that's my temperament. Maybe my personal views haven't shifted so much.

But I think a lot of other people who are more temperamentally cautious, who heard these arguments and thought, "Oh, yeah, that sounds compelling on paper, but I'm just going to stick to what seems like more commonsense to me..." I think many of those people have shifted over a period of years where they just explored it more and became more convinced. I guess as a group it's become more possible to take action in that direction, because it's just more of a consensus among people who work in this area full-time.

Julia: As soon as I asked that question it occurred to me that, I think for a lot of people, to some extent there's not that much space between “What am I confident in?” and “What could I make the case for to other people in a way that would make sense to them?” Maybe those two questions were all bound up together for a lot of people.

Rob: Well, yeah, I would have been willing to push on it a bit harder, faster.

Julia: Yeah, I believe that.

Rob: I guess, yeah, you know what kind of a person I am, right? I tend to ... Yeah, I mean-

Julia: A straight shooter.

Rob: Straight shooter, I guess. Perhaps, also, a bit more risk-taking. A bit more of a venture capitalist when it comes to ideas to maybe jump onto. I mean, I think, it takes all kinds in this respect. You don't want everyone ... If everyone was like me then the world would just fly back and forth between ideas.

Julia: It would be an interesting world.

Rob: It would be too faddish. It would be an interesting world. Maybe a riskier world, which is not so great.

But I think you do need some people who are willing to try to stake out new ideas and say, no, actually I do believe this, and I'm going to push it forward, and then see if they can convince everyone else who's a bit more cautious.

Julia: I wanted to ask about one apparent change that I read about on the 80,000 Hours mistakes page -- which is really great. I'd recommend people check it out. I think it's just called "Our mistakes." It's one of the main pages on the 80,000 Hours website. We can link to it.

They list mistakes that they think they've made in logistics or management, but, also, mistakes in having gotten the wrong answer on some question, or given the wrong public position on something.

One thing they say is "We always thought personal fit, i.e. how likely you are to excel in a job, was important. But over the last few years we've come to appreciate that it's more important than we originally thought, most significantly due to conversations with Holden Karnofsky," -- who is a founder of GiveWell, and now runs the Open Philanthropy Project.

Why do you now think that personal fit -- you being 80,000 Hours -- why do you now think personal fit is more important than you had previously thought?

Rob: Yeah, I think the argument there is that most people who end up having a huge impact in their career are very monomaniacally focused on a problem. They're really passionate about the organization that they're building, or they're just focused on influencing some policy agenda in Washington, D.C. and they just work all the time at it.

If you look backwards it seems like most people who had a really large impact, they were not just doing this part-time on weekends. They were putting all of their energy behind it.

I guess, there could be a bunch of reasons for that, but one might just be that you get economies of scale by combining all of the relevant information into one person's head. They can become really specialized and really expert, and then people ask them what to do so they could become real leaders.

I think we always thought that personal fit was quite important, but we perhaps thought that it would be possible for someone who wasn't passionate about an area, wasn't passionate about a particular method, to just stick with it. Just grit it out and say, no, I'm going to do this even though I'm really not enjoying it.

I guess over time we've become more pessimistic about people's ability to stick with that.

For most people, it seems like they have most of their impact later in their career once they've built up a lot of skills. They've built up a lot of connections where

they can influence what's going on. Having someone who grits through it for a couple of years, but then gives up because they don't have enough energy to continue with it is probably losing most of the value from that person's career. They work in it for a few years.

They just stick with it even though they're not enjoying it, and then they leave, and then all of the skills that they've built up, or all of the connections that they've built up, all of the organizational capital they've created then dissipates. If you're playing a long-term game then it seems like personal fit becomes more important.

Now one thing that I would say is that we still think that you should try to find a priority area, and then try to find a key bottleneck to solving that problem. Then within that, look for a role that has a good personal fit for you. Typically, there are so many roles at that point that most people can find something that's potentially suitable to them.

I guess if you can't find anything like that than earning to give is always still potentially quite a good option even if it's not the one that we suggest that people look at first.

Julia: I'm curious, does 80,000 Hours take any kind of official position -- I mean, I'm interested in your personal thoughts as well -- on how to balance... I'm imagining someone who can actually stick it out in a career that's not the ideal career for their own happiness, or intellectual interest, or whatever, but where they can have a large impact. And imagining if someone could do that, should they? Is that the right choice for them to make morally?

Do you or 80,000 Hours have a position on whether someone should take a career like that, if it's not a personal fit in the happiness sense?

Rob: Yeah. I mean, I think, there's a range of views on the team. We don't really have a position on that per se.

I guess, personally, just to be honest -- I think, morally, that would be good. If someone really could just go and save thousands of lives, tens of thousands of lives through their career, even if they didn't enjoy it. As long as they actually could do it, and stick with it, even if it wasn't super fulfilling to them. Then I agree, in some moral, hypothetical principled sense, then they should do that. I'm not going to shirk that conclusion.

But, I guess, in practice I don't really push that agenda very hard. In most cases the roles that actual people in the real world are going to find where they're having a large impact they're going to find very stimulating. It's going to be very rare that someone's best role is something that they find unpleasant or unfulfilling.

Julia: I mentioned the “Our mistakes” page on the 80,000 Hours website. I'm just curious if you guys have noticed any impacts from having that page? Do people get angry about things that you confess you've screwed up that otherwise they

wouldn't have known about? Do they tend to view you more positively? What have you noticed?

Rob: I think probably almost nobody reads it. Or maybe they see that there's a mistakes page and they're like, “Oh, these people are credible, so that's great,” and then they move on.

No, I haven't heard that many reactions to it other than people saying "Oh, it's great that you acknowledge your mistakes."

Julia: Okay.

Rob: ...Certainly no one's given us a hard time about any of the mistakes there. I think by and large people are pretty forgiving if you're like, "We messed up. Here's how we messed up, and how we're going to fix it." I'm sure there are some grumpy people who will still give you a hard time at that point, but mostly, I think, people are sympathetic.

Julia: Do you think that any of these changes that you've been describing are things that you've changed your mind about, or is it other people at 80,000 Hours who've come to see the wisdom of your views?

Rob: Well, I think I got very lucky somewhat early on when I was exploring effective altruism, and trying to figure out how to do the most good with my own career.

A lot of the views that I've been describing are basically the worldview of Professor Nick Bostrom, who's the director of the Future of Humanity Institute at Oxford. I found out about his work back in 2008 and 2009, and read many of his key papers and was like, "Yeah, this basically seems right." I'll describe what those views are in just a second.

Being the kind of person who is perhaps easily persuaded by new ideas, or new papers that I read -- I think, I basically got lucky by taking this package. And then over the last 10 years that worldview has mainstreamed itself. And a lot more people have gradually just become convinced that Bostrom's view is broadly correct, even if they disagree with some specifics, as do I.

Nick Bostrom's view of things -- he's a philosopher -- one part of it is long-termism. So, thinking that most of the moral consequences of our actions, or the most important ones, are effects that will occur after our natural life spans are over. More than 50 or 100 years in the future.

Then you think, well, how could we actually affect the long-term? A common objection is that, “Well, even if the consequences of our actions hundreds of years in the future are really important, I can't predict what they're going to be. So instead I'm going to focus on improving the short-term.”

But then it does seem like there actually are things that we could lean on now that do improve things for hundreds, thousands, maybe even millions of years.

An obvious one would be preventing global catastrophes from which we never recover. You could have a huge war which, say, takes us back to the Stone Age. And then we never develop technology again, and eventually humanity goes extinct. Or you could have an even worse disaster that causes humanity to go extinct in the 21st century. It's pretty obvious that has consequences that affect how the world will look in a million years time, or a hundred years time, or a thousand years time.

There's other potential ways that you could try to change the long-term that aren't extinction focused. That might be, for example, you could imagine a global dictatorship locks us in that we could never escape from that has bad ideas.

Julia: So you guys would be anti-that, just to clarify.

Rob: Anti-that, exactly.

Julia: To be completely explicit.

Rob: So try to prevent that. That seems to be probably the most prominent thing that could happen within our lifetime that would have very long-lasting effects that we could try to change.

Julia: Other than a nuclear war, or an epidemic or something.

Rob: Oh, sorry, I'm saying that whole category, global catastrophic risks.

Julia: Oh, oh, I see. You're counting the dictatorship as one of the catastrophic risks that could-

Rob: That doesn't involve extinction, but still involves most of the loss of value.

Julia: Right.

Rob: Then the next step, I think, in the argument, is wondering – "Well, where do these risks come from?"

It could be from asteroids. It could be from supervolcanoes. Or it could be from things that we make like nuclear weapons, or advances in biotechnology that would be dangerous, or changes in information technology that could disrupt government or disrupt society.

Julia: Right.

Rob: Basically, there's very strong reasons that have been written up to think that the vast majority of the risks come from humanity itself. That it's new things that we're going to do. That the probability each year of those things screwing up civilization is much larger than the natural risks.

In part, simply because we know roughly the annual risks from supervolcanoes, or asteroids, and so on. Because we can look at the historical record and the risk seems to be incredibly low.

We might be optimistic that we're not going to have a nuclear war, but do we really think it's a one in a million chance each year? That would seem way too confident.

Julia: Way too confident that it's not going to happen in any given year.

Rob: Right.

Julia: Yeah. It's funny, people are used to hearing the term “overconfidence” or “highly confident” in terms of predicting that something *will* happen. It's a little bit jarring, or hard to parse, when people talk about overconfidence in terms of thinking that we're *not* going to have a nuclear war.

Rob: Yeah. I just think to say that the risk of a nuclear war in a given year is one in a million or lower would just require you to really incredibly well understand the process that generates this. And that in every respect, every link in the chain is incredibly unlikely.

Which we just don't have much reason to think. I think the annual risk is more like one in a thousand than one in a million. Which basically already means that the risk of humanity destroying itself is larger than all of the natural risks combined.

Then if we think that most of the risks to the long-term future are humanity doing stupid things itself, or failing to coordinate itself such that we have a huge war, or we misuse some new technology, or discover something that would be better not to know that messes us up -- how can we get to a good future that's potentially very big, where people are having excellent lives?

That's going to require a lot of technology to do itself. Basically, Bostrom and I think that we need to basically order the things that we invent, make sure that we invent things such that we're inventing new technologies and ideas that enhance safety sooner, so that we're ready when we later invent more dangerous things that could screw us up.

One easy example is that we invented nuclear weapons -- I think that's the point at which, for the first time, we had the ability within maybe a decade of the first nuclear explosion to kill billions of people very quickly, and potentially really throw civilization off-kilter in a way that might be permanent, where we would never recover.

Now it took actually decades for us to invent permissive action links that make sure that someone can't just go to a nuclear weapon and arm it and launch it, and potentially use it.

Julia: How does a permissive action link work?

Rob: Okay, so basically this is a gadget that you have in the nuclear weapon that ensures that it can't be used unless you have a specific code. A code authorization from the president, or the Pentagon, or whoever else who said "Yes, absolutely, we want to use the nuclear weapons." Basically, for the first decade or two there weren't even to begin with physical locks... Then, eventually, they added these permissive action links which required a code to use the weapons. But they set the code to 00000, famously.

Julia: No!

Rob: Yeah, because they were really worried about not being able to use them in an emergency. So the Air Force, or the strategic command wanted to make ... Basically, the thing they were worried about much more was that they would need to use them and couldn't, rather than that they would be used when they shouldn't.

Julia: Why use them at all then? That's just embarrassing.

Rob: Yeah. But, basically, my point is that it would have been great if we'd invented permissive action links that we were confident would work, and figured out the technology for that *before* we scaled up nuclear weapons, or invented nuclear weapons in the first place.

Julia: Right.

Rob: I think with many new technologies that we can envisage creating in this century, we can foresee the risks to some degree. And we can foresee technologies that we could invent beforehand that would make them safer once they arrived.

I think that is one of the key things that we can lean on, is both inventing technology like permissive action links, but also social technology, or ways of coordinating humanity. Such that when we invent things like nuclear weapons, or whatever the next version of that is, we'll be in a much better position to make sure that they're not accidentally used really badly, or deliberately used really badly.

Julia: To recap, and feel free to jump in and correct me: The updated position of 80,000 Hours about how to maximize your positive impact with your career is to look for opportunities to preserve or maximize the long-term value of civilization. Most of which, or a lot of which, flows through finding ways to prevent humanity from destroying itself in the next couple hundred years, or dealing a severe blow to our growth trajectory in the next couple hundred years.

Rob: Right. I think that's probably not the only way that people can have really huge impacts, but it seems like one where it's clear how the scale of the problem is very large, so the benefits would be really large if we could make this change.

Also, just very few people are working on this, really. We're talking about millions, maybe tens of millions of dollars going to this framework for improving

the world. We think there are just a lot of really high impact opportunities for the kinds of people who read 80,000 Hours within this area.

Julia: Do you direct people to anything besides research, or donating to research organizations?

Rob: Oh, well, we're encouraging a lot of people to go and get experience in the policy world either in London or D.C. We're not sure exactly what government policies we'd like to promote, if any, in these areas. But it seems like it's going to be important to have people who have a lot of experience understanding what impacts different policies would have, when it comes to regulating or deciding not to regulate new technologies.

Also, just focusing on international relations, right? One of the biggest risks is new dangerous weapons that are used by one country against another. Or just that there's a normal war between America and China, which would just be absolutely devastating. So going and getting experience in international relations and diplomacy also seems really valuable.

Julia: Great. Do you try to do back-of-the-envelope quantifications about why... I don't know, I could imagine making a case for "Improving education is going to create a more educated populace, that will then be less likely to vote for a president who will launch a nuclear bomb," or something.

You could tell a plausible sounding story for why a bunch of other things that aren't on 80,000 Hours' list actually do serve the goal of reducing these global catastrophic risks. It sounds like you would have to do some kind of rough quantification to say, yes, you should go into these political avenues instead of these education startups, or something like that.

Rob: Yeah. So this is the question. If you want to improve the long-term future should you do very targeted things or should you do very broad things?

The benefit of the targeted things is that in a sense you have a lot of oomph, because you're focusing on specific organizations, or people, or policies, or technologies where it's very clear what impact they might have on the long-term future. For example, let's say that the U.S. and China are negotiating, or they're at one another's throats and considering going to war, and you're in the room there, and you're trying to negotiate to make sure that they don't have a war at any cost. It's a very, very targeted intervention, and very focused on specific circumstances.

Another approach to improving the long-term future would just be as you're suggesting -- to improve education, maybe grow the economy, to just make people more reasonable in general. Or to improve science across the board in the hope that this would make things better.

I think some of those broad interventions do help somewhat. Others that people are hopeful about I'm less optimistic. I think the main problem there is just -- for example, you talk about improving eduction. There's a lot of effort that already goes into improving education. There's a lot of reasons other than worrying about

the long-term future that people already dedicate their careers and their time and their money to improving that. It just seems really hard to move any.

How much effort, how many careers would it take to improve United States education by 10%, or to make people more reasonable across the board as voters? It seems like just extremely hard. It would take a lot of money, a lot of effort. It's not clear that it would happen.

With other things, like trying to prevent a war between the U.S. and China, there's definitely people working on that. So, some people in the government. There's some non-profits that have some small programs about this. But it's nothing like the movement for improving education in the United States or other countries.

Basically, we think as it just pans out, these broad approaches to improving the world are just already very crowded. People have strong reasons to go into them. If you're looking for somewhere where one person can really move the needle by going into it, you typically want to look at more targeted approaches, where it seems like there's just actually really useful stuff that can be done. Like things that could be invented right now, conversations that need to happen between people, that very few people are working on.

Julia: It still seems like there's a tension -- even if we exclude the fields that are already extremely crowded, like education, especially U.S. education -- it still seems like there's a tension between interventions that have a clear path towards how they could help reduce global catastrophic risk, or increase the long-term expected welfare of civilization... versus interventions that aren't really aimed at anything in particular, but just would fit into the category of “exploration.”

This is a general argument that some thoughtful critics of effective altruism sometimes make. Which is that if you look back at the history of things that have improved the welfare of humanity, most of them were not intended to improve the welfare of humanity.

There was some dude in 18th century England who was like, "I want to build a better textile mill, not because I think it will spark the Industrial Revolution and raise living standards for the next 10 generations of people, but just because I think it will make me more profitable."

Or a scientist who studied electromagnetism, or genetics, or something -- just because that was really interesting, and not because he had some story about how it was going to help humanity.

I think it's hard to argue against a lot of effort, or at least more effort than already exists, going into interventions that seem like they would reduce our risk of catastrophe that no one else is doing. But I'm curious whether you *also* see a role for a lot of people doing random exploration of stuff,that isn't actually intended to help the world, but if you look at the track record of such things, at least some of them do, and those things end up pushing humanity forward.

Rob: Yeah, there's a lot of arguments potentially to get out here.

Julia: Sorry.

Rob: I'm just thinking there's lot of interlocking arguments here. It's hard to get the whole worldview all out at once. We might run out of time.

One thing, as you said, people point out that if we look historically it seems like most of the good was done incidentally by people who were just trying to improve their own business, or explore something that was interesting. That is probably true.

But I think it's a terrible argument, because there was so much more effort that went in to that. There are so many more people who were just trying to improve their business, or studying science because they're interested in it, or because it was their job.

Julia: You're saying we don't have a strong track record of people trying to help humanity and failing?

Rob: I mean, I think lots of people have tried to improve humanity and failed. And some have succeeded, also.

My point is that it could be the case that people who try to do targeted things were a hundred times more impactful on average, because they were so much fewer. Their share of total good done would still be swamped by the people who did good incidentally.

I don't find that argument very persuasive. I would want to actually look at people who tried to do targeted good things who were like, "I'm going to try to figure out what research topic is going to be valuable," and then act on that. Then look at, how well did they perform relative to the base rate of everyone else?

Julia: Sorry, but what you just said does seem like an argument for why we shouldn't have *no one* doing the targeted interventions. But is it an argument for why we shouldn't have a mix of targeted interventions and random exploration of stuff?

Rob: I don't quite see that. I mean, I agree it would be a pretty strange world, I guess, if everyone was trying to do this targeted stuff. One thing is it would exhaust most of the targeted options. They would cease to be neglected, because there just would be too much effort going into that style of doing good. As it is, because most people are trying to do good in a very broad way, most people are trying to improve education, grow the economy, make the world more reasonable...

Julia: Right.

Rob: That's where 99% of humanity's effort is going. Which means that if you're part of the 1%, or the 0.1% percent who are looking for really targeted opportunities, there's a lot of money on the table there to grab, because no other people are looking for it.

Julia: I think this is a really important point, actually. This is a misunderstanding that's in the background of a lot of conversations about effective altruism.

I think when a lot of people hear the arguments that 80,000 Hours and other EA organizations make about the best way to help the world, they're imagining what you think everyone should do. Whereas, I think a lot of your advice is given from the perspective of what, *on the margin,* for the next hundred people who want to help the world, will be the best thing for them to do.

Rob: Yeah.

Julia: And that isn't necessarily the advice you would give if you were giving advice simultaneously to everyone on the planet. Is that right?

Rob: Yeah, exactly. I mean, imagine that we said, "Oh, well, if you want to do a lot of good you should become a surgeon." I mean, obviously, it would be farcical if all seven billion people in the world try to all become surgeons. No, that's not what we're saying. We're just saying on the margin it would be good to add some more surgeons relative to everything else, or relative to what that person might have done otherwise.

Julia: This might be just too hard to answer off-the-cuff, but if you could wave a magic wand and cause some percentage of the world to follow 80,000 Hours' career advice, what would that percentage be?

Rob: Oh, interesting. Where the advice has to be constant, so we can't make it any more broad than it is now, or add more options?

Julia: Oh, I see, so you can't change it as the people-

Rob: Right, so, yeah, I mean, if we could change it as we went, I'd say 100%. Or maybe 50% for some risk aversion thing.

I guess, the advice as it is... maybe one in 100, one in 1,000 something like that.

Julia: Oh, okay. Would you select a particular -- let's say you could filter for some characteristics. What's the group of people that you would want to follow 80,000 Hours advice?

Rob: I guess it's people who are analytical, cautious, curious, trying to be very informed, care about not just going ahead with their own intuitions without listening to other people at all.

Julia: Yeah.

Rob: Yeah, I guess those are some of the criteria that are really important.

I mean, one thing is we think it's very possible to go into trying to solve the problems that we're very concerned about and cause harm. We wrote this article last year about various ways you can accidentally cause harm in your career.

People who have a “run in and break things” mentality actually might well go in and make things worse in a lot of these areas. They're very fragile problems to be dealing with.

But that was all prelude to this discussion of, yeah, so what about approaches to improving the world like inventing new technologies in general?

Inasmuch as humanity’s main problems that we face come from the natural world like super volcanoes, or asteroids, or natural diseases -- then improving technology and growing the economy all makes us larger and more imposing relative to those problems, and puts us in a better position to rebuild after an asteroid, or deflect the asteroid, or control diseases.

But if this fourth point from Bostrom that I was saying is correct -- that, in fact, most of the ... Actually, sorry, it was the third point -- that most of the risk to humanity comes from ourselves, comes from new technologies that we're going to invent, or stupid mistakes that we're going to make... Then it becomes less clear that just empowering humanity at the broadest level is actually sensible.

Because while you improve our ability to solve the problems that we're creating, you also potentially grow the problems as well. Because the whole problem in the first place was that we're running ahead of ourselves inventing things, changing the world in very dramatic ways, running the risk of destabilizing everything and ending it.

Julia: That sounds like it would be an argument against broad interventions to increase technological or scientific progress. And it would be an argument for individuals who want to have a positive impact going into scientific or technological research, but specifically the research that would produce the safety promoting technologies, instead of the safety decreasing technologies. Is that right?

Rob: Exactly, yeah.

Julia: You did a great episode a few months ago with Tyler Cowen, who wrote the book *Stubborn Attachments*. That was Tyler's argument for long-termism, but the main intervention that he promoted in his book to promote long-term welfare was increasing economic growth, as opposed to reducing global catastrophic risk.

Did you feel by the end of the episode understood why your prescription and Tyler's prescription for maximizing long-term welfare was so different?

Rob: Yeah, somewhat. I mean, it was very funny reading that book because I agree with 90% of this, and then we diverge pretty seriously in the conclusion.

One thing is that it's not clear that Tyler and I really disagree all that much. Because he doesn't actually say that economic growth is the best way that one person could take to improve the long-term future, or to prevent human extinction. In fact, I think he agrees with practically everything that we said, about that the risk of extinction is higher than people think, and that there's

useful stuff that could be done to reduce the risk of human extinction, that people could go and work on in a targeted way.

He explained in the interview that the reason he was talking about economic growth so much was that he thought that many more people would be likely to take that advice that it was a lot easier to get people to go out, and just try to make more money, and be more innovative in their jobs, invent new things. That's something that potentially a very large fraction of the population can do. Whereas, the advice that 80,000 Hours is giving is something that it's hard for most people to know exactly how to act on.

Julia: That's a tough trade-off, though. Because as you said a few minutes ago people already have incentive to go out and be innovative and make money. There's already a lot of effort going towards doing that. Whereas, there isn't as much effort going towards figuring out ways to reduce catastrophic risk.

Rob: Right, yeah, so that's the argument that I made back to Tyler.

Julia: Right, yeah.

Rob: One can argue even about whether speeding up economic growth is good or bad. It's not entirely obvious that it make things safer rather than riskier.

Although I think Tyler does offer some pretty good considerations in favor of thinking that faster economic growth, on balance, makes the world more secure rather than riskier. But that's an active debate that people have.

I think the main argument against this is just that it's an insanely poorly leveraged approach to reducing human extinction. Let's say that you were thinking, yes, I want to make sure that human civilization persists for hundreds of years so I'm going to start a business and try to make it bigger and grow GDP.

It's true that maybe, that more people can see opportunities to grow GDP than to reduce the risk of war between the U.S. and China. But you're losing so much oomph in the fact that the causal connection between growing GDP or growing the economy, or even inventing new technologies towards preventing human extinction, I think is very weak. And it's unclear whether it's positive or negative.

Plus, just the fact that we spend $1 trillion already on R&D globally. And about $60 trillion is paid out to people to do their jobs to engage in economically productive activities. So in a sense it's the very background cause area, is just growing the economy.

Certainly, it's very hard to say that it's neglected relative to other things, all things considered. Especially in a market economy where people have such strong incentives to make money, and in the process do the things that just grow the economy in a very general sense.

Julia: I mean, you could argue from the perspective of, say, a philanthropist or maybe a policymaker or something -- as opposed to an individual who's a participant in

the economy -- you could say an area that's neglected is figuring out how to increase technological growth. Which then feeds into productivity. Which is something that Tyler's written a lot about, that our productivity is stagnating.

And there isn't really a field yet devoted to “Why is scientific progress slowing down?” I did an episode a few months ago with Michael Webb who wrote that paper, on Are Ideas Getting Harder to Find?

If you weren't worried about global catastrophic risks you could make the case that figuring out why scientific progress is slowing down, and how to speed it up again, is high-impact and neglected, and at least somewhat tractable or plausibly tractable.

Rob: Well, yeah, so if you really thought that global catastrophic risks were impossible, that civilization was just going to continue... actually, it's not clear to me why any of it really matters. Because as long as we just keep growing each year then we're going to get there eventually, and there's no particular reason why we have to grow super quickly, I suppose.

One thing is that the universe is expanding. Interestingly, that's the most important argument, is that galaxies are receding from us. So the longer we wait, or the slower we grow as an economy, the less value they'll be available to harvest it at the end of it.

Julia: Wait, I'm surprised to hear that you don't think that the rate of growth matters. If growth makes people better off, than isn't more people being better off sooner, better?

Rob: Yeah, does it? I mean, I guess... It's very hard to shift frame from thinking about the flow of goodness in any given year, and that we're trying to increase the flow so that next year more value is generated than this year, to thinking about it more as an endowment that we have. That we have eternity, essentially, or we have as long as we want in this universe as long as we don't destroy ourselves. The limiting factor is how much energy and matter can we harvest, can we reach in the galaxy, and then use in the entire universe? And then at some point, whenever it's ideal, to convert that into value at a later time.

As long as you're continuing to grow it's not clear. The reason to go faster then would be, well, we managed to capture more of the universe before it recedes from us outside of the accessible universe. But it's less obvious to me the fact that we would generate more value next year than this year is really so key.

Julia: Interesting.

Rob: I guess this is taking the long-term, the utilitarian consequentialist long-term view very seriously, I suppose. On other values, where you're concerned about people alive now in particular, there's more of a commonsense view why we want to speed up improvement a lot.

This is one thing where I think that the global catastrophic risk crowd deviates from potentially commonsense. Or potentially from people who are working in business who are thinking, yes, we're going to grow the economy as quickly as possible so we can generate more value next year.

Whereas, I think we're thinking more "How in the very long-term can we get the most value out of everything?" From that point of view it's much more about stability than it is about growing really quickly. Growing more quickly, but with a greater risk of catastrophe, is a terrible trade-off in this view. Because the universe is only receding at a rate of one billionth per year. Basically, if you could get there a year sooner than that would only be worth one in a billion risk of the whole thing ending.

Julia: Right.

Rob: It's much more focused on stability than speed.

If you thought that just improving technology did stabilize things, and one could make arguments in that area, then working on science and technology policy to figure out how we can do innovation more quickly could potentially be really valuable.

Because I agree not many people are thinking at that level. Patrick Collison has done a bunch of interviews lately. You read that article in The Atlantic, that we can link to.

Julia: Right, yeah.

Rob: There are surprisingly few people thinking at the policy level about how to make science research proceed much more quickly.

But just personally, I'm not convinced that increasing the total amount of scientific research that we do each year is even positive, or certainly is among the more leveraged ways to improve things. I'd be much more focused on, "How do we improve societal wisdom and prudence, so that when we have more advanced technologies we're more likely to use them well, and not use them against one another, or just in very stupid ways?"

Julia: One other thing I wanted to ask you is: When we were talking about who should follow 80,000 Hours advice, and what percentage of the world would we want following your advice versus doing other random stuff like exploration... do you worry at all about people following 80,000 Hours advice who otherwise would have pursued some kind of eccentric passion that could be the next 18th century textile mill of the 21st century? Do you worry about getting rid of the potential innovators?

Rob: Yeah, so I suppose inasmuch as I'm very much within this frame of "Advancing technology in general doesn't help," I suppose I wouldn't be so worried about that. But that is a somewhat kind of intuitive view that I'm not sure about.

We suggest 10 priority paths which are the 10 career paths that we think are most likely to make a really big difference to improving the long-term future, currently. One exception we have in your whole process for deciding a career is if there's something that you're incredibly well positioned to do, that no one else is able to do, that seems like it would have a really large impact, then there's a pretty strong case for just sticking with that, rather then switching into the paths that we suggested.

Probably the first thing that we think people should do when they're planning their career is to try to figure out what problem they're trying to solve, and potentially to do that before they figure out what method they're going to use, or think too much about what they're specifically passionate about.

Just because we think there's a hundredfold, thousandfold, possibly even greater variation in how much bang for the buck you get trying to focus on solving different problems. So just making sure that you work on something that is enormous in scale that other people aren't working on, where you can make a difference, just seems like it's one of the prime considerations.

Now all of that said, people do sometimes worry that we reduce people's creativity or exploration in their careers. I think if people actually read us very closely, because we're so focused on people doing stuff that's neglected, because there'll be low-hanging fruit there that other people haven't taken -- in a sense, we are extremely in favor of innovation and exploration.

But one way that creating a career guide in general limits you is that we have to put something on the page. Suggestions that apply to more than one person, that can be generalized somewhat. Which can cause people to think it's only these things, these very generic stock positions, that are available.

But very often the best opportunity for you with any problem is going to be something that only you know about. But, of course, we can't put that down because that's specific to each individual. So we can think of that in the one-on-one career coaching.

That's something that people should watch out for is that we're not saying that you should just go into some position that's extremely codified and well understood. Often it will involve finding something that's very unique to you and your unique circumstance.

Julia: Okay, well, we'll link to the 80,000 Hours career guide and just ask people to just mentally insert those asterisks after all the advice that 80,000 Hours gives.

Rob, before I let you go I wanted to ask you to nominate some resource, whether it's a person, a book, or an article, that has influenced your thinking in some way. Or that you have substantial disagreements with, but that you've gotten value from engaging with. There's a lot of possibilities there, do you have anything that fits that?

Rob: Yes. So a lot of your readers will be familiar with James C. Scott who wrote what's the classic anarchist book about? Oh, Seeing Like a State the book about projects to improve the world. Modernist projects by governments where everything is standardized have often failed and failed really catastrophically.

I'm something of a defender of high modernism, of these very organized ways to improve the world. I think people underrate, for example, how much high modernism just improved agriculture enormously, and made us much richer in many ways in the long-term. Even though, obviously, the Soviet Union and Stalin were very catastrophic to begin with.

Anyway, he's written another book which was maybe my favorite book of last year, called *Against the Grain*. It's a deep history of the first states. He goes back and looks at how did the very first countries of just thousands of people form in 5000 BC. And it's just an absolutely fascinating history, and includes many unexpected things about the nature of those very first city states to begin with.

Julia: Excellent. Yeah, James Scott is great. At least one other guest -- and I, on a different podcast -- have recommended Seeing Like a State as a book that really influenced our thinking. So it's nice to get a contrarian perspective on this contrarian book.

Rob: Yeah, I can stick up some links of some people writing reviews where they critique and review.

Julia: Oh, great, that would be great.

Rob: It hasn't been rare for people to just say “No, actually, I want to defend high modernism against this”?

Julia: Yeah, it's really not trendy.

Rob: Not fashionable.

Julia: Not fashionable, exactly.

Rob: Just one other really quick one is *Destined for War* by Graham Allison, which is about trying to assess the probability of a war between the U.S. and China in the 21st century, and looking at historical analogies to try to guess at that.

Julia: Excellent.

Rob: I think it's an underrated book. It might be interesting to ... I'm hoping to interview some people on the 80,000 Hours podcast about how do we make China and the U.S. get along and cooperate in future.

Julia: Excellent. Well, Rob, thank you so much. It's been a pleasure having you on the show.

Rob: Yeah, it's been so much fun. Hopefully, we can talk again soon.

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