New York City Research Event, June 20, 2016 – Open Philanthropy Project

This transcript was compiled by an outside contractor, and GiveWell did not review it in full before publishing, so it is possible that parts of the audio were inaccurately transcribed. If you have questions about any part of this transcript, please review the original audio recording that was posted along with these notes.

00:00 Speaker 1: Recording again. And, again, anyone just let me know if you want me to take anything out of it. So now I'm gonna talk about the Open Philanthropy Project. So the Open Philanthropy Project is... It has a lot of similarities and some very important differences to GiveWell. So, basically, we were working on GiveWell when we met Cari Tuna and Dustin Moskovitz a few years ago. And they were, in some ways, in a similar situation to what we were when we started GiveWell, and in some ways very different. When Elie and I started GiveWell, we were kind of like, "We have full-time finance jobs. We would like to give a few thousand dollars this year. We'll have a couple hours to think about it. We'd like something straightforward, and really good."

00:42 S1: Cari and Dustin were saying, "We are looking to give away a large personal fortune. We have several decades to do it. We have the resources to hire our own staff, to spend decades thinking about these things, and we'd like to do it really well." And the commonality is that, like me and Elie, I think Cari and Dustin were not finding a lot of public intellectual discussion and debate about how to give as well as possible. And so, the similar call for someone to do the best job you can, figuring out how to give, and then write about it publicly so others can learn from it, and we don't all need to reinvent the wheel, was there. But a big difference is the kind of giving you wanna do, because I think there's much less call for things that are proven, cost-effective, scalable, verifiable, and much more argument for doing things that might be very risky, or very unconventional, or things that no one else can do. Because they might involve creating organizations instead of funding existing ones, or transforming organizations, or just generally doing things that require a lot of trust and context and staff to see the case for.

01:47 S1: And so, the Open Philanthropy Project has a similar mission to GiveWell. We try to find the best giving opportunities we can, and write about it publicly so others can learn from our work. But it does have a completely different model, and different approach, and increasingly different staff, and day-to-day. So I was co-founder of GiveWell, but I now work on an Open Philanthropy Project almost exclusively. I relate to GiveWell a little bit more as a board member, so I get involved in the mid-year and end-year discussions, but it's not what I spend my time on. And Open Philanthropy Project is currently part of GiveWell, but we are working on changing that, because we believe the two are fundamentally separate missions, and in particular, Open Philanthropy does a lot more debatable things than GiveWell. And we want the two to be separate organizations, with separate brands, so that you can give to one without giving to the other, for example.

02:44 S1: So we're working on the process of separating them. When they separate, I will work for the Open Philanthropy Project. I will be the executive director, and then GiveWell will be a separate organization. And where GiveWell looks for proven, cost-effective, scalable charities, Open Phil', instead of looking for charities, we look for causes. And instead of proven, cost-effective, scalable, we look for important, neglected, and tractable. And so, in other words, if we can find an issue, a problem in the world that is not getting enough attention for how important it is and for how tractable it is, then what we can do is we can hire around that issue. We can build expertise and context around that issue, and get to the point where we're able to look for the best grants in that issue. Less by prioritizing evidence, though we do still care a lot about evidence 'cause that's who we are, but more by having the deep context, having the relationships, having the expertise, having the background to be able to spot what the best giving opportunities are. Which is a very subjective thing, and really relies on the people at Open Phil building trust relationships with each other. So Open Phil' looks for, like I said, important, neglected, tractable causes.

04:00 S1: Another thing to know about Open Phil' is we have an interest, it's not the only thing we do, but we have an interest in what we call hits-based giving. So we have a post about this on the Open Phil' blog, which, by the way, recently became separate from the GiveWell blog. So there's a website, openphilanthropy.org. It has its own email list. It has its own blog. If you're not subscribed to it, then you're not seeing any news about Open Phil', so that's something to keep in mind. But we have this interest in hits-based giving, which is this idea that you might try a lot of things that have a low chance of working, but if they worked, would be a big enough deal to make up for everything else.

04:40 S1: And this is similar to how a lot of venture capitalists approach their work, where you might invest in 10 companies at an early stage, and nine of them flame out, and one of them is Dropbox, and you did great. And that's similar in philanthropy, that I think, I can point to some cases where a philanthropist did something so forward-looking, and so huge for the world, that it would have been easy for them to have 10 other failures and still come out ahead, in some sense, of just doing the easy thing. And so, that's why we're interested in hits-based giving. And when you do hits-based giving, things look very different. And so, in my opinion, when you're looking for unlikely but potentially huge

successes, you often are gonna be attracted to things that are not already popular with expert consensus or with conventional wisdom. You're gonna be looking for things that maybe only a small number of people can appreciate, or have the context to really see the case for.

05:35 S1: And so, a lot of the way Open Phil' is structured is, we pick our causes and then we decide which staff are gonna be specializing in those causes, and often, we hire specialized staff. And then, once we have someone... Let's say we have a cause, like criminal justice reform, we'll have one staff member who is really leading the way on that cause, and the rest of us will look to defer to her and ask good questions, and make sure things are making sense, and make sure the whole case holds together. But in the end we want to have one person who has a great deal of context, who has all the relationships, who has all the background knowledge, who can consider all the arguments but ultimately making their own personal subjective call about whether something is the best thing. We think that's a more promising route to doing really high upside, and perhaps high risk things, than trying to have a system that's more based on universal standards of evidence or boxes to check, or needing any universal staff consensus which is not something we aim for at Open Phil'. So that's a basic introduction to what we're about and our values.

06:40 S1: In terms of where we are, we spent the first few years of Open Phil really trying to pick causes. And so we had a small staff, and the thing we wanted to do was find areas, issues to work on, that we could really commit to. And so in 2014 we announced a variety of... By the end of the year, we picked a variety of focus areas. And these are areas that we've chosen as important, neglected, tractable. I'll give examples in a second. And then in 2015, most of what we did that year was just hire, was just look for staff who could help us pursue those causes. And now finally in 2016 we are in a position to do a fair amount of grant making, which we were never able to do before. So total grant recommendations in 2015 were about $16 million. In 2016, I'm expecting something north of $50 million, because now we have the staff to work on a bunch of causes and do things this way. So that's where were at.

07:42 S1: I'm gonna start by just talking about what causes we're prioritizing on the US policy front. And that's an early category that we did some investigations on, picked some priorities, did some hiring. That'll give you a sense for just how we do things. Then I'm gonna take a break and do questions, and then I'll talk about some of the other causes we're doing and Open Phil generally.

08:03 S1: So one of the broad areas we wanted to look at, or one of the broad categories, this is bigger than a cause, this is a category of cause, is this idea that when you're doing philanthropy, if you can fund people who make arguments that then change the direction of public policy, like what governments are doing, you could have a big multiplier on your giving. And even if you again, if you fail 90% of the time, but 10% you play a real role in a policy change you could end up with a pretty good value for money overall. We started with a focus on the US for totally pragmatic reasons, and as we've gotten deeper into our focus areas we've gotten more open to looking at other countries as well. But we did feel that when we were getting started we needed to stick with a policy landscape where it was feasible for us to network, to meet the right people, to do the right hiring, and to just understand the general background.

09:01 S1: I think it wouldn't have been very practical to start a policy program that was, let's say, focused on Germany. I think that would have been a travel and logistical and linguistic nightmare, and wouldn't have bought us really a lot of advantages that I can name. But it's not because we value the US more. It was a pragmatic decision to chase high value in this area. And what we did is we made a list of issues we could imagine working on that seemed like they could plausibly be important, neglected, and tractable. So, there's some areas of policy that I could just look at and say, "I don't think there's any chance that that is as important as some of these other things we're doing." But for plausible contenders we tried to investigate and say, "Here's a policy area... Let's say criminal justice reform, 'cause that's one we actually picked, "How much good could we accomplish with plausible reforms to the criminal justice system?" And this was something we did with our cost-effectiveness calculations that use a flexible framework. That could be a long conversation, but we look for importance in terms of how much good could you imagine coming about from policy change.

10:09 S1: And then we look for neglected-ness. We say, "Who else works on this issue? Are there things that people could be doing that no one is doing because there isn't money to do them?" And then finally we look at tractability and we say, "Could we imagine a win here? Could we imagine an impact here? Are there special reasons to think that this cause is promising in that way, is politically doable or is not?" And so causes that we ended up picking... Criminal justice reform. Briefly, the US has a very high rate of incarceration compared to any other country. Really questionable whether all that incarceration is getting us much of anything on the margin in terms of public safety. And so the idea is that if we could have better policies, we might have a lot fewer people in prison that could avert a lot of suffering, as well as saving the society money. And could also hopefully not have resulted in much of a hit, if anything, to public

safety.

11:06 S1: David, who's here tonight, is working on a very thorough, maybe arguably unprecedentedly thorough, review of the literature on the effect of incarceration on crime. Seeking out the highest quality studies, and in many cases rerunning the data and code, reconstructing them himself, and looking for all the holes in them. And trying to nail down this question of whether we really do believe we could reduce incarceration without affecting public safety, or with having only a minor hit. And preliminarily it looks like, and this is where our intuition go in, is that the answer is yes and that there is room for improvement there. The reason this cause ended up as a focus area of ours, in terms of importance we scored it as a moderate relatively. Which means it's very important, it means there are some other causes we thought had higher importance. Neglected-ness we also had as a moderate, it's not the biggest policy field. It's not as big as...

12:02 S1: Well, there's a lot of areas that are bigger. But in terms of tractability, it's a bit of a standout because most US policy issues you can imagine working on, a lot of the really big ones, a lot of the really important ones are kind of federal issues only. So immigration being an example, very hard to get much done by just advocating in California for better immigration law. And DC is not a very exciting place to be pushing for change right now, as some people may know. With Criminal Justice Reform, a lot of the juice is state and local. And so you get a chance to pick your battles. Other issues, I think, just there's a stably changing dynamic around criminal justice that we think presents opportunities in a real and long term way. And so when some people think of political opportunity, they might think of something that's in the headlines and maybe something that'll happen in the next few weeks.

12:52 S1: We think that's not the right time-frame for philanthropists. So when we think about tractability, we think about durably changing factors that could matter for the next several years. And with criminal justice, over the last couple of decades, the rise in incarceration, the rise in expenses, the tightening of state budgets, the lowering of crime rates, which we actually think is not necessarily connected to the rise of incarceration, have created an environment where there's been a lot of reforms and we think there could be a lot more. So criminal justice was chosen on that basis. And we spent several months looking for someone to hire to lead our work here. We hired Chloe Cockburn, formerly of the ACLU. And basically our process is that Chloe is very well connected in the field, has a lot of background, has a lot of expertise, talks to everyone, tries to find the most exciting things and then investigates them and answers a series of questions about how much value we're getting per dollar.

13:48 S1: Probabilistically, questions about how good the leadership is, what the downsides of a grant would be, why we should think the grant is special or outstanding and then, we review the case and we discuss with her, and then generally... In most cases, we end up making a grant when Chloe recommends it. And so that's the system, that's the process for Open Phil'. Other things that we're working on, another focus area of ours is farm animal welfare. So animals are treated horrifically on factory farms. These are the industrial agriculture that gets you your food. And we believe that there is actually a fair amount of interest in America in animal rights and animal welfare but it does not tend to focus on farms, on factory farms where the vast majority of animal suffering is occurring. And we think that's a cause that has high importance depending on exactly how you wanna value animals, but certainly the numbers are huge and the amount of suffering is enormous, and also is relatively neglected for reasons I just said.

14:54 S1: And the tractability, we actually didn't really know much about it when we started our search and I think on a lot of policy issues, the tractability is just a huge question mark 'cause you never know what's gonna happen. But we hired Lewis Bollard who came in and really convinced us that there's a particularly promising area of farm animal welfare right now, which is campaigns against corporations to encourage specific reforms in terms of treating animals better. And this is unsatisfying to a lot of people, I think a lot of the big donors in the space right now don't like it, because they don't like the idea of acknowledging or working with the current system. They don't like the idea of treating animals maybe less badly instead of just not having them in the farms at all. But the fact is, we think that there's been enormous successes of these corporate campaigns and it's just been momentum and wins beyond what we would normally see in any issue.

15:50 S1: Some time and basically within the last few months, there's been just a complete epidemic in some ways of cage-free pledges by groceries, by fast food companies, campaign against by the groups that we are supporting. And our role in that victory, I think, is partial and as ambiguous and I can get to, but that I think is going to affect an enormous number of chickens. These pledges tend to only sell cage-free eggs. We are convinced that it's going to affect the chickens very positively and it's a win. It's momentum, it's the kind of thing that we think can pull more people into the field of pushing for animal welfare in a way that just kind of encouraging people to be vegan, and mostly just not having an effect with that I think would not have the same kind of promise. And we believe that getting wins and

getting momentum is historically very important social movements and political movements, and for pulling in talents. So we're incredibly excited about cage-free campaigns.

16:50 S1: We've done a series of grants to support groups working on them in the US. And now that we have tried to cover everything we can there, we're looking internationally. So we're looking at China, India, EU, Latin America, Japan, and trying to see how much campaigning can we get done for better standards for animal welfare in those countries. So there are other focus areas in US policy. Those are the only ones we have full time people working on. We have some kind of quasi focus areas that Alexander Berger spends part of his time on. He's a policy generalist. But I've been talking for a while so I'm gonna stop there and take questions and then after a while, I will talk about some of the other causes we do. Yup.

17:31 Speaker 2: Is this kind of to give [17:32] ____ or what are the other organizations that you see as having like minded approaches to things that you are fan of [17:43] ____?

17:45 S1: Are there other organizations that I see as having a like minded approach to things that I'm a fan of? Certainly. We make grants so there's a lot of organizations that I look at and I feel very excited about and I think are doing great work. I think it really depends in what way and in what kind of alignment. In terms of some of the groups that we felt excited to support in a very general way, Center for Global Development is a think tank that works on encouraging rich countries, especially the US, to take on policies that are more friendly to the global poor. They have a lot in common with GiveWell in terms of they're really into transparency, it's very easy to tell what they're doing, what they're arguing, and why they're arguing it. We think they have high intellectual standards and have had a great deal of impact. And we've actually done a historical case study in how they were founded, which we think a philanthropist had a major role in, and study of their impact over the 13 years or so, maybe 10 years, since they were founded, and we were very impressed with what they do. So I guess that's one example.

18:53 Speaker 3: So it seems like the approach with Open Phil' is different from GiveWell in that rather than hovering above the world of a particular area of charitable giving you kind of immerse yourself in it and you go and you find experts in it and so forth. And then I'm wondering if you're worried about being captured by their conventional wisdom in a particular area, which it seems like GiveWell is super duper skeptical and tries really hard to avoid that. Is that something that you think about and how could you mitigate it?

19:24 S1: Sure, so, immersing ourselves in the world and listening a lot to experts versus being above it and being very skeptical, I think in some ways the two represent different world views that I think both have something to recommend them. People can pick which one they're more excited about. I personally think there's something to both. It's a little bit more of a gray area than maybe that sounds like. GiveWell does some immersion, we visit our top charities, we go check them out on the ground, we get to know the leadership, we get to know the development economists who work on the studies, we want to have a rich contextual sense of the issues so that we interpret the data intelligently.

20:04 S1: We don't think the whole thing can be reduced to a formula even on the GiveWell side, and on the Open Phil side we definitely immerse ourselves and we definitely try to find people with context, but we also try to cross the gap from people with deep expertise to more people like me who are making these high-level cross-cause decisions, who don't know everything. I do ask a lot of critical questions and I point out when things don't make sense and I spot-check things and I push on things and I say, "This isn't adding up, what is the reason for that? This is a big, important-seeming claim, can we push on that?" And it's not just me, and it's not even just the staff in that kind of position, the cross-cause selection position.

20:46 S1: It's also just a general habit that we try to identify claims that seem important and that don't seem obvious and that we feel we could learn more about and we try to learn more about them. So, David's work is an excellent example of that where, sure, if you talk to people in the criminal justice reform field, they all agree it's completely obvious that putting people in prison does absolutely nothing to reduce public... This is a little bit exaggerated, but there's certainly a lot of people who you can get a sense of an echo chamber, why would you ever even put someone in prison?

[laughter]

21:16 S1: But that's why we have this incredibly thorough literature review going on that I think is very skeptical and is very "Let's look at the data and be data-based," and similarly we've had some arguments about whether cage-free systems are really better for chickens, well there's big literature on that and we're about to start diving into it. We already know a fair amount about it and we've already got our view, but we're gonna try and dive into it more. I think

there is a mix, that said, I think if you want to do the hits-based giving approach, you should be less afraid of being in a bubble.

21:50 S1: If you want things that definitely work and are hard to argue with, you should be very worried about being in a bubble, be very skeptical, look for evidence in everything, that's more of the GiveWell way. And if you want some massive wins that no one else saw coming, sometimes you want to be in a bubble. Because sometimes you want to be believing something and seeing something that most people in the world don't see and the people who do see it are the ones who are most interested in it and so those are the ones you end up talking to. And so I think these bubbles, they can be scary but they can also be productive, and I think some wonderful things have come out of people in intellectual bubbles.

22:24 S3: You did move to Silicon Valley.

22:25 S1: Yeah, exactly.

[laughter]

22:28 S1: So we do try to strike a middle ground, I do think we are doing something different form just saying "Screw critical thinking, we're just gonna listen to our friends." But I also think we are doing less of the "We're going to pass up this giving opportunity just because it feels like an echo chamber." I think some things do feel like echo chambers and you have to ask yourself on the merits, what's the best argument you've heard against and evaluate that. Yep?

22:57 Speaker 4: I was just interested in a little more about, you had just mentioned the evidence for cage-free actually being better and that was something I... I'm vegetarian and I'm pro the approach of making incremental changes, but my current impression was cage-free meant instead of like 100 chickens in one foot by one foot cages they were in a 100 foot by 10 foot cage. I didn't do the math right but whatever. And that it's approximately as bad and also they can peck at each other and stuff and I'm curious, what's your, I don't know at a high-level understanding of that?

23:31 S1: High-level understanding of cage-free chickens, so cage-free has this specific meaning, it requires this specific certification, there are definitely questions about whether cage-free pledges translate to using those systems and meeting those specifications. Those are things that we believe continued advocacy will be needed to stay on top of, but in a nutshell, there is a lot more space per chicken under a cage-free system. It's per-chicken, not just, you know.

23:56 S4: Okay. [chuckle]

23:56 S1: Yeah. There's a lot more space per chicken under a cage-free system and the main literature we're working off of is the best literature review we know of, that basically tries to survey all the studies of different systems, and then score them according to which factors there are for chicken welfare. That could include things like mortality, where we think it's ambiguous and maybe even worse in the cage-free systems. But it also includes, are the chickens able to engage in chicken behaviors, in natural chicken behaviors? I think that is where you see a big bump up on cage-free. There's just several really core chicken behaviors that chickens are able to do with that. The details escape me and this is why we have Louis. I've definitely reviewed this and looked at this and that part is pretty convincing to me. You get a serious change in welfare there, but that is something we do want to look into more. Yep?

24:52 Speaker 5: You kind of described the dichotomy of the factory farming system as either the approach of working with companies and trying to get them to lessen the horribleness of what they're doing versus telling everyone to go vegan and changing the cultural understanding that way. But there is, I think, a third prong of the issue that I'm wondering if you've looked into at all. Especially because you mentioned that this was one of your political issues, which is just the fact that there is almost no prohibitive legislation on factory farming whatsoever, both federally and state by state. As a result a lot of small farming towns kind of get huge many thousand animal factory farms coming in and diluting the water system and having the animal welfare issues. Is that a side issue that you're looking into at all?

25:46 S1: Sure. Are we looking into legislation on factory farming? It's definitely something we could do. It's definitely something that's on the list of things to consider, and I think there has been some important legislation or legislative campaigns in the past that have put certain issues on the map. So the factory farming strategy is to have a list of all the things we could do and go down the list in order of what we guess is best, and if it looks amazing we should recommend some grants right away, and if not we should move to the next thing and make a landscape. And so far, it just feels like... Louis came in very excited about these cage-free campaigns and since he's come in we've seen them get

even more successful than any of us had anticipated. And just the wins there and the value for money. I'm personally, my own best shot at the philosophy questions, do not make me think that animals should be given a large moral weight relative to humans. But even I look at these numbers and I just say, "Wow".

26:50 S1: It's like multiple chickens moving out of cages per dollar spent or something like that, is our calculation. The corporate campaigning looks really good and it just looks really effective. It's just very different from passing a bill, because you only need to convince one decision maker who is worried about their brand, and it's often not a huge extra expense for them and things like that. And as soon as we finished that we said, "Is there much of this going on in other countries?", and it was like, "Well there's a lot of other opportunities to fund better work in other countries." I think the legislative stuff is definitely on the table but frankly I think there are other things that just look more promising to us, and that is a subjective judgement but I think it's not possible to get to an objective view on that. And I think instead what we wanna do is move quickly to other things that are most exciting to the person that we've selected as our best shot at the person who's gonna make good choices here. Yup?

27:49 Speaker 6: With the open felonies you're getting into issues that are more political people, have you thought about what kind of issues that's gonna present for your organization funding structure? I'm assuming the fund structure will support ending up with things that don't fit conventionally together to piss off one side or the other, and just have you thought about how you'll manage or deal with that, those kind of issues?

28:11 S1: Sure. How are we approaching potentially politically controversial issues from the stand point of our brand? Point one is to separate GiveWell and Open Phil. Because I think GiveWell does non-controversial things and doesn't need to start getting heat from everyone who has an issue with controversial decisions Open Phil is making, and this is a part of that trade-off again, I think. If you want things that are non-controversial that work, I think GiveWell is a good choice, and if you want to be really ambitious and shoot higher I think it would be a mistake to just say, "And we can't do anything uncontroversial." It's separating them is answer number one. In terms of do we want to try and be bipartisan or take sides? The way that we try to do it is we try to look for issues... We try to look at each issue and try to decide on the merits of the issue.

29:02 S1: We are not really interested in trying to balance the number of partnerships we have with each political party. We're also not interested in announcing that we are loyalists to one political party, and we're just gonna do everything to help that party and everything they believe in. So to us it's really just case specific and we aren't really working to be balanced and we aren't really looking to be loyalists, we're just looking to do the right thing on a given issue and finding things that are important, neglected and tractable.

29:30 S6: Yeah I think, I was trying to ask more like... I'm assuming you want to work with things even if they have to be controversial, and that the odds are they won't all line up in one direction, so you can just have a consistent set of funders. Do you feel confident to be able to pursue that sort of direction or is it just day-by-day?

29:48 S1: Do we feel we'll be able to pursue multiple directions? I think we can, I think our position, what we do does not really require loyalism. In other words if you say to someone, "We love your work. We would like to support you financially." They don't usually come back with, "But are you... What's your position on this?" They don't say... As a funder I think that's a reasonably okay way to be. I can certainly imagine that there are issues where we could imagine that getting involved in the issue could jeopardize a lot of our other work very seriously, and if we did that we wouldn't necessarily be public about that. Because our mission is to share information so that other people can learn, which is different from sharing everything we ever decide. So we think it's perfectly consistent to say, 'We want to be open, we want to talk about this stuff intellectually. We wanna help people understand how we're thinking. Also some of the things we do are not necessarily going to be talked about in public because when we perceive that the things we could do could jeopardize our other work." Yup?

30:55 Speaker 7: So, in the area of say farm welfare, there's a lot of things you could do that doesn't involve political lobbying, trying to get laws changed. Are there such opportunities in the area of criminal justice reform or is it mostly just trying to get laws changed?

31:10 S1: Are we just trying to get laws changed in criminal justice reform? Because it's true that in farm animal welfare, there's a whole set of important decisions makers that, they're corporations and if they change their minds then that effects animals. I think that's not as true in criminal justice. This is a policy issue and it's really hard to do much about that issue without getting policy changed. Certainly it's not the case, there are a lot of things we can do that are not just about trying to get policy changed. For example prosecutors, there's a general feeling that a lot of prosecutors

are trying to be maximally tough on crime and lock up as many people as they can and that's how they relate to their job and that's what they believe makes them a good prosecutor. If you could change that culture by highlighting some of the ways in which a prosecutor can be good other than getting people locked up, such as saving money for their community and averting unnecessary suffering, then I think that is an area that we have active interest in and have made some grant recommendations in.

32:14 S1: So there is some of that. But generally everything intersects at the legal system because all the choices about who goes to jail and prison are in that system. So, I'm going to now talk about some of the other causes we're working on and then I'll take questions 'til the end. Another kind of broad category thing we're interested in is what we call global catastrophic risks. These are, climate change would be an example, asteroid strikes would be another example, pandemics would be a third. These are things that could happen that could be very bad for global civilization. The theory here is there's a couple of arguments for why you might expect these to be good things for a philanthropist to work on. One is there's comparative advantage. So if we're all worried about some kind of global issue knocking all of human civilization off course, it's not always easy to point to one company or one government whose job it is to worry about that. And that makes a natural fit for philanthropy.

33:12 S1: Another part of the reasoning here is that a lot of the basic dynamics of disasters is that as they get bigger they don't necessarily get less probable as quickly as they get bigger. So it might be that a really bad pandemic is fairly unlikely and a super, super, super bad pandemic is only more unlikely than that but not a great deal. And so the idea is that because of how interconnected the world is certain things going wrong could have really amplified global self reinforcing effects. Therefore we can imagine that if we did things to help global civilization become better prepared for something that threatens all of global civilization as an interconnected whole, we might be able to get a multiplier and get good value for money. You know.

34:00 S1: The thing we did was similar. We looked at all the things we can imagine being big threats to human civilization. We ranked them according to importance or scariness, in this case, if you prefer, neglected-ness, and tractability. The two that we've really honed in on that I really do believe are especially strong on these three criteria and by a decent margin. So one of them is bio security and pandemic preparedness. Basically if you told me that human civilization was either going to go completely extinct or fundamentally get knocked off course in the next 20 years, or the next 50 years, or the next 100 years, and I had to guess how, a pandemic would be one of my top guesses. And that could be either a natural pandemic, there was one after World War I that actually killed more people and in a much shorter time than World War I itself.

34:53 S1: Or I could guess at a synthetic pandemic as the state of biology advances and you see the risk that people are able to craft their own pandemics. That would be very high on the scariness scale. There is a big government infrastructure around preventing pandemics, but I think there is a lack of philanthropy and I think philanthropy can have a special role to play here, especially around looking at far out worst cases.

35:22 S1: I think a lot of the government work is around preventing tragic, but not civilization threatening pandemics and in terms of trying to think through what the very worst case could look like and whether the preparations for that look any different. I think there is a lack of attention there. So we recently, we basically spent an entire year trying to hire someone for this. And we recently did hire Jaime Assif who's now gonna be working with Howie Lempel. The two of them, I think this is a big complicated field, and the two of them are both gonna be working full time on this. This has only been going on for six or seven weeks, so we expect in the next few months they'll have a bunch of grants to make the world more prepared for the next pandemic.

36:04 S1: And then the other cause is potential risk for advanced artificial intelligence, which is probably an example for many people of something that sounds way too wacky to actually work on. It's kind of a hotly debated issue these days. The question is there's been a lot of exciting progress in machine learning and artificial intelligence research in the last five years or so, and the question is whether we're getting anywhere close to a day when you can imagine AIs or machine learning systems that are really intelligent in a broad array of domains that are able to process information and make super human decisions, not just in board games, not just in math, but in things like making predictions about how society will evolve and what kinds of actions one can take to achieve a desired result. That could speed up science a lot.

36:53 S1: That could lead to very powerful technologies coming a lot quicker than anyone is expecting. And some of the concerns here, I think there's a lot of reasons to expect this could be a very good thing, but some of the concerns, one of them is if you have... Arguably, human intelligence is the thing that has changed the world the most to date and if you had something that was more powerful than human intelligence at finding technologies, finding solutions, finding

strategies, that is a top candidate for something that would possibly really change civilization and give a great deal of power to whoever has access to that technology first. And so we have two categories of risk. One of them is misuse risk and this is what happens if you have a very powerful artificial intelligence that through hacking or through government programs becomes a tool of, let's say, an authoritarian government or someone who has bad intentions.

37:47 S1: That's one of the worst things I can imagine happening. It's very speculative, the technology's not there yet, but we won't necessarily get a big warning shot when the technology comes. It's not like medicine where you have 10 years of randomized controlled trials while you're figuring out whether you can put a drug on the market. These things can make big leaps in capabilities and I don't think anyone has much of a road map in front of them. Then, the other other kind of risk is accident risk and so the question is what if you have an AI system that in many respects is better at accomplishing goals than any human is, but that also has some kind of fatal bug, that is missing some crucial piece of the world that didn't have the correctly specified value function that it's optimizing for. In some ways, one of the worst cases I can imagine for human civilization is that you have a superhuman intelligence that has a non-human friendly goal, whether due to accidents, bad programming, bad user, whatever.

38:42 S1: So I know a lot of this probably sounds very speculative and sci-fiey. I think in some ways it is, but I have become convinced over the last year or so that there is a reasonable chance that this is not a far off, never to be occurring event. And I think that would be a long conversation and we wrote a very long blog post about it that still only scratches the surface, but I do think... The thing that I'm saying is I believe in non-trivial probability, so at least 10% and definitely over 1% that within the next 20 years, we'll see the kind of artificial intelligence that has really superhuman abilities in a large number of very important domains and can fundamentally transform the world because of it. And once you believe that this could be a 20-year issue, as a philanthropist that kind of means, act now because the time scales we work on trying to build fields, encourage people to start organizations, help organizations grow to the point where there can later be impact, I think is a major issue.

39:49 S1: Potential risk for advanced AI is a very challenging cause for us. It's not like criminal justice where we could hire one person who kind of knows everything that's going on. It's a field that a lot of people don't even really believe is or should be a field. The field of technical research to prevent the kind of bugs I'm talking about is not really big in mainstream academia at this time. Hopefully that can and will change. So we haven't really been able to find one person that we think can do it all. We think it's a time intensive cause at this time to lay the groundwork for ourselves to make better grants in the future. And so we have multiple staff working on it this year and I am personally spending a fair amount of time on it. And one of the reasons I'm in New York is because when we work on an issue, we think about something a lot in the abstract and I like opportunities to see the work we're doing up close like a site visit. And so I'm here at ICML, which is one of the premier machine learning conferences and that's a chance to just see the whole AI research community in one place, see them all interacting, see them presenting research to each other, talking and get to know just how they're thinking about these issues, something I already know a fair amount because we've done this less systematically.

41:00 S1: That is something we're spending a decent amount of time on this year and we've kind of set it up as a large priority, partly because I think this cause really does score very high on importance, neglected-ness and tractability, and also because I just think the challenges of it are such that it needs more time than other causes at this stage in the game and hopefully that changes later. So, those are some of the things we're doing. There's a lot of other stuff going on Open Phil too. We recently brought on three science advisors, who are going to try and help us do this same important neglected tractable exercise for fields of scientific research, which is very challenging, it could take years. We have a history of philanthropy project, trying to learn from what philanthropists have done well in the past. We recently put out a couple of case studies on that, but at this point I'm going to take questions for another 10 or 15 minutes. Questions? Yeah.

41:54 S?: So for me, this kind of [41:58] ____ with my question from before. So where GiveWell may seem like it's a little bit too targeted to some people. This can possibly be seen as going the complete opposite end of the direction.

42:12 S1: Arguably.

[chuckle]

42:13 S?: Is there more of an opportunity to go almost that far? Maybe you can back it up a little bit?

42:20 S1: Yeah. One question would be is there something in between GiveWell and Open Phil? In some ways, there

are. The GiveWell experimental work is looking for new feature top charities, It's more speculative and high risk than GiveWell. It's less controversial than some of the Open Phil stuff. I think Open Phil has quite a range in terms of controversiality awareness, I think pandemic preparedness is a much more mainstream and normal idea than potential risk from advanced AI. So I think there is a range we're covering, I will say in general that I have an easier time seeing the case for doing something really rigorous and self skeptical and totally evidence based.

43:00 S1: And I have an easier time seeing the case for doing something where you're really just going all out for the biggest home run you could possibly hit. And I see a less compelling case at this stage in my intellectual development for this middle ground where you're taking risks, but also holding yourself back. So it's not as exciting an idea to me, but I think there is a span within GiveWell and Open Phil and there's also plenty of opportunities for other people to try and find their middle ground themselves.

43:29 S1: I think in some ways one of the best public goods we can provide is exploring what the extremes look like and giving demonstrations of how to talk about giving publicly when you're being very evidence based. How to talk about it publicly when you're being more out there. I think it makes the middle ground easier to explore, but we are most interested in the two ends of the spectrum. Yep?

43:51 S3: So this is a little philosophical, but how do you guys value utility across the spectrum between GiveWell and Open Phil? I know you guys have looked at that like adjusting light years, but have you also taken into account psychology studies and happiness in general? Just social types of connectivity in your decisions?

44:12 S1: Yeah. How do we trade off all the different... You look at these different causes and it's like some of them are trying to stop human civilization from going off course, some of them are trying to reduce chicken suffering, some of them are trying to reduce malaria. How do you compare all that? The most general answer is that at ActiveWell and Open Phil, we generally try to cultivate an attitude of being very deeply engaged with these tough philosophical questions, but also very pragmatic and not letting them hold us back or stop us or dominate our work.

44:43 S1: So we don't want to be in a position where we can't do anything until we decide exactly what we think about philosophy of consciousness, even though philosophy of consciousness matters a lot. And we don't want to be in a position where everything we're doing only makes sense if you endorse one particular philosophical view. And so we try to strut the middle ground and what it often means is taking our best guess, diversifying a bit across different defensible world views. So in other words, I think there's an argument that animal suffering is the most promising direction because it depends on your philosophical views about animals and other things.

45:18 S1: There's an argument that this global, catastrophic risk prevention is the most promising direction. As long as I feel legitimately uncertain, and feel like I could change my mind and take on any of these pretty soon, I think there is a case on multiple dimensions for doing some of each, especially given the amount of resources that we're recommending at this point, that we can hit diminishing returns. We can be one of the biggest funders basically in all the fields I just named at once and that seems better than just picking one and trying to put everything in there.

45:47 S1: So we try to balance them. In terms of the cost effectiveness calculations, we calculate cost effectiveness usually either in terms of lives or life years or in terms of economic value created in terms of dollars, but we also will entertain arguments that under standard models of utility, money is worth more when you have less of it, and so in other words, you can argue for example that give directly is 100 X return because you're taking money from a country where the average income is something like $30,000 to a community where the income of the recipients is something like $300 a year.

46:29 S1: And so you're taking money from one place that has 100 X as much of it and on some models that would be 100 X return. And so those are some of the main assumptions and views and metrics that we use to compare things, but a lot of times when things are incommensurable, we'll just say, "As long as either one looks like it could really be the best, we'd really be happy to do both." And I think we are interested in getting more rigorous about how to make these comparisons and that might be a future project, but it's not the main thing we're working on right now.

[background conversation]

47:04 S?: And this is kind of related to the... What's that? Either organization, how do you measure your impact [47:10] ____, saying, that it can be ambiguous?

47:16 S1: Sure. How do we measure our impact? One measure of impact is money moved, so it's how many dollars did we cause to be given to things based on our recommendation? That measure is pretty easy. I think the question then becomes, "How good is our recommendation, and how much more good is the money doing under our recommendation than otherwise?" That's the much harder question to answer. I will say we've made some attempt to answer it and I would mostly say it's like if you look at the money moved and you look at our operating expenses, and you believe that we're making the giving two Xs effective or more, it comes out looking very good. And I think there's good reasons to believe we are doing that, for example variation even between our top charities, looks like arguably 10X.

48:08 S1: But, another thing to realize is that, under Open Phil, I believe that philanthropy is one of the areas where you're going to gain the most by being non-insistent on always being able to get the feedback loops, and get the metrics. I think one of the things that distinguishes philanthropy is that you are able to fund things where the only argument is that this seems like a really good idea, and it may not pay off for 20 years, it may pay off with only 10% probability. You may never really see the results, and you're not gonna have a profit, bottom line. I think that is the comparative advantage of philanthropy, and I think it'd be a mistake to leave it on the table.

48:44 S1: So, our general attitude is to evaluate ourselves wherever evaluation seems feasible, but not cut off work just because it can't be evaluated. And so there is a great deal of Open Phil work where we say, "Honestly, this grant has a 10% chance of being great over the next 10 years. We're not gonna know anything about it any time soon." But we can review our whole portfolio every five years and say, "Did we get some big wins? And if not, we've got a problem." And at least that is worth something. And we can take the big wins, and we can vet one or two of them with an actual historian, which is another thing we're currently doing for other big wins with the history of philanthropy project. Yeah, Colin?

49:22 S?: Yeah, I was asking [49:22] ____ can see the potential [49:28] ____. What can we actually do now? Can you talk more about that?

49:32 S1: Sure. What can we actually do about potential risk for an advanced AI? So, there's the two classes of risk, misuse and accidents. So, first I'll talk about accidents, this idea that you could mis-specify a program, or that you could have a bug in it that causes it to be a really scary combination of intelligence and stupidity, in a sense. Or intelligence and nonhuman-friendly values. There, I think, there are a series of technical questions. We're already seeing some of the opportunities to have this kind of screw up in today's promising machine learning system. So for example, a reinforcement learner is a kind of system that learns from reward. And so, it does something, maybe it's playing a video game, it presses a button, it looks at its score, it presses a button, it looks at its score. And without having that score, it's not able to have any way to direct its experiments and learn. And so, it's kind of a popular framework, to have a machine learning system learn that way.

50:33 S1: The problem is, once the machine learning system gets broad enough, and instead of just looking at the video game, it's looking at the whole world, it may notice itself and its own code and its own reward channel. And it may realize that the best way to get a high score is to hack into itself, and set its score really high. Which could actually be a bigger problem if that also involves making sure that no human ever shuts it off, and then it gets to run as long as possible, and maybe as fast as possible. So, that's an example of something that I don't think anyone has a totally satisfying, easy solution to how you get a reinforcement learner to not hack itself. But there are a bunch of ideas. I think the ideas need to be worked out. There are some concrete technical challenges. There will be, basically in the next couple of days, there will be a paper that I think is probably better than stuff that's been out there before that sheds light on a lot of this. So, hopefully we'll cover it in our blog within the next couple of weeks. But it will have a nice, very concrete, technical description of some of the challenges.

51:35 S1: Now, what about misuse? How do you stop this from getting into the wrong hands when it doesn't even exist yet? I think basically there are a series of potential policy and regulatory challenges with a general artificial intelligence, or just a broad scope artificial intelligence, that I don't think, by default, would be well handled by current legal and regulatory frameworks. You have potentially a system that is simultaneously a medical research tool and a weapon. Things like that. And figuring out at what point you need to start regulating it, and how you should regulate it, and who should regulate it. And what international agreements should exist on who gets to control and deploy such a thing, I think is basically something that no one is thinking about. We would like people to start thinking about it, I think that would be productive.

52:26 S1: And I think, even though these things could be a long way off, starting to analyze the potential scenarios and starting to have... Realize, what are some of the things we could actually do, and what would be some of the non-

obvious pitfalls and some of the non-obvious options, that's something I'd like to see happening. So, that's also work we'd like to fund. And that could be academic think tanks, and things of that nature. I'll take one or two more. Yeah?

52:51 S?: It's just a quick comment on how you evaluate yourself. I think that, at least in the cases that I'm aware, you should also consider that you might be increasing the amount of money people donate as they read your research and become comfortable donating more, because they know what the effect is. But I really wanna ask about Open Phil. I have no idea how Open Phil is funded, and I imagine it'd be really hard to get GiveWell donors, generally, to fund it, even though it is exciting. Just because the alternative, like you said, is low hanging fruit. But have you considered that people who currently donate significantly to GiveWell could be useful beyond their donations? Just as people who may work in a related field, or have a relationship with their Congressman, or something like that.

53:39 S1: Sure. So, how do we fund Open Phil, and have we thought about how donors could help in ways other than giving? Open Phil currently is just part of GiveWell, and it's funded out of the GiveWell general budget. That's something we want to change. We anticipate that it's gonna be funded when it separates by Good Ventures, which is our main partner for the time being, and we think that'll be probably the simplest way for things to be. Because that is where a lot of the grant making is coming from and that kind of makes the most sense. In terms of donors helping us out, anyone who has ideas for how they might be able to help us out, definitely get in touch. Usually we look to our program officers to lead the way on figuring out what needs to be done and who are the best people to contact to do it. I'll take one more question then I'll hang around and talk for people who wanna talk. Who's got the last question? Yeah.

54:28 S?: Have you thought about, as a catastrophe scenarios, where the sun gets blocked out and you need to think of new ways to make food in a hurry, which is something that Dave Denkenberger from GCRI kind of proposed.

54:45 S1: Sure. Have we thought about scenarios where agriculture is severely affected and you need quick ways and unusual ways to make food? I think this is an interesting idea. I think it cuts across a few potential global catastrophic risks including some of the worst possible effects of super volcanoes, nuclear war and climate change. It's something we would consider. I think we would need to get more into the science to be really excited about it and it's not super high in the priority list because I think if I'm naming the two things that could really knock civilization off course, they'll look more like pandemics and AI in my opinion. That could be a long conversation I'd be happy to have with people afterward. But I think the set of risks that I would associate with primarily operating through agriculture and the food supply I think is to me is adding up to a smaller risk in terms of that massive disruption.

55:39 S1: Now I think, climate change is a huge global issue and I think one of the most important issues in the world and something that I really wish more were being done about. But I think in terms of that knock civilization off course probability I think that's on a longer timeframe with a lower probability than the other two things I said, though still pretty scary in the scheme of things.

56:00 S1: So, with that said, thanks everyone for coming. I really appreciate all the questions and more so appreciate all the support that you folks provide to our top charities and to us. You make our work possible and we really appreciate your sort of openness to this being your charity event. It's very different from most charity events and we love that. So, thanks everyone and I will stick around a little bit to chat.

[applause]