GiveWell NYC Research Event May 22, 2017 – 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 Holden Karnofsky: Recording. If anyone doesn't want to have the things you say be on the recording that we publish, then just let me know after or let me know when you speak up. So yeah, I'm Holden. I co-founded GiveWell. I'm here to talk about, though, the Open Philanthropy project which I also co-founded and now serve as the Executive Director of. And the first update on the Open Philanthropy project which I'll start with, is that we are going to be a separate organization actually soon. So, we've been saying this for a while, that we're working on separating them, turns out that having this spin off happen the way we needed it to happen is pretty complicated, it took a while. The target now is two weeks. It really shouldn't be that long, and I think that gets to the fact that Open Phil and GiveWell have become completely different things. I think very much sharing the same spirit, the same goals, a lot of the same philosophy but to understand one you might want to delete a lot of what you were thinking about the other from your memory.

00:57 HK: So I'm going to talk about how we came to be different, where we came from, what we do at a high level, and then I'll take questions for a while, and then I may get a little bit more into the details of the causes we work on and some of the stuff that we're funding.

01:12 HK: So first off just who we are, where we came from. Back in 2012, GiveWell, Elie and I met Cari and Dustin, Cari Tuna and Dustin Moskovitz. Dustin is one of the co-founders of Facebook, and they were looking to give away a very large fortune in their lifetimes and we kind of started off by saying, "We would love to help you do that as well as possible and create a resource to help people do that as well as possible," just like we did to help the kinds of donors Elie and I were in 2007 which is... But what we realized quickly was that, maybe not as quickly as we should have, that they're very different kinds of operations, that one kind of product GiveWell is aimed at the sort of person who says, "I want to give away money, I want to do it now, I don't have a ton of time to think about it, I could be giving $100, I could be giving $1 million, but it's not what I'm doing with my life, it's something that I want to do now and so I'd like something reliable, preferably linear, something that doesn't take me a huge amount of time to get up to speed on and that, where the case can be demonstrated, written up, established."

02:21 HK: And so GiveWell has these criteria of looking for things that are proven, cost-effective, and scalable. Whereas Open Phil is more targeted at the kind of person who's saying, "I want to give away very large amounts of money, like billions of dollars, have my whole life in which to do it, can hire my own staff, can build my own trust relationships." And I think where that ends up pointing you certainly when you're in that position, you still should consider GiveWell top charities to be outstanding giving opportunities. But you also should consider a bunch of things that the first kind of donor really couldn't have thought about at all. And a lot especially of what that is, as a major donor you can create new organizations, you're not stuck with the ones that exist. You can change the priorities of existing organizations and you can also just build up a lot more context and a lot more of your network and your own trust and your own approach.

03:11 HK: And I think that's important because what we've gotten pretty into as one of the philosophies we follow, and it's not the only formula for good giving, but we have this idea we call "Hits-based Giving", which is this idea that if you try as a philanthropist, you take 10 big bets, and nine of them fail miserably and even comically, and one of them goes really, really well, that in aggregate could be a really good return and that's similar reasoning to what a lot of venture capitalists use. For example, that they might invest in 10 companies, nine of them might totally fail, one of them becomes enormous and overall you do very well. And I think historically, philanthropy has big enough successes to its credit that it's at least plausible that you can get that kind of win that makes up for a lot of losses. And coming into philanthropy with that attitude, I think, requires a big frame shift from the GiveWell mentality.

04:04 HK: Because if what you want to do is swing for the fences, in a sense, go for things that would be really huge if you got them right, such that it's okay to get several failures. You're going to want to take a different approach and I think among the things you want to do differently, one of them is: Have less of a requirement for evidence. Have less of a requirement for organizations that already exist, have more of a willingness to form your own view of a situation, form your own network of trusted people and do things that makes sense to you and make sense to the people you trust and really don't make sense to other people whose case really can't be made quickly, that you really can't create an easy compelling write up on the internet for.

04:48 HK: And so I think that ends up being in some ways the opposite end of the spectrum from GiveWell in terms of

the mentality we bring. It's the same exact mission, there's some money, we want to do as much good as possible with it. But it's a different mindset and it comes down to saying, how can we aim really high and how can we do things that are special? And I think that still requires a great deal of empirical research of self criticism, of self skepticism but it does take a different form. The thing that we've... What I'm going to walk through right now is just the structural set up of Open Phil, just how we've designed the whole thing to decide what to give to, what kind of staff to build or what kinds of activities to do, and then I'll probably pause for a while and take questions, and then maybe we'll talk a little bit more about some of the specifics.

05:35 HK: But Open Phil was, for several years, was really just trying to find our way, hire the right staff, build the right frameworks and we worked in a place where we weren't really giving it scale yet, so we didn't really know what Open Phil would look like yet. And I would say this year is the first time that we have a really good picture of what... What it looks like when Open Phil is giving away a lot of money.

05:58 HK: 2016 giving was over $100 million, and that does not include any of the money that was given to GiveWell top charities by Cari and Dustin, which we also recommended, and you know that number will be equally high, or higher probably for 2017 and so you know now we're in a stage where we're kind of stepping back and saying, "We know we have a method we've used to choose our causes, to choose our people, to choose how we do things, and now we're trying to step back and examine a lot of assumptions we made along the way that may be needed a little bit more attention, build a stronger association to get the whole thing to be more robust."

06:38 HK: So, the story how we got there and what we've done so far, the choices we've made, we started off by looking for areas, causes, that are important, neglected, and tractable, so that is a very different set of criteria from proven, cost effective, and scalable. So we look for basically fields or areas or problems, and we look at a whole lot of them and we say, "How important is this area?" in that how many people would it affect and how much if we got a really big wing in this area. How neglected is it? So who else works on it, and then how tractable is it? So, when we look at what there is left for us to do how promising does it look. And we did a lot of really intensive investigation looking at a lot of different areas we might work in rating them on these criteria.

07:24 HK: Instead of running through all of that, instead of running through how we did the ratings, I'm going to group them into two broad themes to give a sense of where we ended up, what causes we ended up thinking are the most promising. And we did start very deliberately, by picking areas to work in, causes to work on so that we could build expertise around them. So, we didn't start by looking for organizations, we didn't start by looking for grants; we started by looking for things to become more knowledgeable about in the sense and to build more staff for.

07:55 HK: And so two big themes that most of our work fits into in one way or another, and this is like how it turned out not how we originally set it up. The theme one is what I call radical empathy, we have a blog post on this, and the basic idea is that there's been sort of a history of people expanding the moral circle of concern. So, it used to be that, if someone wasn't from your city state or your tribe they didn't have human rights from your perspective, you didn't care what happened to them, may be not you specifically but people. And as time has gone on, I feel like, in many ways the set of people who are considered to have rights and who are considered to be fully equal with everyone else has grown, and now when we look back, and we said, "What do we wish we were doing if we were trying to do a lot of good in the past."

08:49 HK: I think a lot of what we wish is that we've been early on some of these expansions. So, early to feminism, early to civil rights would have been really good. And so a lot of what the causes we found promising are kind of finding populations that are marginalized in the sense that most philanthropists don't really pay much attention to them, don't really prioritize them, may be don't consider them to be persons, or to have rights at all. An example of that is the global poor I think get much less attention from rich countries than the American poor for example. And so that's a logic for thinking that GiveWell top charities are excellent and that they're good place to give.

09:30 HK: Another example would be our work on criminal justice reform where a lot of people when they debate criminal justice and what our system should look like the whole debate is about public safety and what's going to make us safer. And when you consider the considerable human cost of that system that we have I think things look quite different, and I think decarceration looks like a very good idea.

09:52 HK: And then the final theme in this category is farm animal welfare. So, this is... Basically there is enormous number of animals being treated with incredible cruelty on factory farms, and if you're going to look for something where you might be able to do an outsized amount of good where you might be able to make a huge difference

considering whether these animals might be appropriate objects of empathy and moral concern is a very interesting place to go and thinking about when we look back a hundred years from now will we wish those were the creatures we were trying to help. And furthermore we've found that cause quite tractable. I think it's very neglected and we believe we've already seen a lot of wins on that cause.

10:33 HK: So, we believe we've already funded a lot of work that has tangibly gone after fast food companies, grocers shamed them into taking pledges to treat animals better and get, sometimes minor welfare improvements but for a very large number of animals and with incredible speed, such that if you did decide that you value this as much as humans, you would think this is really a good thing to do in some sense.

10:57 HK: And then the other big category for us is global catastrophic risks. So, this is a different way of thinking about things, but if you're trying to do an outsized amount of good, if you're trying for something that might really... We might look back and say, "That was the best thing you could've done." Another major factor in how much good you do, is how you contribute to the overall trajectory of civilization, and I know that's a very dramatic way of putting it, but if you'd gone back three hundred years you... I think the industrial revolution arguably did more to reduce poverty than anything you could have done in 1700's that look like alms or charity. And so if you can think of ways in which either things might get a lot dramatically better in that way, or giving the trajectory we're on, it's more interesting to talk about ways things might get dramatically worse.

[laughter]

11:51 HK: If you can stop some of those bad things from happening or make them less likely, even a small contribution to something that could really derail society, if you think about how that contribution ripples throughout the rest of future generations, that's another way in which if you do well, you could do really really well. So climate change is the obvious example of this and we have funded work on climate change. Other less salient examples perhaps and that have become focus areas for us because we think they're more neglected, certainly pandemics. So if you gave me a crystal ball and said humanity gets wiped down in the next 100 years, and you made me guess how? I would probably guess something with a pandemic and we believe the world is under-prepared for these, we especially believe that the pandemic preparedness community under emphasizes the possibility of really really big pandemics that can catch on, that can kind of reshape the way the world works for the worst.

12:46 HK: And then, the other cause in this category for us that's a major focus area, is potential risk of advanced artificial intelligence, which is a whole can of worms that we'll get into in the next part of the session. But the short story is that we think this is a... Artificial intelligence is a surprisingly dynamic part of science right now, it is advancing, it is hard to tell where it's going to go and when it's going to go, and we do believe that most of what has happened to kind of Earth so far in a sense has been driven by human intelligence. And if there were... The more that we build computers that are able to outdo humans at certain intellectual task, it becomes very unpredictable what the consequences of that would be, trying to get out ahead of it is something we're very interested in. So once we pick these causes, what do we do? And our basic MO is one, work really hard at hiring the right people. So what we're trying to do with these causes is not just bring our own lens and do what seems right to us but what we're trying to do is learn enough about them, that we can find true experts who will then lead our work.

13:49 HK: Once we hire people who are, the usual title is programme officer, and those are the people who lead our grant making in an area, it becomes their job to understand everything about the cause obsessively. So, Chloe Cockburn who works on criminal justice reform, it's her job to know everyone who works on criminal justice reform who's we're talking to. Think about what ought to be done here from all the potential grantees, be an expert on it and then communicate with us. And then from there... So, she leads the way on strategy, she leads the way on what we consider and then from there we just asked her a ton of questions. So we have this internal process where there's a internal write up that comes through and it's focused on a lot of things, like for example, "Can you give a rough quantitative model of how much good you're expecting to do here that can help us understand which parts of what you're doing are the parts you're most excited about." And that's kind of a similar thing to the GiveWell cost effectiveness analysis, but usually a lot rougher.

14:46 HK: And then another thing we look for a lot is forecasting, so can you make predictions, can you look at your grant and say, "If this goes well, this will happen, this is what I think will happen with this probability," so that we can look back and assess ourselves. And then once those things look pretty good to us, we will make the grant, there's a general guideline called the 50-40-10 guideline, that if we have a programme officer, we want 50% of their portfolio. 50% of the grants we make to be things that we the decision makers, Cari and I, feel really convinced of that we feel we

could personally defend, that we think that are good ideas. And then it's okay if another 40%, bring us up to 90% total, is just things that I wouldn't quite defend myself, but I could at least see how I could think they were a good idea, if I knew more. And then the remaining 10% is deferral, is trust and our programme officers can make some grants where they don't even have to argue them to us, they just say, a certain percentage of my portfolio, you're trusting me as long as there's no major risk or downside here we can just go ahead and make it.

15:51 HK: And so that is how we're set up, and the overarching philosophy here is that we are not trying to centralize all the knowledge in one place. And that makes Open Phil very different from GiveWell and much harder to talk about and much harder to summarize, 'cause GiveWell in the end, it has these recommendations, and everything that you need to understand the recommendations is on the web. You might not have time to read it, but it's all in one place.

[chuckle]

16:14 HK: Open Phil, I have my spot checks and I have my trusted advisors and I have my background beliefs, but there's no one person who actually could explain every grant that Open Phil has made. A lot of questions people ask I'll say, "Yes, we could get into that, but also there's other people at the organization who knew much more than I do, so it can get very hard to summarize. We work on a lot of different areas, we have about 20 people on staff now, but I'm going to do my best and take questions about whatever people are curious about. Yep.

16:47 Speaker 2: You said [16:47] ____ that philanthropy has big enough successes in it's past that you think you can justify this hits-based model. So can you give me three really big hits with philanthropy in the past that you would love to [17:01] ____.

17:01 HK: Yup. Yeah, three really big hits for philanthropy in the past. So, I'll name the three in the blog post we wrote on this, although one of them is not quite what you're asking for, but I think it'll be good enough. So, one of them is the Green Revolution, so the Rockefeller foundation funded Norman Borlaug to do research on seeds with improved productivity, agriculture, wheat that would be resistant to pest and things like that. And originally they did it in Mexico, later did it in India; they went from being in the middle of a famine to be an exporter of wheat. And generally this development of seeds kicked off a massive period of economic growth growing agriculture productivity that has been credited fairly credibly in my view with saving over a billion people from starvation with potentially making a major contribution to kicking off the East Asian tiger, like miracle, where several countries went from poor to developed. So that is a big win.

18:00 HK: And then, another one that I'll give you that I think is pretty massive is the pill. So, there was Margaret Sanger, a feminist, thought it would be good to have a pill that people could just take if they didn't want to have children and convinced another feminist philanthropist, Katharine McCormick, to fund this person who'd been doing trials in rabbits and figuring out what chemicals could manipulate their cycle. They wanted to try it in humans, and that I think is another... It's one of these things that it wasn’t going to be government-funded. It wasn't something that everyone in society wanted. It was controversial. They actually, I believe, were not allowed to advertise it for birth control. They had to get it approved for this other thing, but then they put a warning on it, that you couldn't have children while you're taking it.

[laughter]

18:50 HK: I think they might have even had to put the warning on there. And so it was not a popular thing. It wasn't the kind of thing where everyone was on the same page. It really, I think, was philanthropy taking the lead, and I think made a huge contribution to the world and to that person's goals.

19:05 HK: And then, another example is Steve Teles has written a book on the conservative legal movement. So this is a case for... I might not agree with the goals, but I think you can look at this and say this is a big win for the person who did this. And this is philanthropist's funding of a huge amount of intellectual work and college associations and graduate schools and just everything they could to get a certain way of thinking about policy, particular conservative attitude to be more intellectually respectable, to give it a home outside academia, where they felt there was hostility toward it. And it's hard to say exactly what that changed that it's discussed at great length in this book. But I think it did lead to the rise of a number of important conservative think tanks, conservative intellectuals, conservative ideas that have really dominated American discourse and really changed the way we make policy, for better or for worse. Those are examples and if you think you might hit it that big, I think you should give it a shot, at least for your values. Yep.

20:07 Speaker 3: What is the biggest failure that you funded so far?

20:12 HK: The biggest failure we funded so far. So a lot of things we do are not... There aren't long-enough timelines, that it's often not that interesting to talk about failures yet. Giving on a high level is very new to us. So there's nothing I can point to yet, where I'd say, "We sank a ton of money into that and failed," because most of things we put a ton of money into have not been around that long. We have failures already and we have things that are pretty concrete. An example I'll give you that I think will satisfy you is there was... I think the Howard G Buffett Foundation funded this experiment where they were matching farm workers with farm visas. And the idea is that for every person who comes from a poor country into the US to work on a farm visa to get a huge boost in income, and so you can get a massive return. And these farm visas were not capped, and so there's no limit to them.

21:07 HK: And we basically tried to repeat this experiment, to see if it was a one-off or to see if it'll work again. We did it specifically with Haiti, specifically with Haitians. I don't remember exactly what happened but I think they weren't able to make the matches between the employers and the people seeking the visas. And in a lot of cases, they weren't able to get the US government to make the approvals they needed to make and to play along with it. But I think now when I hear that story, there was a little bit of a media cycle about how this was an amazing return on investment. And when I hear that, I'm kind of like, "Yeah, it was, but I don't know how repeatable it really is". Yep.

21:52 Speaker 4: You had mentioned you have... Your researchers make predictions about how things are going to turn out. Is that publicly available or is that private?

22:01 HK: Are the predictions that researchers make publicly available or private? Probably most of them are private but some of them are public. So if you look at grant pages, there will be predictions on there. A lot of times, it's like a minority of the predictions. Definitely, Open Philanthropy has moved in the direction of sharing information that we think will be interesting, that we think will help people understand what we're doing and learn about philanthropy but not having a default to share everything. Because we just don't think we're a desirable partner when we do the latter and we don't think the benefits justify it. So, Open Philanthropy, definitely a smaller percentage of our thinking is online and public. And Givewell, there's a lot of stuff we just won't talk about because we have grantees and partners who don't want us to. And a lot of predictions just wouldn't be appropriate but there are some that are public. Yep.

22:53 Speaker 5: A lot of your themes, within your themes, your cause areas are really connected, so low poverty and climate change, farm animals and climate change, farm animals and pandemics. How do you sort of tease out the connections between those cause areas? Do you think of them as separate or are they looked at as integrated?

23:10 HK: Yeah, how do we tease out the connections between cause areas? So, first comment is that when we're choosing focus areas, focus area is a very vague term and it could mean narrow things and it could mean broad things. So a focus area, it could be climate change, it could be geo-engineering, it could be environment, it could be supporting great leaders who are unconventional in way X no matter what cause they work on. You can come up with anything, call it a focus area, and then figure out if it's important, neglected and tractable. So, anything people think we should be doing that we're not, it's not the focus area idea's fault, it's just our fault for not looking at that area and thinking about it.

23:46 HK: So, in terms of the connections between them, I would say it's just a huge part of our philosophy is it's just every cause has its point person and that person is really the go-to person, and my job is to continually assess them, to continually understand them, to spot check, to have a view of their strengths and weaknesses, to know who we might be missing, to think about if we have the best people and the best causes. But they're really in charge of driving the strategy. And so, if there is a connection, for example, between farm animal welfare and climate change, that's going to be up to the people who work on those two causes to get together and talk. And that's really all there is to it, 'cause those focus areas were chosen to be the best we could make them.

24:29 S5: Are there attempts within the staff? Is there a lot of cross-pollination between those program officers?

24:34 HK: Is there a lot of cross-pollination? I would say mostly cross-pollination right now is methodological, so our program officers meet occasionally. I don't actually know how often, 'cause I'm not there. [chuckle] But they'll talk about things like, "How do I give feedback to a grantee about something I wish they were doing differently without turning this into me pushing them around with unhealthy power dynamic?" or, "How do I look back on my grants and determine what went well and what went poorly?" Or like, a huge topic of discussion with us had been, "How do I build a field?" So, similar to what I said about building the conservative legal movement, there's a lot of places where we say, "Here's an important problem, we wish it was a popular topic for top academics to debate all day, but it's not. What do

we do?" So we have a whole case study on field building that's public, and we talk a lot with each other about, "All right, should I do a fellowship now? Should I fund professors?" So we try and share knowledge like that. That's probably the most of that, we do. Yep.

25:29 Speaker 6: Thanks, Holden. Are there any focus areas that you wish that you guys can be focusing on and haven't yet just because of capacity? Do you want other people in the EA movement to focus on?

25:39 HK: Sure. Are there focus areas that we don't have but we... That I would love other people to do? Yeah, there's a really long list of those, so I could just go on for a while about... And some of them are not... They're not all weird, different, clever ones, like, "Wish more people would work in climate change." It's not the highest priority for us, but it's something we have done some funding on, and it's another one that I think if someone were to go into philanthropy, I think that would be a good one. But if you go to our process page under research and ideas, there's a couple bullet points that go to spreadsheets. They list a lot of the things we looked into. And most of the things we looked into I think are actually pretty good causes, 'cause they had to clear a certain threshold of plausibility. There's a lot of stuff out there and we are going to update, 'cause there are certain causes too that we're just spending less time on them than we used to, and we'd love someone else to pick up the ball because we've re-prioritized.

26:30 HK: For example, we're not working as much on land use reform and macroeconomic stabilization policy as we were. We're going to be writing about why that is, and we have already written some about why it is, but those are good causes, still. Yeah, Jacy?

26:42 Jacy: So you talked about the morals circle kind of pushing beyond that is an important thing for philanthropy. And for farm animal work, which I'm familiar with, that's a lot of what you're doing, [26:52] ____ concern for those animals. But have you thought about doing the same thing for the global poor? So any sort of, maybe an immigration thing that you think would help people in high-income countries think more about the global poor and anything like those?

27:05 HK: Sure. Have we thought about trying to raise awareness of the global poor and broaden the circle for them? Right now, Open Philanthropy, our work on global health and development really just comes down to recommending continued support of GiveWell and GiveWell's top charities, and GiveWell incubation grants. So that is... We have our hands full, and we don't have special staff for... There's other things we could do in global health. We could do things that are structured differently that aren't fit for GiveWell, like what you're talking about, that'd be one example. But we do feel we have our hands full, and we also feel that the GiveWell stuff is really good on that axis. If your goal is to help the global poor... I've looked at and thought about and discussed a heck of a lot of ideas, and there's some stuff that could be better than GiveWell top charities, but nothing that clearly is. And so we feel pretty good there with just... Yeah, we have other priorities at the moment, and we feel that we have really good things to support on that front. Yep.

28:02 Speaker 8: Have you guys already started working on any programs related to AI or you'll be starting [28:05] ____?

28:07 HK: Yeah, we're very busy with AI. Yeah.

28:08 S8: Can you just go through a couple?

28:10 HK: Sure. So the work we're doing on AI... To even characterize the problem would take me a while, but I'll kind of see what I can say about what we're doing. So one major thing, and the thing we've spent most of our time on, is the idea of trying to build a field around academic technical work on AI safety. And so the basic concept here, is that if AI advances in a very dramatic fashion, which... "Could it advance in that fashion? When will it?" I'll leave those questions aside for now. But if AI advances in a very dramatic fashion, you might end up with some sort of principle agent problems or coding challenges. An example of how this might work is current AI paradigms really need well-defined rewards and goals to learn from and optimize around. And the problem is that most things that we humans want most are not very well-defined.

29:05 HK: And so an example of a bad situation would be something where you have incredibly powerful, and intelligent in a sense algorithm. You give it a goal, it figures out how to get the goal. Anyone can give it a goal, like, "Maximize the amount of money in this bank account." But that would be a really bad goal. You don't know how it's going to do that. It would take you very literally, because the goal's been specified very definitively. And if you want to say, "Please protect us from other jerks?" using AIs to do that, that's a poorly-defined goal. And so that's a situation

where you could be in bad shape where you have well-defined goals that would actually be really bad for AIs to pursue, and poorly-defined goals that would be good for them to pursue.

29:47 HK: So we want to fund a field of technical research that works on problems such as A, "How do we get AIs to learn fuzzy human goals from humans?" Now, example of this is a paper coming out soon, not funded by us, but good one. The normal way for an AI to learn how to play a video game, like an Atari game, is it plays to the end, it watches it's score, the scores is very well defined and it learns how to play. And an alternative way that's being experimented with is, it plays for a while, it samples itself, it shows a human two different video clips of itself playing the game and the human goes, "That one's better." And then you do it over and over again and so you're trying to get it to learn what the human wants instead of just using this metric. So one of the things they tried to do with this framework is teach a locomotion simulation robot to do a back flip, which is a not a well-defined thing in that environment but they can try and do it by trial and error.

30:42 HK: And then another area of this work is robustness or adversarial examples, which is that AIs often don't know what they don't know, so they might get trained on a big data set of images, but then they see a new kind of image with a new kind of noise or fuzz or filter and they not only will get it wrong, they'll be completely confident and they'll be wrong. And so, having some kind of way for AIs to be better at knowing, "Hey. Something weird is happening. Instead of telling you this image is a turkey, I'm going to tell you I don't know what's going on and maybe we should skip this one."

[laughter]

31:16 HK: Would be really nice. And so research on those kind of topics and we're trying to support research on those topics, partly because we think they're good topics, but partly because the real intervention that I'm excited about with AI safety is field building. The real thing that I'm excited about we don't know where this is going. We don't know if very powerful AI is ever going to exist. We don't know when, but if and when it does, I think we'll be a lot better off if there's already 100 great people who've been thinking about this kind of thing for the last 10 years. No matter what form it takes, no matter what challenge we're facing, that seems good. And it seems like we're much better off with that, and that's the kind of thing a philanthropist can do is build a field.

31:52 HK: So things we've been doing to build the field, we did the field building case study, we talked to a lot of people, and one of the things we thought was, when you're building a field, one of the challenges, it's hard to get senior people to switch topics. It's hard to get junior people to go into a topic with no senior people working on it. That's tough. And so we tried to figure out the right order there. We ended up designing a request for proposals. We funded several top senior AI researchers to devote a portion of their lab to this kind of research, and our next move is going to be to fund more junior people now that we think there's a proof of concept and a career path there, and we also fund workshops for people to discuss and try to improve on their work. And then that's all in academia.

32:37 HK: And then our OpenAI grant is an attempt to say, another challenge of the AI field is that it's not an academic field. It's becoming more and more an industry dominated field. The good jobs are in industry. The cool stuff, not all of it something like half of it, maybe more, is in industry. And so unfortunately, we can't really fund Google Brain or Google DeepMind to work on safety. But Open AI is basically an industry non-profit and so, we made a major grant to them, took a seat on their board and have been basically looking for spending time with them and trying to think about how to get that culture to change and how to get these safety topics to be as interesting to people and as prestigious as the things that AI researchers currently work on. Cool. Yup?

33:25 Speaker 9: This may be a little bit related to the organization. So if you've got an Open Phil [33:28] ____ here in New York and then you're going to move to San Francisco?

33:32 HK: Yeah.

33:32 S9: Can you talk about some benefits or drawbacks of that? [33:35] ____.

33:36 HK: Yeah. Benefits or drawbacks of moving to San Francisco from New York. So we moved for a pretty simple reason. We wanted to work more with Cari and Dustin, and they didn't want to move, so...

[laughter]

33:46 HK: I think it was good for us. I think it's been more good than bad, so I think that certainly our partnership with them has grown. Also, it's been much easier for me to interact with the Effective Altruist community which I think we don't necessarily, it's a diverse community. We don't necessarily agree with everything that you might have heard coming out of it, but there's a lot of really great people there who have helped us think through what our top causes should be. And a lot of my heuristic these days is I'm always, I don't have time to think through anything as much as I want to, I'm always looking for the person who's thought about something kind of how I would if I had a zillion times more time. There's a lot of people like that in the Effective Altruist community. So that's been good.

34:30 HK: I think that we are also trying to meet and just get a sense for what people are interested in of people who are going to be major philanthropists in the future. So people who founded companies that are worth a lot of money who might be doing philanthropy in the future, and it's much easier to do that in San Francisco. There's a lot more of 'em, and the weather's nice. Glad we did it. We were five people when we moved. Now we're two organizations and each of em's like 20 people, so I think that's where we are now. Hi, yeah.

35:01 Speaker 10: There was a comment earlier about how GiveWell's Incubator grants avoids some of the administrative burden traditionally placed on grantees and I'm wondering if Open Phil works on that model and maybe at a broader level...

35:16 HK: Sure.

35:16 S1: How do you see innovating in the space of the traditional foundations?

35:21 HK: Sure. How do we keep the burden on grantees low and are we innovating on the traditional process? So first off, we don't usually go into anything looking to innovate. We just try to do the best thing, learn from what's already out there and do what makes sense to us. Our grant process does have, it's very similar to the GiveWell Incubation grant process. They really kind of grew up together. And the Open Phil process is also designed to reduce some of the overhead relative to grant processes that I've gone through myself for example as a grantee. So I think one of the big differences is a lot of foundations, they will say, "You've gotta submit this detailed application. You've gotta do all this writing, and then we will start to take a look."

36:02 HK: For us it's more like it's the program officer's job to do the write-up. It's the program officer's job to answer the questions, to convince the decision makers that the grant is good and the program officer asks the grantee for whatever they need to do that, and no more. And so it's the same king of thing. I think for some organizations it's actually much worse and much harder, and for some it's easier, but it's a different kind of thing. The expert in the field is talking to that organization. They might have to take a lot of their time with questions, but they're just trying to talk to them and figure out what they need to know to make the case instead of asking that organization to have one of their grant writers write up the case.

36:39 HK: I find that model better, I don't know if it's easier on grantees overall, but I think it's easier on them given how much info we want. So I think the other method would be like a disaster given how much, how tough we like to question grantees. So I think it is better given that. And I also have a little trouble seeing the first process result in grants that I would want to approve.

37:07 Speaker 11: Has GiveWell significantly increased the, sorry. Has Open Phil significantly increased the spirit of transparency within the philanthropy community?

37:15 HK: Has Open Phil increased the spirit of transparency within the philanthropy community. And I'm repeating things for the recorder, by the way. I wouldn't say so. First off, I think that foundations generally, they're organizations. They have culture and practices and it's a big lift for a foundation to fundamentally change what it's doing. So I think if we're going to really change the way philanthropy is done, it's going to be much more through our influence on people who aren't philanthropists yet. And I think there's been some of that but I wouldn't say that I've seen real movement on that front right now. I do think it's true that we share a lot more information about what we're doing than other foundations do. I don't really see them moving that direction right now. Not that I can tell.

38:05 Speaker 12: On a somewhat related note, so other foundations that are evidence-based, the Arnold Foundation or the Gates Foundation, do you think they've done research that will be useful to you guys that you might be replicating? Are they friendly about sharing it? And do you think that you are fundamentally different about them in from them in some way in your approach?

38:23 HK: Sure, so do other research-orientated foundations or results-orientated foundations, do they share information with us and do we think we're different from them? First off, yeah, we're pretty friendly with all those groups, and we generally try to have good relationships with other foundations in our space, or other foundations we can learn from. Usually funders do share information with each other. It's kind of nice to talk to someone who isn't asking for money. [chuckle] And I've generally found funder relationships seem to be pretty good with each other generally.

38:54 HK: A lot of times we just work in different areas, and so there's just a limited amount for us to really learn from each other. And a lot of times that's by design, because we're trying to work in neglected areas. Do I think we're different from other foundations? Yes I do, I think the biggest difference is, the big one is the way we've chosen our causes, so that's just fundamental. Conventional wisdom about how to choose causes in philanthropy is to pick what you're passionate about and go from there. And that's just like the opposite of what we did. We spent years trying to pick the best focus areas to work on. Certainly, we're not the only ones who've done anything like that. Definitely people consider what causes seem important when they think about what they're excited about. But I think the level of intensity and intellectual picking apart of what causes we should work on and why, I don't believe anyone else has done something like that.

39:44 HK: And the upshot is that we work on very different causes. I know that there aren't any other major foundations in farm animal welfare. I know that there aren't any in potential risks of advanced AI. There's some in biosecurity, but none focused on the sorts of pandemics we're most worried about. In other words, the really big ones. I think it's a difference in philosophy and methodology that's led to a dramatic difference in what we work on. The program officer model is common. We kind of more like stole it from convention. It's not something that we came up with. I have heard, and this is where things get into nitty-gritty, and it gets really hard to generalize, I've definitely heard that we have a different relationship with our program officers than other foundations do with theirs. That we give them more autonomy, it's more a system of they're in the lead and we come in and question them rather then they're carrying out something we told them to do. Those are some differences. And I think there's lots of little differences as you get further into it, but I think those are the really big ones.

40:45 Speaker 13: For the AI grants you mentioned more long term things, what about the short term economic impacts?

40:51 HK: Short term economic impacts of AI. So first off, I'm not always on board with the way people distinguish between short and long term AI risks 'cause I think there's so much uncertainty. One of the things I think is that at the point where AIs are able to do 50% of human jobs, there's a good argument that we're really really close to 100%. Like really really close to the day where they're able to sort of do anything a human can do and do it much better, because the difference between the most capable and least capable humans is not terribly dramatic if you kind of zoom out and look at the brains or something like that. Even the difference between humans and chimps, there's not a lot of evolution in between. There's not a lot of difference in brain size.

41:37 HK: I don't think that we can or should confidently say, "First this is going to happen, then that's going to happen, then that's going to happen." And part of the reason we're interested in these very dramatic transformative impacts of AI is because we don't think they're necessarily further away or coming with more warning than some of the less dramatic stuff people say. A lot of times I talk about slow takeoff versus fast takeoff.

42:08 HK: In other words are we going to gradually see AI on more and more tasks and gain more and more economic value or is something going to snap into place. A lot of things humans do, like we need to do a lot of things that all relate to each other in a complicated way. For example if I'm trying to decode the word someone's saying, a lot of times my knowledge of who they are and what they do for a living matters. I think there are scenarios in which AIs take on a little more time and there are scenarios in which they get some basic abilities and there's a snap into place where they're able to interact with human contexts. I don't think it's obvious which one is coming first and that is why we're more interested in the transformative stuff. I do think that this sort of automation challenge is a real challenge, I think it is already happening, has been happening for a long time.

42:57 HK: I don't think we're necessarily going to be looking at unemployment but we might be looking at a situation where returns to productivity just go more and more to the elite knowledge workers who are benefiting from the AIs instead of the people who are sort being shifted from job to job to do what the AI still can't. So I think that is something that's already been happening; you can already see that economic growth over the last couple of decades has been going

much more in a skewed way, much more to the best off than it did before. That creates challenges we're already reckoning with. I do think those challenges are fairly likely to get worse, and that is something we are interested in funding more forward-looking analysis on, especially forecasting, when will AIs be able to do what and how can we plan for this. And what kinds of retraining will we want and what kinds of education will we want. I think it's something that people could be doing more on than they are, but it's not our top priority in the area. In the back.

43:54 Speaker 14: You talked a lot, a little piece about forecasting, and I know you're not GiveWell, so you're not doing, say, specific recommendations for specific causes but at the same time, are you at a point where you think you could put your money where your mouth is and your mouth where the money is, and between the different catastrophic risks and assess some relative importance or forecast probability percentage towards where these things might go?

44:24 HK: Sure, do we think we can make good forecasts yet of where things might go and do we feel like we're reliable yet? And I would say no. I think we've thought through the relevant figures and the relevant questions to an extent greater than most others but a lot of the stuff we do, we're just accepting the fact that we're working on areas that we might not be very good at, that maybe no one is very good at. Looking 20 years into the future and seeing what's coming is just not something that humans have a very good track record of. We've tried to study that track record ourselves a bit and are looking into it at the moment and we try to just do things, we try to pick areas that are worth working on and do things that are worth doing that don't rely on a highly precise probabilistic understanding of the future. I think it's especially hard to make any kind of forecast in between zero and one percent or in between 99 and 100%, like what are the odds that someone releases a horrible pandemic and wipes out half of the planet in the next 100 years. Is that 1%, is that 0.1%, 0.01? That is a tough area to get into quantification and I wouldn't say that we've done done it or that we pulled it off but we've said "That's an important area, let's look for some things to do that are robustly good." Yep.

45:43 Speaker 15: [45:44] ____.

46:20 HK: Sure.

46:20 S1: How do think about the difference between safe bets in some sense where nothing very bad will happen versus bets where the expected value is very high but your [46:31] ____ expected value calculations, the worst case scenario is that you make things significantly worse.

46:38 HK: Yeah, so how do we think about adversarial or controversial causes? Cause it's one thing if you're just saying "Well if we're right we'll do a lot of good, if we're wrong, nothing happens," and it's another thing if you're saying "If we're wrong, we can really make things much worse." I would say it's true, that it's a difference between Open Phil and GiveWell. Not everything we work on is controversial, but some things are and some things give us the opportunity to make things worse. I think all we can really do there, one I think it is true that else equal, if you've got two things that look equally good in isolation and one of them is uncontroversial and the other one is controversial, you should go with the uncontroversial one, I think that's the easy call. On the other hand, like I said our philosophy of trying to swing as hard as we can for the really big hits, I think means you can't just leave things on the table because they make you a little uncomfortable. And there's a lot of different things that you can't...

47:35 HK: If you start saying, "Well, we'll take risks but not if they might do harm and not if someone might look at this and think we did it for the wrong reasons, and not if this, and not if that and not if that," you're going to lose a lot of your best opportunities. We don't want to shy away from that stuff too much, the thing we do want to do is, A: Really do our homework, do our homework more than anyone else, try and be as informed or be connected to the people who are as informed as anyone on the topics we care about. And B: Just be good citizens.

48:04 HK: So we'll fund people who are working for the things we care about, but we're not going to do or really fund things that involve deceiving people, lying, breaking the rules, using coercion. That's us kind of saying we're going to advocate for what we want, we're going to try and do it fairly, we're going to try and do it honorably, and we're going to try and be as knowledgeable as anyone else doing it, and that's pretty much what we can try to do. On incarceration specifically; you do rightly point out that maybe if we succeed in reducing the amount of incarceration in this country, maybe that's bad, maybe that is a public safety cost. And I can't promise you that that's not the case.

48:44 HK: But we did; David Roodman who is our senior adviser who has done a lot of research projects for us, he spent like a year going through the literature on the relationship between incarceration and crime, has a paper coming out on this soon that I think is really excellent, I would basically recommend everyone read it; especially anyone who

picks apart science, takes studies, picks them apart, exposes their flaws, and I think David is that is the best person I've seen anywhere really at going into a large confusing academic literature, sorting out the good studies from the bad, then taking the good studies scrutinizing them really hard, I think studies have like a 50% mortality rate when David looks at them. And then putting it all together making it understandable, and telling you what the most sober, even-handed, critical assessment would say.

49:32 HK: That's the kind of thing we try to throw at problems that are confusing us, where there's the literature to look at, and that's the general attitude we try to bring. And David's conclusion at the moment is that on the current margin, so where we are now in the US, further incarceration, his best guess is that would have zero benefit to reducing crime. So the positive aspects would be cancelled out by the negative aspects because there are ways in which putting more people in prison can create more crime. So that's his best guess and he has a whole lot more analysis in that paper that's coming out soon. Yeah, Tyler.

50:11 Tyler: Can you give an update on the science areas?

50:13 HK: Yeah, update on the science area. So for a long time I would just show up with these things and we'd be like, "We haven't really figured out how to start doing science yet." So finally I have some news. We have a couple of senior science advisors, one of them has just stepped down as professor at Berkeley, and what our basic approach has been, broadly speaking in science, is we did a lot of things, we did a lot of experiments to try and find cool science stuff; and I could go through them all but I won't, and where we kind of landed was, we landed on saying, "Look, the science we'd most like to do is science that directly relates to other focus areas that we've already chosen," so the things I've talked about before.

50:57 HK: Science that could help prevent the next horrible pandemic, science that could help create replacements for meat that would lead to less animal suffering, things like that are our top priority. And then after that, what we kind of believe is that our advisors are able to look at an area and sometimes just see something ridiculously good, and if they do and it's an area that we think is pretty good then we'll do it, and if they don't they'll just move on and look at another area. So it's it's very opportunistic, it's not so much cause-based. It's more like we look in a lot of corners of science and we just look for great studies or approaches that fell through the cracks.

51:33 HK: So examples of things we've done, so one thing we did in this spirit the National Institutes of Health, the big government funder that spends like tens of billions of dollars a year on medical research, they have a transformative research award program. So they ask people to submit things that are too ambitious, too big, too risky, for the normal funding mechanisms and then they fund the best ones. But they don't fund many of them, I think they fund like probably something in the $10 million a year range of this stuff, out of tens of billions of dollars in their budget. So we basically got their reject pile, we got all the things they hadn't funded and we invited; well we basically invited everyone who had been rejected to apply to us and just send us the same application, then send us the reviewers comments if they had them, and we went through them.

52:16 HK: We found definitely some interesting stuff there; such as someone who is trying to build Nanopore sequencing for proteins, so basically if you put a protein into this machine you should be able to cheaply and quickly find out what the protein is, that can have a lot of applications. So we're funding a number of those, that's an example of how we explored for things that are really outstanding.

52:41 HK: And then other things we found while kind of poking around in areas were excited about, we are working on a potential broad-spectrum antiviral drug that would, might not work out and even if it did, wouldn't be useful for mild viruses 'cause it would have side effects, but could be really useful in a very bad pandemic situation. Another thing that we've done is an investment in impossible foods; which is our best guess at the most promising thing for replacing meat with other products. So the impossible burger is a burger made out of genetically engineered plants in a sense, or it involves genetically engineering plants to produce a certain protein which they then put into the burger, and as someone who likes burgers I have found it the most impressive veggie burger.

53:28 HK: So you can take that for what it's worth though, we're investing, so disclosure. We've also recently announced a major grant, like something in the 20 million dollar range for gene drives, which is a technology for genetically modifying mosquitoes in a way that they transmit the modification to their offspring highly reliably. So if you can get one mosquito to be malaria resistant, they can't carry the malaria parasite, and to transmit that quality to their offspring, and those offspring to transmit it to their offspring this is the thing we've seen that we think is like, "If we wiped out malaria, this would be how we did it." And so we made a major grant. This is what the Gate's foundation's

been funding for a long time. We made a major grant to try to speed that up so those are... That's the basic MO in science and those are some of the things we've been doing. I'll take one or two more. Yep?

54:19 Speaker 17: On the topic of controversial causes, is there any concern that, let's say you take cause like macroeconomic policy, migration, and you figure out it would strategically have an outside impact on the policy and you're sort of subverting democratic process in a sense, so even aside from your belief about yourself being correct, do you have a concern about some process should go through democracy...

54:44 HK: Yep.

54:44 S2: Maybe it'll lead to a populist backlash like migration or whatever it is just...

54:49 HK: Yeah. Yeah, do we have concerns about subverting democracy? So I definitely...

[laughter]

54:55 HK: When we work on legislation and politics, the way that I have seen the work we've actually done, I think this is a thing that definitely could be of concern in certain cases. The work we've actually done I would generally characterize as like we're going into an area that's already a war of interest groups and there's no interest group on the side we're on for some reason that is fairly straight forward to see, so there's not a lot of lobbyists for farm animal welfare because the animals can't lobby and that the farmers, they have a very powerful lobby and they want to stop the... They have these things that are hard... It's hard to imagine these are things people would've voted on in referendum, these ad gag laws, that you can't secretly go to a factory farm, take video, and publish it to show how people are treating the animals. That's illegal.

55:45 HK: So I think when we go into an area like that, and most of our work on farm animal welfare is not policy, but most of these areas we work on, it just doesn't... The way the system is set up, there's a huge number of interest groups. A lot of them are incredibly well-funded, a lot of them are very powerful. We're not very big in the scheme of things and all, and we generally fund people who make arguments for their ideas who organize people affected by policy, who try to make the case again in a fair and honorable way and yeah, I basically just don't...

56:18 HK: I think that we are taking a system that is skewed against certain ideas and against certain populations and trying to defend those ideas in those populations within that system without ever being the big gorilla in the room. That's how I've felt about it so far and there may be exceptions to that in the future and I don't think anything is automatically off the table for us, but that is how I see the work we've done. Alright I'll take one more, who's got it? You already have one.

[laughter]

56:48 HK: Wait, did you already? Alright, go for it. Yeah.

56:51 S?: Well, I have a similar question so... In terms of lobbying, so when you think about [56:54] ____ global health and, for example, how much money Against Malaria Foundation had raised, you could argue that maybe it would become more cost effective to spend a small amount of this money to convince a [57:05] ____ or a department or a state department somewhere of one government to fully fund the initiative.

57:10 HK: Yeah.

57:10 S?: So do you have a more structured way to try to lobby different government to apply more effective giving?

57:17 HK: Sure. So, could you make the argument that instead of giving to AMF, you should give to someone who lobbies for more foreign aid and how do we see that thing in our own work. I think you could make that argument for AMF. I don't think that it would be a giant clear win. There is already a foreign aid lobby. There's a lot of these NGO's that lobby for foreign aid to continue to be generous and generally, it seems to be kind of a war that keeps ending at the status quo. And Gates funds can work in that area too so I don't feel like I would expect if we just put that AMF money into lobbying I wouldn't particularly expect much to happen as a result. I don't think that the ROI would be better. Might be worse, it might be like the same, it might be better but it's not clear to me.

58:00 S?: [58:00] ____ from existing projects to more projects?

58:03 HK: Yeah. So you could advocate for re-allocation of aid. Yeah, we did look into that space a little bit and didn't come out feeling like there were any easy wins there. Most of that aid is being spent the way it's being spent for some reason. You can try to fight it and you might win, but it doesn't look like there's any super low hanging fruit to us, like anything obvious. But certainly a lot of the work we do is trying to change minds and organize people and influence the way policy plays out, and we do try and be as strategic about it as we can and I think it is definitely a more complicated thing than just sort of like you pay lobbyists and the lobbyists buy the results you want. I think it's... We wrote a post called I think "The Tools of Policy Oriented Philanthropy," and it lists several different things you can fund to try to influence policy. And I think you want to come at it probably from several directions because you need the ideas to be well formed and the legislation to be worked out and you need the constituents who care to be loud and a lot of that stuff matters a lot compared to some of the more quick pro quo type things people picture with politics.

59:12 HK: So we try to be strategic. I'm not aware of any policy issues where I think if you just walked in with $10 million or even $100 million that I could guarantee you some kind of a win, but I know a lot of areas that seem worth worth working on. Cool, well I'm going to call it there and I'll hang around for a little bit. Thanks everyone.

[applause]