

**GiveWell San Francisco Research Event December 4, 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: If anyone at any point wants to take something off the record, you can either tell me then or you can tell me after, just remove it, but by default, we're recording, and I'll be repeating questions for the recorder.

00:10 Speaker 2: Holden, can I say one quick thing? We did this thing where you submitted questions on your phone, Holden is old school, so just raise your hand, he'll call on you and verbal question asking in this session.

00:17 HK: Yeah. I was going to say.

[laughter]

00:22 HK: We're doing the traditional format, and so I'm just going to talk for 10 or 15 minutes, and then take questions, and it's just me. Maybe we'll try to give away in the future, but want to see how it works out first. [chuckle] So, basically, like I said, I'm going to go through the basics of Open Phil. I think a bunch of you know the basics or have been here before so I'm going to go through on the quick side and happy to take more questions and try to talk a little bit more about what's been new at Open Phil most recently, but I am going to go through the basics, just so everyone is caught up and just like what we are, what we're about, how we're different from GiveWell. I obviously was at GiveWell, co-founded it with Elie, and after several years there we met Cari Tuna and Dustin Moskovitz, who were trying to give away a very large fortune within their lifetimes, and this could be billions of dollars.

01:15 HK: And we wanted to help them and saw that the problem they were wrestling with was, in some ways, very similar to the GiveWell problem, and in some ways, very different. So the way in which it's similar is they were looking for... Had a certain amount of money to give, they wanted to do as much good as possible with that money, and we're having trouble finding resources, finding guidance, and so we were looking for a way to help them with that, and simultaneously create a better public knowledge base or common knowledge base for other philanthropists in a similar position, to get up to speed more quickly in the future.

01:47 HK: So far, that's very similar to GiveWell, the difference is just the scale and the style of the giving that they could and were going to do. As people who are giving away possibly billions of dollars in the course of their lifetimes, that is a very different thing from when we started GiveWell, it was originally trying to solve the problem of, "How do I give away, let's say, \$5,000 and I only have a few hours to think about it?" So those are obviously different and a lot of GiveWell donors do give a lot more than that, and Cari and Dustin do give to GiveWell Top Charities, but they're different kind of animating or original problems. And so some of the differences are that, with Cari and Dustin, they can spend decades themselves, getting immersed in the issues, learning to be full-time philanthropists, and can also build their own networks, hire their own staff, and start doing things that smaller donors might not be able to, such as creating new organizations or transforming existing organizations. And so the options look pretty different.

02:45 HK: And one of things that we've started to hone in on at Open Philanthropy, and I don't think it's the only good kind of giving even for Cari and Dustin, but it's this idea of hits-based giving, and in some ways, is the opposite philosophy from GiveWell, in that rather than focused on things, where we've checked them out and we understand the case for them and we've established some strong case that something is likely to work or has high expected value. A lot of times what hits-based giving looks like is more like trying a lot of things and expecting a lot of them to dramatically fail and hoping that there will be a few successes that are big enough to justify the whole rest of the portfolio.

03:27 HK: And so its an approach that may be familiar from venture capital, and when we looked at the history of philanthropy, we did feel, and we've written about this, we did feel that there were enough good examples of major wins in philanthropy that could have made up for many failures. We actually have conference rooms over on that side that are named after what we think are some of the biggest philanthropic wins. But one example would be that the Rockefeller Foundation funded this research on improved crops, and this is widely believed to have led to the green revolution that led to several countries going from very poor to developed or quasi-developed and pulling, saving more than a billion people from starvation. That's the kind of thing where it would be fine to try 10 things like that, and have only one of them working. [laughter] It would still feel that you got a pretty good return on your portfolio.

04:13 HK: So that's the basic philosophy. And in terms of what we do and why we do it and where we've been, Open

Phil started by studying the history of philanthropy, studying what existing foundations do, and in our first few years, we really spend our time trying to pick good focus areas. So just topics to focus on, topics to hire around, and we looked for problems in the world that we can work on that were important, neglected, and tractable. Important means, if we were to do something about this problem, if we were to get a win, it would be a big deal, it would affect a lot of people in a very positive way. Neglected means, there's not too many other people working on this. All others equal, would rather work on a cause where we're one of the only funders, and we can be one of the transformative funders in the field, rather than working on some cause where there's already a zillion people and a zillion dollars going into it and so we just have to play around the margins. And finally, tractability means that we see some sort of path to impact.

05:13 HK: To give an example of this, we looked at a lot of different US policy causes, we did a lot of short investigations, where we tried to get a basic idea. Here's a policy issue, maybe it's Criminal Justice Reform, maybe it's gun control, maybe it's education policy, and we'll ask these basic questions, if we can get a win on this issue, how big a deal would it be? How many people would it affect? That's importance. We ask neglectedness, who else works on this? Who else is advocating? How much will we be adding? And then we ask tractability, like do we think we can get a win?

05:45 HK: And so for our for our opening years, we mostly focused on picking good problems to work on. We call these focus areas based on importance, neglectedness, and tractability. And then we focused on hiring people who could specialize in these focus areas, become true experts. A lot of times, come in with a lot of existing expertise and then lead our giving. And the overall Open Phil philosophy is pick the best issues to work on, pick the best people to work on them. That's where a lot of the hard work for people like me comes in and then from there, let the people you've hired lead the way.

06:21 HK: Obviously, ask them tough questions, evaluate how they're doing, assess how the portfolio is performing, but overall, our strategy in a given cause, for example, in Criminal Justice Reform, our program officer is Chloe Cockburn, in Farm Animal Welfare it's Lewis Bollard, and if someone asks, "What's your strategy?" To me if someone asks, "What's your strategy in Criminal Justice Reform?" my first order answer is Chloe, or whatever Chloe thinks our strategy should be. Obviously, we interrogate that, we ask tough questions, but the idea is that we're not set up to do one kind of intervention. We don't say, "Well, we do impact investments, or we don't do impact investments, or we do advocacy, or we don't do advocacy." That's not how we slice things. We believe in picking the right issues, the right people and then from there, there's all kinds of different ways you can do things, but we want that person to be the truly most knowledgeable expert who can do that.

07:14 HK: And then another core part of our philosophy is the minimization of decision makers and veto points. For hits-based giving, you want to take big risks, and that means sometimes doing things that are very bold, things that maybe are not easy to explain to others or that others would look at and not immediately understand. And if you want to do that, I think the worst thing you can do is have the whole organization vote on everything you're going to do. I think the best thing you can do is really empower the people who know the most, and so a lot of how we've set things up. For example, our program officers generally follow what we call the 50/40/10 rule, which is that when they're giving away money or recommending grants, when we look at the portfolio, we want 50% of the dollars to be things that the ultimate grant approvers, and those are myself and Cari right now, things that the grant approvers really buy into, understand, could argue the case for themselves. But it's okay if another 40% of grant dollars bring us to 90% total, are merely things that I can see how you could think that was a good idea if you knew a lot more than me.

08:22 HK: And so a lot of times I'll read a grant, I'll say, "I don't feel sold on this, but you know the area better than I do and I see where you're coming from, so we're going to approve this one, but it counts against that 40%." And then the final 10% of the grant portfolio tends to be discretionary. And that means we have an expedited process. The program officer doesn't really need to make a case, they just need to tell some basic stuff about the grant and whether or not we buy into it, it goes forward as long as we don't see any major red flags. And so the idea there is that we're trying to focus a lot of the decision making in the people who know the best, and we're not trying to let grants get held up by needing them to be legible to everyone involved in the organization, and even sometimes to the decision makers. So that's the basic Open Phil philosophy.

09:06 HK: And now what I'm going to do is I'm going to quickly go through the causes that we work on, and discuss what we've done so far. And so I'm not going to connect all the dots between the process we've used and how we arrived at each cause. That's something I've talked about at previous events. Instead, I'm just going to list the issues we work on. If people have questions about why those are the issues, I can answer them, but they were chosen based on importance, neglectedness, and tractability. And then I'll talk about how they're doing. And so one of our causes that I'll talk about first is Farm Animal Welfare. There's billions of animals being treated incredibly poorly, all across the world

every year on factory farms, in the food industry, and we believe this is a cause that I, myself, I'm not naturally someone who would have thought to prioritize animals, I'm not a big animal lover, but when you look at just the magnitude of the suffering, how poorly they're being treated, and how few others care. I can consider us by far the biggest, and in some ways, the only really major funder in this area.

10:08 HK: There are really big opportunities to help improve the welfare of animals. And they're so big and they're so much bigger, I think, in some sense than a lot of the other causes that it starts to look like in some sense, it might be a good return on investment or a good philanthropic deal, even if you value animals and only a fraction or even a small fraction as much as you value humans. So to give a sense of that, a lot of the work we funded so far has been corporate campaigns. And so people will go after a fast food company or a grocer, and they will say, "We do not appreciate the way you're treating chickens who are producing the eggs. The chickens are in battery cages. The cages are incredibly confining, incredibly inhumane. They don't let the chickens practice natural behaviors and for a tiny amount more money per egg, you can take the chickens out of battery cages. It still won't be a good situation, but it'll be a lot better." And what's happened so far is these campaigns have been incredibly successful because you go to a corporation, you point out something that would not cost them much extra money, but would have huge benefits for chicken and you basically make life difficult for them until they make that pledge to switch.

[laughter]

11:16 HK: And so we've been funding groups that do that. And so far in the couple years we've been at that, these groups we've funded have generated hundreds of corporate pledges, potentially benefiting about 300 million hens live at any time. We've also been working on broiler chicken welfare. So that's chickens that are raised for meat. And there it's earlier days, but over 50 corporate pledges, potentially benefiting 65 million chickens at any given time. That's from the latest report from Lewis. The size of the wins here is massive, just the leverage on your dollar for the amount of incredible suffering you're preventing per dollar. Again, you don't have to think that chickens are the same as people to think this is a good deal. You could think they're a small fraction and still say, this is impressive.

12:01 HK: This is one of the only areas where we've seen, I would say big large scale wins already. And I think in some ways, it's because it's such a neglected area that there was this low hanging fruit, and there were these wins available for the taking, but we're also working on the longer term. So a lot of the other work we do is research on animal conditions to try to create better standards and a better understanding of what further interventions could improve animal welfare at scale, and then we're also doing a lot of funding overseas so most of those pledges I said are in the US and in fact, as I understand it, every major grocery and every major fast food company in the US has now made a cage-free pledge. And so hopefully, 10 years from now or so, you won't even be able to get non-cage-free eggs, but we're also looking to other countries such as China and India where as wealth goes up, as meat consumption goes up, the scale of additional suffering could be massive and it could make a huge difference whether there's an active animal welfare movement or not.

12:56 HK: And so for example, we're the only major farm animal welfare funder, maybe the only funder working in China, and so that's an example of playing a longer game and it may take a very long time to see any fruits from that, whereas on the corporate campaigns, we came in when there was already a little bit of momentum, and I think our funding largely turbocharged things. Criminal Justice Reform, it's a US policy issue that we picked because we thought we saw political opportunity. We thought we saw the opportunity to get again some shorter term wins and to make a difference and so far, I think we do have some initial signs of having impact. Though, I hope to have more in the next couple of years. But one of our grantees, one of our biggest grantees Alliance for Safety and Justice played a crucial role we believe in an Illinois bill that passed by Partisan support that will lead to a 3,000 person reduction in any given time in Illinois incarceration.

13:50 HK: We've also been the largest supporters of the CLOSERikers campaign, which has been advocating for the mayor and the governor of New York to commit to close this horrible Rikers jail which as we understand it, would necessitate actually making reforms and lowering the jail population because there's not really any other way to do it. And since we did start funding that campaign, there has been an announcement from the mayor that they do want to close Rikers, there's been a plan. The plan is 10 years. That's too long in our opinions and the campaign's opinions. The governor has now publicly gotten in a spat with the mayor where the governor says they should be trying to close it faster, so that's all been interesting, and none of this, don't want to give any impression that we think this is all due to Open Phil.

14:30 HK: All the impact we have, we have grantees, the grantees have partial credit for the good things that happen,

and we have partial credit for the good work our grantees do. So that is the basic philosophy there. Other major causes we work on. I would say we haven't seen as much results yet because they're kind of inherently longer term in the way that we think about the time horizons. So, one of the major concerns for us is what we call global catastrophic risks. So can you imagine anything that might happen in the next 50 or to a 100 years that could be so far-reaching and global and catastrophic in scale as to derail civilization in a way that could affect very large numbers of future generations. 'Cause that would be another way to get very good bang for your buck... [laughter] Is if you could prevent something really bad like that. So to give a well known example, nuclear war or climate change, those are both things where if you play a small role in reducing the risks that one of those really knock civilization off course, you could then have a small role in better lives or existing lives for many many future generations.

15:34 HK: The two biggest global catastrophic risks that we see are pandemics and artificial intelligence. It would be a bit of can of worms to get into exactly what we're thinking behind both of those, and I'm happy to answer questions. But briefly on pandemics, we think that... First off, there was a pandemic immediately following World War I that killed significantly more people than World War I had killed and it only happened in about a year. And so the potential for pandemic in today's interconnected world to go global and really do a lot of damages is big, and especially as science and synthetic biology advance, the kind of nature of the risk we think could be getting worse. And so some of the things we've worked on there, we're working on trying to support and keep operational the Biological Weapons Convention, which is an international agreement that basically no government should be working on developing or potentially deploying biological weapons. We think that's very very important. Governments, state actors are best positioned to design the things that could really cause worst worst case scenarios and we believe that that is a bit of a vulnerable institution that we are trying to help out, basically in some sense, they're kind of in a liquidity crunch and we're trying to figure out a sustainable way to do something about that.

16:51 HK: And then we've also funded some major organizations to free them up more to think about worst case scenarios in bio-security, and that includes the Nuclear Threat Initiative which had a dormant bio-security program that we recently decided to make a grant to bring that program back. And then on potential risk for advanced AI, I won't go on too long about this now. I might have to take a break in the middle to get to it if I get a zillion questions. But basically we believe that if artificial intelligence was designed that was able to do certain things such as science in a superhuman way, the way that today we have artificial intelligence that can play Go or Chess better than humans. That could be a civilization defining moment, and there are major risks there.

17:37 HK: Risks both pertaining to, it would be really bad if let's say, North Korea was the first to develop or deploy that kind of AI because it could have military implications, or two, any AI that was incautiously designed or deployed could lead we think to globally catastrophic accidents. And so a lot of what we do here is we try to build a field of people who are preparing for potential risks. We have no idea when this might happen or if it will ever happen but we think there's a decent chance it could happen within our lifetimes and if it does we would like the world to be prepared. We would like there to be fields full of people who have spent their entire lifetimes thinking about what could happen and what we would do about it.

18:17 HK: And we have studied past histories of philanthropists trying to build fields and we're trying to apply our lessons. For example, we've been funding both senior researchers to devote more of their labs to AI safety research and we've also been... We recently had a fellowship program, that deadline just closed, to get more junior researchers to be able to base their careers around the idea of AI safety and not just around AI capabilities.

18:42 HK: Finally, we do a lot of scientific research funding. We have a science team. And that's something I know at past events, I've often talked about how we are still figuring things out. And we still are to a degree, but we have more of a system now, and again with science we look for things that are neglected, that have huge potential to change the world. Some of what we do is directly relevant to our other causes so we've made an investment in Impossible Foods, which is just trying to design a burger, essentially from plants, that can really compete with a meat burger, and that can greatly reduce animal suffering as well as other issues.

19:15 HK: And then we're also now funding some work around trying to create a broad-spectrum antiviral. A drug that could effect a lot of different viruses and therefore could be very useful to stockpile for a worse case pandemic. We also did a program where we went to the government main funder of biological research, The National Institutes of Health, and they have this high risk research program where they fund especially bold ideas. That is, very small, and they don't fund the vast majority of them. So we basically took all of their proposals they had rejected, went through them again...

[laughter]

19:48 HK: And funded some of those ourselves and that led to some interesting stuff. That is most of what we're working on now. There's a lot going on at Open Phil and because of the fact that we often try to empower people to be the experts, there's a lot of things that are just happening that I only understand the high level. But I will do my very best to answer questions about what we're doing and what the progress has been so far. And then I'll probably take a break in the middle to talk about just what we're looking at going forward and what our major priorities are now.

20:24 HK: Questions? Yeah.

20:25 Speaker 3: Can you give an example of something that's not legible at your level of knowledge and walk us through it a little bit.

20:33 HK: Sorry, can you clarify?

20:34 S3: When you talked about how the 40% are things that you think may be plausible but you can't really tell for sure or you're not 100% sure you agree.

20:42 HK: Yeah. A lot of grants... The question is, can I give examples of grants that are not fully legible and that might get classified in this 40% bucket. The bucket of, "Sure I can see where you're coming from but I don't really buy into it."

20:55 HK: And there's just a lot of grants that, the biggest case for the grant comes down to how you feel about the people involved. Whether you think they're good and whether they're going to get things done, and that's something that a lot of times the best way to get a read of people, on the people involved is to know everyone in the field and to hear what everyone is saying about everyone else, and triangulate and try to figure out who is good, based on who you've seen do good work, what they say about other people, what other people say about them.

21:20 HK: And so a lot of times the write-up comes to me and it says, for example, "I want to fund this group that organizes people to create more of a voice for Criminal Justice Reform." And I'll say, "Well obviously, if you did a great job of that and you mobilized a lot of people, and you deployed them very strategically, you could make an enormous difference. But I don't know these people and I don't really know how to evaluate whether they're the best people, and I don't fully understand how all the pieces fit together in the strategic picture. I can understand at a high level that having grassroots advocacy, having people stand up and make noise, is good. But a lot of times, this group might be in Florida, [chuckle] and I don't know everything about the situation in Florida and how important it is to have more grassroots advocacy versus more lobbying." And so sometimes, I really ask all the questions I've got, and I say, "Okay I think I get it and I think I would bet on this myself."

22:14 HK: And sometimes I say, "Okay, this basically makes sense to me but rather than trying to investigate the whole thing myself, we're just going to put it in that 40% bucket." Does that make sense?

22:22 S3: Mm-hmm.

22:23 HK: Cool. Yeah in the back.

22:26 Speaker 4: Thanks so much. On your website you list one of the focus areas also as global health and there you say that you support a lot of the areas that GiveWell does...

22:36 HK: Yeah.

22:36 S4: But I was wondering if you could reflect upon that. Particularly in light of your criterion on neglectedness, given that many of the issues that GiveWell focuses on are not neglected.

22:47 HK: Sure. So what's our take on global health and particularly with respect to neglectedness, first I'll register a bit of a disagreement. I think global health is extraordinarily neglected. I think the things that GiveWell works on are extraordinarily neglected so if you look at how much developed countries spend per person on health care and then you look at how much is being spent for various places that GiveWell's trying to work in, I think it's out of kilter and I think when I look at the GiveWell situation, there's situations where you might have whole regions of Africa where 50% plus of children have these parasitic infections and it costs \$0.50 a year to treat those infections with these pills and it would

cost maybe a few hundred million dollars to cover the whole world in this stuff. And that would be peanuts compared to what is normally a number, a line item in a federal budget.

23:40 HK: And even to some degree, not a huge number in the context of foreign aid generally, but that gap is still there. I would call those very neglected areas. There are areas we work on that the field is even smaller. There's even less money going into it, but I definitely think that global health and GiveWell's Top Charities pass the neglectedness test.

24:05 HK: I've talked about hits-based giving, and I've talked about the kind of things we support, and I'm very excited about it, but I also don't think that we have established yet, or necessarily ever will establish, that this is actually better than the kind of work GiveWell does. I think it's still very much an open question, and from everything I've seen so far, I think there is a reasonable case that the good you can do through GiveWell, especially if you're not really on board with some of these arguments about animals, or about the value of preventing extinction. The value you can do through GiveWell is quite competitive with a lot of the stuff we found. Basically, we think that GiveWell's recommendations are some of the best giving opportunities out there, and we continue, as we recommend things to Cari and Dustin, we continue to recommend a significant level of support for those Top Charities. We haven't recommended maxing it out, we haven't recommended filling the whole gap, but we've recommended a significant level of support.

24:56 HK: And, just to be clear, that is our work on global health right now, so we don't do anything else. It's possible that at some future date we'll say, what if we took an Open Phil approach to global health that had more to do with, let's say, advocacy or doing something that might not pay out for 30 years, etcetera. We might do that some day, but we might not, and we may just continue to say, look, GiveWell's Top Charities look approximately as good as anything we'd be able to find, and we'd like to continue to recommend support to them. It's also the case that GiveWell increasingly branches out, and gets more into high-risk, high-reward stuff themselves. I don't know if GiveWell talked about this at the session, but they're certainly looking into things that revolve around advocacy, and revolve around taking bigger risks. So, in some ways, they might end up doing what we would do on global health anyway, and I think for the foreseeable future, our global health strategy is just to follow GiveWell's recommendations.

25:47 Speaker 5: I apologize if this is easily available on your website, but just roughly, how much money does Open Phil grant per year, and roughly how does it break down between the categories?

25:58 HK: Sure. How much are we are granting per year, and how much does it break down between the categories? Last year was between \$100 million and \$200 million, and this year I predict it will also be between \$100 million and \$200 Million. In terms of how it breaks down between the categories, generally most of our cause areas are going to be tens of millions of dollars. I rattled off a few cause areas: Farm Animal Welfare, Criminal Justice Reform, Scientific Research, Potential Risk for Advanced AI, Pandemic Preparedness. Those are the bulk, and those are tens of millions of dollars each. So that's to give some idea. I don't really want to get too much more precise than that, but that's the idea.

26:39 S5: That's great.

26:39 HK: Yeah.

26:40 S?: That's not counting GiveWell Top Charities also?

26:42 HK: Oh, yeah, that's right. That's not counting the gifts to GiveWell Top Charities. So the \$75 Million this year that the Good Ventures is giving to GiveWell Top Charities would be additional to that.

26:52 S?: So do you have a handful of areas that are not current focus areas that are on your radar, but for some reason they're not, because they don't meet one of the three criteria? What are some of the top ones of those that just, they're not quite in there yet because for some reason they don't feel... There's somebody else doing it right now, and you think they're doing a good job, or because you think they're impractical or whatever?

27:19 HK: Sure. So, what are some potentially promising areas that are not on our list of focus areas? Certainly, we're big fans of focusing, picking a very small number of things to work on, and focusing on them very intensely. So there's lots of great causes that are not on the list I rattled off. There are some areas that we work on in a bit of a more restricted way, or we don't have a full time person on them. We have done some work on climate change, reducing the risk of climate change. We have some of the other policy areas listed on our website, macroeconomic stabilization policy, land use reform, immigration reform. These are three areas that we think various aspects of them are really under the radar,

and could do a huge amount of good, and we think they're very promising but for one reason or another, often because we haven't found the right person to take them on full-time, we don't have a full-time person working on them, and so we might give something more in the millions to a cause like that.

28:17 HK: And there's a zillion causes that we don't work... Any cause you name that I haven't named, probably we don't really work on it, or we've worked minimally on it. And there's tons of causes that I can say they're reasonably important, they're reasonably neglected, they're reasonably tractable. The things we work on are the ones where they've either been such a standout on some of those criteria, or they've been very strong on those, and we found the right person to work on them. The choice of person is really integral to everything we do, and so that can affect what causes we go into. We have in the past said, we're interested in multiple causes, we're going to try and hire for multiple causes, but if we get the right person, we'll do it, and if we don't, we won't. Yup?

29:03 S?: I'm trying to understand a little bit more the difference between the GiveWell model and the Open Philanthropy model, and really the longer term view. As I think I understand it, you have different origin stories. You certainly have slightly different strategies in terms of point of view, and what you're trying to invest in. The GiveWell story is about how you direct individual donors in general...

29:28 HK: Yeah.

29:28 S?: And direct their funds in a smart way, whereas the Open Philanthropy origin story is around high, high wealth. Is there any reason why you wouldn't also adopt more of a kind of GiveWell strategy, in terms of trying to encourage just regular individual donors to invest in what you're investing in, [29:49] \_\_\_\_ that?

29:49 HK: Sure. So talking more about GiveWell versus Open Philanthropy, the contrast in styles, and the contrast in audiences. So, you're right, the origin stories are different, but that doesn't mean that the target audience is always have to follow that exact pattern. So, I do think GiveWell Top Charities are an excellent giving opportunity for people like Cari and Dustin, for major donors. And also, it is the case that individual donors might be interested in some of the stuff we do. I think that one, the second one, is trickier. Because I think, if an individual donor says, "Hey, I'm interested in Criminal Justice Reform, but I don't really understand why you did this. Can you tell me? Will you tell me, where to give and why?" We're going to largely say, "That's not how we work." I've described the philosophy. We work hard to pick the cause, we work hard to pick the person. And then there's a lot of deference and there's a lot of trust-building internally.

30:38 HK: A lot of times, for the individual donors, your option is sort of you can decide, largely based on pretty limited info, that you trust Open Phil's judgement and you're willing to go with it, or not. Open Phil does not provide the same level of thorough, legible research. With GiveWell, you can really go to that website and whatever you're wondering, whatever you're skeptical about, is probably addressed. And you can probably find it and it's probably addressed out in the open.

31:10 HK: Those are different models. People in either camp can give to either side. But I think that is a difference in style, which means that if you're an individual donor, you probably don't want to do the Open Phil stuff unless you just have a lot of trust in the people and the worldview and the approach. It doesn't have that same robustness and legibility of the case. Now, I also know that a lot of GiveWell donors give to GiveWell's Top Charities because they trust GiveWell's approach. And not 'cause they personally have checked everything out, but at least they know that the details are out there. And that one could check them out and that the whole case has been spelled out. I think there's just a different level of having to cross that chasm, when you deal with Open Phil. But that said, yes, there are more kind of different styles. One of them, I would say is, more hits-based and more based on this deep context and relationships and trust. And the other one is more based on this very thorough research that kind of covers the whole chain of thinking. Certainly a donor of any size, could be interested in any of those giving opportunities. Yup?

32:21 S?: How much control do you try to exert over the organizations you fund? Suppose there's a group that, let's say a program officer, they like what they do and they want to give a grant to them, but they feel pretty strongly that they can do better than their current approach. [32:37] \_\_\_\_ And they try to steer it in that direction?

32:40 HK: Sure. So the question is: How much do we try to control how the funds are used or how the the organizations we're funding behave? Let's say that we like an organization, we want to fund them, but we also have big disagreements with what they're doing. To what extent would we say, "Here, you can have the grant, do what you want", versus, "Here, you have the grant. But only if you do A, B, and C." This is an example of something that I would say, my first-

order answer again is we outsource this to the experts; to the program officers. 'Cause I think there's no one right answer here. So, there's no one right answer and that sometimes it's better to just let the grantee do what they're going to do, and reason that they know their work better than you do. And sometimes it's a good idea to recognize they have a weakness. That weakness has been spotted, and we need that weakness addressed in order to fund them.

33:28 HK: Then there's all kinds of in-between. A lot of time what we end up doing is, there might be an organization that has slightly different values or emphasis from us, and we try to make sure that certain people within the organization are in control of how the funds are spent, instead of it being the CEO, or we might try and make sure that the money is used for a certain cause. So, there might be an animal welfare organization that does work on endangered species and on farm animal welfare. We think, the farm animal welfare work is more important. That's our priority. We think it affects more animals, it affects them more deeply. There's more good to be done there. So, we would say, "Please use this money only for farm animal welfare." There have been all kinds of places on the spectrum. One thing we do is we try to encourage our program officers to take a little bit of a similar mentality to their organizations, as what Open Phil takes to the program officers, which is like, "When in doubt, trust. When in doubt..." You know, they probably know more about their work than you do.

34:22 HK: You should always be asking yourself, "Am I doing more harm than good by trying to be too controlling here?" We have tried to nudge program officers away from being too controlling, and tried to establish a strong default of freedom and deference and being skeptical of yourself, when you think you know how to do someone else's job. And that said, there are times, when a program officer makes a really convincing case that, "Here's a new, strappy organization. They've got a lot of energy, they've got some great ideas, they are not experienced to running an organization and they are making this rookie mistake and we are going to ask them not to make that mistake. We're going to ask them to do things differently." There's all kinds of situations like that. And I think it's a judgement call. Sometimes, you feel that you understand the situation well enough. And since our program officers are experts, a lot of times they really do understand the situation and can see that something needs to be done and a grantee needs to have some weakness address, and other time we need to just say, "We will register our opinion", but it's ultimately the grantee's call.

35:26 S?: I struggle to find political giving recommendations of any data behind them at all. Given that individual philanthropists are limited in how they much can give, have you considered starting an organization to give public political giving recommendations?

35:38 HK: Sure. Have we considered starting an organization that can give public political giving recommendations? It's not a priority. I think we definitely have our hands full. And GiveWell is generally is not political, and they are the ones that are more focused on disseminating information to help lots of individual donors. We are more in a mindset of we try to figure out the best thing to do, spend the money as well as we can, and then give people enough window into our work that they can see the basic way in which we're thinking and learn some things from us. But at the same time, if we're, for example, if we're working on Criminal Justice Reform, there might be something in the middle of a legislative session where if we put out too much detail in what we're thinking, you might have opponents who would go and adversarially use that.

36:24 HK: And so I think that intersection of trying to, for example, do advocacy and also trying to be totally open about every step of your thought process, I think there is a bit of a tension there. So that's, yeah, basically not our priority. And I think we have these two... Open Philanthropy, does do some work on advocacy, but would generally not be trying to blast out a recommendation to the whole world 'cause we think that often will do as much harm to the ends we have in mind as the recommendation could do good or more. So yeah, I think someone could potentially find a way to make that work. But it would be a different mission from either what GiveWell is doing or for what Open Phil's doing so we're not currently doing it.

37:08 HK: Yeah in the back, Jake.

37:10 S?: Can you give an update on how you're thinking about comparing across the different cause areas, and how you think about the value of diversification [37:21] \_\_\_\_\_?

37:21 HK: Yeah, sure. Can I give an update on how we prioritize across different cause areas. Maybe now, I'll go through my final little set of points about just what our priorities are from here and what we're trying to do this year. Obviously, a lot of what we work on is just the whole grant-making flow. And so trying to give away triple digit millions each year is always work and our program officers are totally busy with it and it's something that I work on as

well. But also, aside from that, in terms of organization-wide priorities, some of the biggest open items for us and some of the biggest things we're working on, one of them is what Jake just brought up. So we did pick causes that we thought were especially promising but what we didn't do yet is decide what is the relative priority of the different focus areas we have? In the end, when we reach peak giving, which we're not at yet, how much money should be going to each different kind of cause and how much money should be going to each cause. And I think that is a major project. It's been going on all year. I expect it to go on for a while longer. It is challenging because in a perfect ideal world, you would have some units for good accomplish, like lives helped or something, and you would compute them per dollar for each grant and you would compare and do the best stuff.

38:36 HK: But in reality, A] we often don't have a way of estimating that and we have to just look at more heuristic approaches like funding good people working on important issues. And two, there are these giant judgment calls. So you heard me mention the thing about animals. There are some people who feel that we should care about chickens, some fraction, as much as humans, maybe 10%. And those people might conclude that Farm Animal Welfare is a far better use of money than other causes. And there are other people who think we basically shouldn't care about chickens at all and it, for multiple reasons that we have written about and will soon write more about, we do not want to just take the entire organization in a direction where everything we're doing comes down to resolving that very kind of hazy question one way or the other. And so especially because we believe we see outstanding opportunities to do farm animal welfare work, and we also see outstanding opportunities to do other work. And so it would be a shame if we left half of those on the table because we just took our best guess on this one very hazy question that we could easily change our minds on every week or something.

39:40 HK: And there's other big picture judgement calls like that too. When you think about trying to prevent events that might derail civilization for future generations. If you were to make a contribution to preventing an event like that, do you want to count that as seven billion people helped, which is the world population, or do you want to count it as seven trillion, trillion people helped which is how many people could exist someday in the future, or do you want to count as a much bigger number than that. Those are some of the challenges we face, and I think in order to get a coherent story of how much money do we want to put into each philosophical position, on what counts is doing the most good, and then within those, how much money do we want to put into each focus area, I think in order to get that right, we need to grapple with the whole variety of issues, some of them relate to the animals versus humans question and touch on issues in philosophy of mind and animal behavior and whether animals feel pain and all kinds of things that we've written about. And we have to engage with also the philosophy of what you do when you're uncertain in this way and when you're making these tough choices.

40:52 HK: There's a lot of literatures to go through. A lot of discussions to be had. Think in the end, it'll end up looking something like the GiveWell cost effectiveness analysis where there's a lot of judgment calls, and the best we can do is lay them out very clearly. I have different staff members make them. Ask Cari and Dustin to make them, put them online so other people can challenge them, and then go with our judgments. And I do expect there will be an update this year that'll have some preliminary high level statements about how much money we expect to put into what kind giving, but I think it could easily be another year before there's something that's really tangible that says, "Okay, this much money and this cause. That much money and that cause." And it could be longer than that.

41:33 HK: So that's a major cross-organization project that we're working on. Other things that we're working on, we're still trying to build out better operations and so we're still doing a job search for a Director of Operations trying to get ourselves to the point where as an organization, we're more self-sufficient, not sharing staff with GiveWell anymore. That's a place we're trying to get. Another thing we're working on is self-evaluation. So it's just starting to be the case the we have major portfolios that have been around long enough that we're starting to see impact or could reasonably expect to start seeing impact, so we've just started working on these kind of portfolio reviews where we go back and try to say, "Okay, we made a bunch of grants." With those grants, we made a bunch of predictions. And we have actually tried to systemize, that we make as many predictions as we can when we make these grants. And then we look back and we say, "How did reality play out compared to our predictions? How much impact do we think we had? Are we getting a good deal for the money?" So developing that process is a major focus that we're working on right now. And the final thing that is a lot of the emphasis of what we're doing is kind of self-critique or questioning the assumptions behind various causes.

42:42 HK: So we work on Criminal Justice Reform, we believe there's too much incarceration in the US, but David Roodman, who does literature reviews for us, basically did a massive report where he went through all the literature he could find on, "If you reduce incarceration, will you increase crime and by how much, and how much should we value that?" And so he did this epic report that I really recommend, it's very interesting if you are interested in the kinds of

things that I'm interested in, which most people aren't, [laughter] like econometrics. But that's an example of going really deep trying to take the basic premise of our work, like we want to reduce incarceration and question it. And we have a bunch of other projects like that, that are either recently wrapped up or going on, so we did a somewhat intensive investigation of how good the case really is that getting chickens out of cages is good for the chickens, 'cause some people were arguing the opposite side, so we have a write-up on that.

43:39 HK: Thinking about how reasonable it really is to think that very powerful AI could happen within our lifetimes, that's something we're working on and in terms of trying to make more specific predictions, so we can tell in the coming years if things are evolving more or less compared to how we thought they would. And then also looking at like how bad we think a pandemic could get and what would be the worst pandemic, so that's another major thing that happens like across causes at Open Phil. So anyway, so those are some of the major projects and Jake the answer to your question is like there will be a small update soon, but it's a big undertaking we're not there yet.

[laughter]

44:17 HK: Yeah.

44:18 S?: In terms of AI safety, what do you think is the role of the private sector? So you talk a lot about trying to change direction of research, I guess both in academia and in industry, but [44:28] \_\_\_\_ so much money. Are you worried that Google or some company puts up even more money than you guys and say, "Let's do research on capability and not [44:36] \_\_\_\_"

44:38 HK: Sure. So what is the role of private sector in AI and I've mentioned kind of wanting there to be a major field of people who think about AI safety, not just capabilities, so how much should we worry that a company like Google could just do so much work on capabilities that it drowns out everything we're doing? And I think it's totally valid. I mean, I think it's increasingly the case that in the field of AI, a lot of where most people most want to work is industry, and a lot of the groups that are doing the best work is industry. And as a philanthropist, there's often not much we have to offer a company like Google and not much opportunity for us to engage with them.

45:19 HK: We do have a major ongoing financial commitment to OpenAI which is a nonprofit that is otherwise very similar to an industry lab. So that's at least a chance to sort of be involved in one case of that, but I would say the answer is, first, what's important to us is that there'd be a field of people who are very thoughtful about safety who can provide good advice to whoever needs that advice. And so even if there is a medium-sized field in academia and those people are helpful on safety, and eventually Google needs to go to them for advice, we'll feel that we made a big difference and did a good thing relative to if there's just no one to go to.

45:56 HK: And second, we think that the more these ideas gain currency, the more quality academic work is done on topics relevant to safety, and showing that there is work to be done and there are important conversations we had. We think that one could reasonably advocate to groups like Google to put more effort into these things and pay more attention to them and hire their own safety teams. And certainly Google DeepMind does have a safety team. And I think part of the reason for that has been the concerns that have been raised by a bunch of tiny non-profits, so I think it can happen. Tiny non-profits can affect the decisions made by big corporations and governments, we've seen it happen many times, and we think we've been part of it sometimes.

46:32 HK: Yeah.

46:33 Speaker 6: Hi, can I donate to Open Philanthropy based on specific categories, cause categories?

46:37 HK: Can you donate to Open Philanthropy for specific categories? Open Philanthropy does not take donations from individuals. We're just not set up to do that. You could check out the effective altruism... I think they're called the Effective Altruism Funds, and those are set up by category, those are set up by cause, so basically there are these funds that were set up by another organization. And some of those funds, basically you give to the fund, and then the fund manager decides what to do with the money. And some of those fund managers are Open Philanthropy staff, so for example Lewis, who works on Farm Animal Welfare and he's here, he is the fund manager for the Effective Altruism Fund for Animal Welfare, and so you could give there, but that is distinct from Open Philanthropy. That is like a fund that people are trying to leverage Lewis' expertise and there's a few of those so you can google that.

47:28 S?: Have there been non-trivial grants yet that you felt like are failures in the evaluation or that they're

meaningful negative lessons from?

47:38 HK: Yeah, have there been non-trivial grants yet that we think are failures that we have meaningful lessons from? Yeah I mean, definitely. I think a lot of the work we do is just so long-term that it can be too hard, it could be too early to assess it but there have been grants where we've either made the grant and said, "Hey, these people just don't seem to have made much progress and I'm not excited to continue supporting them."

48:00 HK: Or in some cases just really didn't accomplish what they set out to, so there's one program we tried that was trying to match Haitian farm workers with these visas that are unlimited in the US. And we were hoping that it would cause farm workers to be able to come to the US and make a lot of money for them, which they could take home to their families. And basically, the recruitment just wasn't where we hoped it would be and the return on investment for the grant wasn't where we hoped it would be. So, yeah. Certainly, that's happened and I think sometimes, we're going to write publicly about that when we think it'll be educational for people, and when we think it'll be interesting, when we think people will be wondering. And a lot of times, we're just going to keep that in-house and try to be very open about it as we discuss things internally because we do have, in that way, a little bit different orientation than GiveWell, where our attitude is sort of, we can do a lot of harm to our programs by putting too much information out in public and we can endanger our relationship with partners and grantees. And so, for us, what we want to do is be really clear internally and then share things externally when we feel the benefits of doing so outweigh the costs. Yep.

49:07 Speaker 7: Do you have cases where you've thought about a cause, and really think it's important, but then you go out and then there just aren't any organizations that you can give grants to that are going to do what you think needs to get done, and so you feel like there's a need there? Have you found those and what do you do or, have you thought about what you will do when that happens [49:27] \_\_\_\_?

49:29 HK: Sure. Are there causes where we think the cause is really important but we just can't find any organization that's doing what we think needs to be done? Yeah, there's lots of that and we handle it differently, a bit depending on how we've prioritized the cause. We've written a bit about science policy and infrastructure, and this is something where we've seen a set of issues and complaints about how the whole scientific research process is... How government funding for scientific research works and how the academic system works that may not be giving people the right incentives and the right support to do really high risk transformative science. And we tried, we wrote about what we wish someone would do and what we'd love to fund someone to do, and we made a bunch of calls saying, "We would love to fund this. Do you know anyone?" And we got nothing and we just haven't done anything. And someday, we might come back to that. We still keep an eye out, but that is definitely an example of that. I think on immigration reform, that's been another one where we've often just felt like there's lots of work we would fund and no one is doing it and we've kind of given up on trying to find it.

50:32 HK: So, that happens sometimes. Other causes that are sufficiently a focus area, like if we have a full-time staff around them and they're important enough to us, with a bunch of the current focus areas we work on, I mean, certainly with AI, there had been some things where we said, "Someone really needs to be doing X," and we'll just pound away at it. We'll say we'll write up why we think someone needs to be doing X and what X is, and then we'll just keep kind of almost headhunting to find the right person to do this. And this is something that philanthropists have done before, so the Sandler Foundation, I think we wrote a blog post about them and they're actually near here. They're like a 10-minute walk, and they've had some interesting successes doing that model, where they say, "Here's an organization we wish existed." They make a lot of phone calls, they ask if anyone knows anyone who could run it, and they kind of head hunt and they kind of pitch to the person, and they say, "Will you do this job, 'cause we will fund the heck out of it?" I think it is an interesting philanthropic model and it's one that sometimes we've failed at, sometimes we've succeeded at. Yep.

51:37 Speaker 8: I'm just curious, given your interaction with GiveWell, shared history and all of that, how much have you thought about the interaction with your priorities, what impact it would have on GiveWell as an organization. Specifically you say, you partially fund their Top Charities, but you've made this decision not to fund them up to their room for funding. Not having thought about it deeply, it's not clear to me if you did go up to their level of room for funding, what effect that would have on GiveWell? Would it be that now they've lost the reason for being... 'Cause it sounds like you have funds at your disposal but you can do that for at least three or four [52:12] \_\_\_\_ months. It's unclear whether that would be good or bad for GiveWell. Would it be that now you would unleash all this sort of productive energy that they have in their organization to do research in other organizations because they can say, "All right, the five most important charities are fully funded. Now, we're going to take all the staffing and building we have, and go look for other stuff," or would it be negative because now you have potential donors see, "Alright, this stuff's getting funded. I'm not going to go to GiveWell for donations anymore." Just curious if that's something that plays into

your decision making process at that levels of funding?

52:48 HK: Sure. How do we think about GiveWell and what effect we could have, positive or negative, by greatly increasing support to their Top Charities and how do we weigh? Would that help GiveWell? Would that hurt GiveWell? That's something we've thought about more in the past. A few years ago, we were in this kind of situation where we said, "Look, if we fully funded the Top Charities, just to be clear, that would not be the end of GiveWell. There's no way on Earth that would be end of GiveWell, that would... 'Cause then, you need GiveWell to keep making those recommendations and keep doing that research." And so, that is not the issue. What it might be the end of is other donors to GiveWell Top Charities. I think it would be quite rational and I think it would be kind of weird not to do this where if you as a donor saw us doing that, you might say, "I guess I'll never give to GiveWell Top Charities again because every dollar I give is just causing Open Phil to spend a dollar less." So, unless you want to donate to Open Phil, which some people do, but which probably most of these donors don't. I think it would be the end of that, and I think that is a bit of a cost and so I think that, for a time, we just, we didn't want to do anything irreversible.

53:52 HK: We didn't want to kill a potentially large and growing constituency and source of money for GiveWell when we weren't sure how they would fit into our priorities and how much money we were going to want to put there. And if in the end we had decided GiveWell Top Charities are the best opportunities we can find, they're better than anything else we can do. I think we eventually would've just swallowed that and said, "We gotta make sure those are fully funded. We can't leave that on the table." And on the other hand we're probably ending up going in a different direction. So I think what we're largely saying is there are, at least under certain assumptions and under certain world views, there are things you can do that if you're willing to take a bit of a philosophical leap, and willing to take more of various kinds of risk, that you can do a lot more good per dollar than GiveWell's Top Charities. And we want to reserve the vast majority of what we're doing for those what we consider potentially more promising opportunities.

54:46 HK: Don't want to say they're categorically better giving opportunities. I think you have to swallow something to get that extra impact. You have to swallow something, for example the idea of helping chickens instead of humans. Or the idea of counting future generations in a certain way. And you have to swallow other things about how confident can you be in a case and how much trust you need to have.

55:07 HK: But we basically do feel that we want to reserve a lot of the capital for giving opportunities that we think are a lot better under those assumptions. And then we think we want to reserve still a chunk of the capital for the worldview that says, "Okay, GiveWell Top Charities are the way to go and we're not comfortable with these other assumptions you have to make." And so, we're reserving some. And so, we don't know what the exact numbers are going to be yet. And when we put out the blog post, which will be sometime in December, it will hopefully give a little bit more detail on the thinking and on the quantification. But the bottom line is that we think they're excellent giving opportunities and we want there to be a substantial allocation to them but we don't want that allocation to be most of the capital. And so at this point we're no longer... It's not really top of mind for us anymore, this idea about how that would affect GiveWell. It's more like we're just reserving that capital for other things and that's why we're not fully funding. I'll take one or two more. Yep.

56:03 Speaker 9: Presumably there are some focus areas you haven't even [56:05] \_\_\_\_ imagined yet. What is your [56:10] \_\_\_\_ focus areas you haven't considered? How do you estimate the value of those focus areas when you're trying to kind of make that exploratory trade-off [56:15] \_\_\_\_ launch versus discovering new ones.

56:17 HK: Sure. The exploratory trade-off, how do we think about focus areas we haven't thought of and what do we do about those? It's a good question, and certainly, I think there are probably focus areas we haven't thought of yet. For the most part, what we try to do is we try and take ideas that are out there and find the ones that are most promising to work on. There's whole institutions that employ intellectuals, that try to have new ideas. We don't really think of that as our main job. We try to think of our job as like, "Take the ideas that are out there, find the best ones, fund them." We do hope that by establishing this pattern where we fund very important neglected, tractable things, we're making it more clear and more exciting to people, this idea that if you have an idea for a new focus area that we're missing and you make that case that good things will happen as a result of it. I think there are cases of people putting ideas out there that we wouldn't have thought of, and now we work on those areas and we're putting tens of millions of dollars into them. I think that's one thing is we're creating that incentive.

57:18 HK: We also fund some of the groups that we think are most likely to come up with the next big focus area we're missing. We fund groups that try to build the effective altruism community. And the effective altruism community has historically been the people who think the hardest and the most about what are the causes where we might be able to do

a ridiculous amount of good, and have come up with some of these focus areas or have helped highlight them to us. We fund some groups in that area including Future of Humanity Institute, which is basically, we're funding people to think about both existing major priorities and how to do better with them and what major priorities might not have been thought of yet. So one of FHI's main goals is to think of unknown unknown causes, new causes. Those are some things we do. I can imagine in the future we might do more intense things, we might announce some prize or something. But at this point I think there are a lot of ideas out there. I think the ones that we're focused on are good ones to focus on. And to me exploit looks pretty good and I think the amount of exploring we're doing is pretty appropriate and we might flip that equation in the future.

58:23 HK: I'll see it's 9:06, so I think I'm actually just going to call it here but I will stick around for questions for 20 minutes or so. Thanks very much for coming everyone, I appreciate it.