**Transcript of 06/08/12 conversation between**

- **Donald P. Green, Professor of Political Science at Columbia University**
- **Holden Karnofsky, Co-Executive Director, GiveWell**
- **Stephanie Wykstra, Research Analyst, GiveWell**
- **A donor interested in this topic**

**Donald:** In my domain - campaign craft - proprietary science is a big issue. I personally have to distance myself from that. I'm happy to do temporarily proprietary science because people don't want their findings released while they're still in their current election cycle. But every time I enter into that kind of research relationship, it's with the understanding that the results will eventually be known.

Ironically, many of my experiences in the context of writing the textbook you just saw are somewhat unhappy ones about data sharing and transparency. Even when people have been funded by federal organizations like the NIH or the NSF, they don't always hand over the data when asked to share a replication dataset. So I'm certainly 100% on board with this general idea of improving science by improving information flow. Once I got my head in the right place doing replication archiving, this led to the revelation that you could often catch errors if you simply had a team of dedicated people replicating things that people were about to put into print. One very easy to grab piece of low hanging fruit is simply to fund prospective replications.

**Holden:** What exactly do you mean 'prospective replications'?

**Donald:** In others words, someone's writing this paper, he's about done, maybe his research funds have run out. But the paper has been accepted at a wonderful journal, he's very pleased about it, and off the paper goes. The journal doesn't have any requirements that you deposit the data with them. And so the data stays on the researcher's computer. Unfortunately the computer's stolen, the backups are gone; that's just too bad for the world. Or maybe the researcher did 500 analyses and only reported 3, and the world will never know about the other 497.

So, whether there are good motives, bad motives, or no motives involved, the sad fact is that pretty much every time I asked for a replication for a data set older than four years I struck out. And it wasn't because people were unwilling to collaborate, it was just - they couldn't get it together.

Without prospective replications you're really just hoping that memories have not faded to the point that the data are unrecoverable. So – this very, very superficial level where we're just talking about the studies that have been done, datasets are disappearing at a rate that's unbelievable. Now in this current moment where all eyes are on websites like filedrawer.org where the focus is increasingly on the institutional failures associated with peer review and replication, all the more reason to start taking action to promote prospective replications, especially in behavioral science.

If you take note of the fact that very often people's excuse for not doing this is that it's costly, the whole thing could be centralized so that if you were to add some programmers, even high level programmers, you could simply take someone's article or manuscript, take the data, and reproduce the tables and figures. That would be a quantum leap for the social sciences.

**Holden:** What's filedrawer.org?

**Donald:** It's a very interesting website that is taking note of the fact that many hard to believe findings are not replicated - not that they aren't replicable, but they aren't replicated. They're essentially creating a knowledge market, it's almost like a crowd sourcing activity. If you pool knowledge about what, from a Bayesian standpoint, seems implausible, then that targets your efforts in terms of revisiting results that are suspected to be false positives.

I think that there's an amazing bifurcation in social psychology right now between the old guard, who if anything have become more entrenched in their determination not to share their data, and the new generation saying, "this is leading to all kinds of embarrassing scandals." And it's only a matter of time before that happens in other fields. In fact it has happened in other fields. Even in economics, there are some embarrassing instances where data disappeared or something like that. So prospective archiving means that data go into a centralized source and there is a routinized institution that handles it. Currently in economics, not so much in political science, there are *ad hoc* little replication archives.

You want to incentivize people at least insofar as they don't have to pay the incremental cost of archiving.

**Holden:** So you're saying to have a dedicated tech support team that takes care of some of the work.

**Donald:** Yeah, they look at every issue of every journal, every one comes with a list of forthcoming articles and so they simply download that and contact the authors – or hopefully the author contacts them – and say "I see that you're the author of this forthcoming article and I'm here to help you deposit your data." You don't have to share anything that's confidential, you don't need to share data that weren't published although that would be a good thing, but anything that reproduces the numbers in your table or anything in the text – it should be totally automated. That's what we try to do.

**Donor:** There are two types of disincentives from sharing data. One is that it's kind of costly in terms of your effort and your time and the other is that it could lead to embarrassment; there are small mistakes everywhere. Trying to make headway on this topic, a lot of what we get back from our grantees is "help us build a ISPS type archive, make it easier for us to deposit data" and that's fair enough, but my concern is that even if you reduce that cost to literally zero there would still be this issue that people actually don't want their data out there.

**Donald:** I think that what happens is that as the cost goes to zero, the kind of excuses that I got when I tried to get people's data just become untenable. You literally have to say, "Don, I'm not sharing my data because I'm afraid that you're going to embarrass me." No one's going to say that.

**Holden:** You could keep a list of people who did.

**Donald:** Or maybe the flipside is – if your data are deposited in the archive, so much the better for you. That's why I do it; I want to be more credible. That's the whole thing about the credibility revolution. Our results should be – as the philosophers would say – intersubjectively verifiable. The idea here is that youcan't completely expunge this embarrassment incentive, but you can overshadow this disincentive with positive incentives. If someone's willing to archive prospectively for me and I'm worried that there's a mistake – okay, fine, I'm sending my data to you, if there's a mistake send it back and we'll correct it before the page proofs are finalized.

There are inevitably mistakes, but the whole point is whether there are administrative errors that are potentially correctable or whether there are fundamental conceptual errors. But the latter are not to be corrected here; those are to be corrected in the fullness of time on review.

**Donor:** When you were working at ISPS you were building this archive; did you also have a policy in place basically saying "you have to use it"?

**Donald:** Yes.

**Donor:** And every ISPS affiliate, regardless of where he or she got funding?

**Donald:** Everybody.

And the only people who didn't do it were people with very, very special data, data provided by the military, basically data that were promised to never be in anything but the highest security. But they're archived, they're just not in a public archive. Fine. I just don't want things to disappear, for a lot of reasons.

That was the policy. It was different from Dataverse, because Dataverse is like a big unstructured dump. And it's not that the data aren't there, it's that who knows what they are, because there's not always accompanying metadata. So the idea was to have both. Obviously producing metadata is the pain in the neck; that's why you have to have support.

**Holden:** And you had support at ISPS.

**Donald:** Yes, we had a dedicated full time person.

**Holden**: Where did the funding for that come from?

**Donald:** It was from ISPS, its own coffers. It was one of my highest priorities.

**Holden:** What is ISPS?

**Donald:** It is an interdisciplinary research institution founded in the late 60's to facilitate cross-disciplinary collaboration, but when I became director in 1996 I was basically the only faculty member left, it had kind of run down into nothingness. What I tried to build there was a lively team of researchers who would be identification-oriented. They have special expertise in experiments or in regression discontinuity analysis or in the kinds of things that are credible these days and the team of people became very, very prominent. I served five terms. In my last term, we had more American Political Science Review articles just from ISPS than any department on earth including the rest of the Yale department. It was a research factory. And the idea of having a research factory is that you've got to preserve your outputs.

**Stephanie:** So is this archive accessible?

**Donald:** You can look at theISPS archive and you see all of the data sets, all of the metadata and all the articles themselves.

**Stephanie:** This is data just from the articles, or all of the data they had?

**Donald:** It really depends. I only required that they archived the data that they were publishing because I don't want to put people under the gun. Because many people want to publish four or five things; I wanted it to be their call. So the idea was, at the end of this exercise you were going to press this button both in Stata and in R so that there would be freeware for anybody in the world that could reproduce all of the numbers.

**Donor:** If you say that this builds credibility and is helpful for a scientist, not just the pioneer who started it, but for the people who say "my paper will look better if it's there" – what was the barrier to other political scientists saying "I'll do the same thing" or "I'll get in touch with ISPS and see if I can get permission to publish my data there"?

**Donald:** We didn't provide support to others because we were maxed out with the resources that we had.

**Donor:** Did you have demand for it?

**Donald:** People would say "could this be a general experiments archive?" and we only had one person with some graduate students. It was overwhelming. But at that point we thought, “Look, it’s going to catch on.” It is catching on. It’s only a matter of time. The problem right now is that it’s very decentralized.

**Donor:** What I’m hearing is “what we did at ISPS should be bigger and should be available to others." But in order to avoid it becoming so that you don’t know where it starts, you don’t know where it ends, and it kind of becomes this state of gunk where you don’t have consistency anymore, you don’t know what the universe of things is that’s in there, what are the right boundaries to draw?

**Donald:** It would be randomized experiments. Maybe I’d be willing to allow regression discontinuities as well. But there’s such a radical difference between what we mean by replication in the context of an observational study and what we mean in the context of a randomized experiment. Because a randomized experiment starts literally with the code that tells you how the units were assigned. Whereas with an observational study – I know that if I replicate your work it’ll be your opinion vs. mine. I’ve almost never seen a data set that was unbreakable from the point of view of causal inference, whereas if you analyze a large randomized experiment you can analyze it upside down and backwards and it’s not going to change very much.

So the idea of replication is especially important for small studies because they’re more vulnerable, they’re less robust. But it’s useful for all studies essentially because the new generation of statistical work that uses machine learning is using techniques that were not often available to the author when he or she wrote the paper. So for example, when you pool together of a lot of my type of work on getting out the vote you can do data mining to see the conditions under which the treatments are more or less effective. That’s what happens in medical research as well. This is taking the data set and going beyond it.

**Donor:** My concern is that if there’s a real disincentive to share data, because it could potentially lead to errors' being uncovered, and so there's a real public goods problem here – I’m concerned that we put all of this money into building this efficiency and people only upload their data when they feel like it.

**Donald:** I think that the answer to that is straightforward. This is a case of regulation. I think that everybody has an incentive to be in a regulated environment, particularly if they’re both consumers and producers of research. Because I want to know before I start a research project what’s real so I that I don’t waste time chasing red herrings. That said, a funder can do two things. One is to say, “It’s going to be option like everything in life including violating regulations, but henceforth no grantee will be exempt from this policy.”

**Donor:** The problem is that some of these institutions get funding from a lot of different places. That’s the reason that we trot around the registration story, to get all of the funders at one table and speak with one voice. On the data story we’ve unfortunately faced a more difficult battle. Some people say, “We’re not sure that’s a fight we want to fight.” So we don’t have a critical mass of people who will say “we’ll make it a policy.”

**Donald:** I think that almost certainly, in 25 years, people will say, "yes, of course, we share our data." The same thing that happened in medicine under Consort will happen in political science.

It used to be that everybody would resist the Consort standards of reporting, which involve careful disclosure of how your experiment was conducted. People said it will never happen. And now everybody does it. Why? Because if you didn't do it, reviewers will say "why didn't you do it?" In much the same way, if you said that one part of reviewing is pressing this replication button, people would say "yes, it's so obvious that that's what you do, why wouldn't you do it?"

**Donor:** But when you say pressing that replication button, in all honesty, you can press the button all of the times that you want, you'll always get the same result. You need to dig in and actually analyze it, you have to analyze those command files and do a very different review from what people are actually doing.

**Donald:** Maybe, but I think that if somebody took the time – the archiving person sitting there with the article going line by line – in my experience that is what expunges the errors. It doesn't necessarily expunge the broader conceptual errors.

**Holden:** So I have an alternative way of approaching this, which I've thought of while we were talking. Which is instead of funding an existing research center to create this thing and hoping that norms evolve, what if you just had a group which that's watching one journal and every time somebody submitted to the journal they said "hey, send us you're data and we'll deal with it and we'll metadata it and archive it" and if the person did, great, and if they didn't, they publish their name on a list of everybody who they requested from and whether they said yes or no. It's an independent group. You don't have to work with an existing research center, it doesn't matter who you work with, it's just a bunch of technicians, all they do is email people and archive data and maintain a list of who said yes and who said no.

One of the things that I would find appealing about this is that I think you run into trouble when you have two grantees and it's up to them what they do. Once you have a group of technicians you can say "we don't have to work with a particular grantee," you have a totally different kind of leverage.

**Donald:** One thing I like about that, I didn't think about it until you mentioned it, is that it brings in a totally different dynamic. If you had a group of outside people saying "I'm constructing this list" – they're totally outside of the norms and mutual sanctions of the discipline so they can't be bought off; they're not graduate students of the people involved. So if you're name is going to be on a dirty list of people that don't share data, you have no one to push around to get your name off that list.

Right now if I were to do this for example, I would have to deal with the blow-back from it, not that I couldn't, but it would be a nuisance, but if it's an outside organization it's kind of a *fait accompli*.

**Holden:** And you could even maybe do something similar with registration. It would be harder because you'd have to get the studies at the point of funding instead of point of publication, but it would be the same thing. You don't have to get a particular foundation to work with you. Once they fund someone, and they publish the grant, you get in touch with the researchers.

**Donor:** The kind of name-and-shame where the default is shame …

**Holden:** You make it easy to not be shamed, that's the point. You say, "here's the list of people who we emailed, and we asked them for 5 minutes to send their data and here's the list of people who said yes and here's the list of people who said no." You make easy for them and I think you have to combine the two.

**Donald:** You could even have a middle position, which is that they've said yes and that they're going to hand in the data by this date.

**Holden:** Sure, make it as easy as humanly possible.

**Donald:** I like that because it makes it almost bulletproof. What are you going to say to that, aside from of "stop harassing me" – all right, then you're on the list.

**Donor:** You suggest drawing that around a certain journal?

**Holden:** I don't think it matters too much what you draw it around. I guess you don't want to take on an entire field at once because you don't have to, because you can experiment on a smaller scale. So yeah, I guess I would go with the journal because it's what's easiest, but I don't know.

**Donor:** That's a question that I wonder about. I'm sure there's no harm in trying to make it easy. There is this other way that could complement it on a standalone basis, which is to overcome the public good problem by making it costly not to share, basically the type of thing that you suggest.

So I'm wondering down the line, if we envision a world in which data is shared differently, not just so much my grantee shares their stuff, but we envision a real systemic change, people rethinking this. I have this concern that to subsidize this kind of thing could be really costly while not support a culture outside.

**Holden:** I think another thing you could do is to back out over time. And I think you would leave the system in place.

**Donor:** But it needs to be funded, this is basically research assistants. I'm wondering, is there a way that we could start with things that don't involve a continuous subsidy?

**Holden:** But I don't think it's continuous, I think it's the kind of subsidy that you could get away with pulling out. Let's say you get it to the point where everybody is submitting their data. So all data is published. Now what you do is pull out from one journal. And you keep doing all the other ones. So now there's this situation where if you publish in this one journal your study's not as credible. So someone, the professor or the journal has to deal with that situation and evolve his or her own way of paying for this stuff.

**Donor:** So I guess that's a question. Do you think it makes sense to start this way? Do you think it makes sense to start with a large project, an ISPS type thing? What you're envisioning is not so much a change among our grantees and in our narrow field but really a change in culture way beyond what we'll be funding in the future.

**Donald:** Like anything, there's a risk. And I think the key thing in doing it is you've got to partner with something akin to a permanent institution. I think it's even questionable whether you would fund an organization as opposed to a university -- or even endow a permanent institution. ISPS is a permanent institution, not that I'm saying that you should fund it in particular, but that's the kind of thing that's permanent as opposed to a group, an organization, a consortium that's newly formed.

**Holden:** That's a good point. That's kind of the model in developing world aid. A lot of times it doesn't work. I think occasionally it works. And here there'd be a plausible mechanism for it to work. You have to have a plan for getting in and out. But if you do it asynchronously... One thing you could do is say, "Here's our order, we're going to go in here first and out of here first." And so the journal knows that the expectation is that eventually it's going to be without support for this. And you pull out in the same order that you go in.

**Donald:** I think that the time horizon should be long though. It's the kind of thing where people are putting their prized possessions in this archive. If they think that one day the whole thing is going to go dark ...

**Holden:** Well, you can commit to maintaining the website. We're talking about the tech support. That I think would be the more expensive thing.

**Donald:** It would be a ten-year horizon, because you would have to think about the tenure life of a junior faculty member.

**Donor:** How much did ISPS go through?

**Donald:** A budget of between $100,000 to $150,000 per year.

**Holden:** So what's the scale here?

**Donor:** If you were to work for an organization that's five times larger would it cost five times as much?

**Donald:** The main staff person does not have to be replicated five times because part of what he or she is doing is setting the policies, developing the broader structure, getting people on board. What you would ideally do is have more and more of those statistician types, programmer types, who are basically turning datasets into reproducible information, creating the metadata. So it's like a factory. The reason I say ideally is because you'd have faculty associated with the project whose role is to make sure that there's a smooth flow. They could get some credit for it because the archive could have their name on it.

**Donor:** How many studies got archived by ISPS per year? By the end?

**Donald:** I don't know the actual number, I think it was in the dozens. You can go there and scroll through them, study after study after study and there are lots of them. And I know that they're pretty much booked up over the summer just archiving things. But I think that part of their activity level is occasioned by the fact that they were archiving retrospectively too. So one way to think about this is you do everything prospectively at a given point in time and then as the workflow becomes uneven you move retrospectively.

**Donor:** Could you explain where this line is that you draw between prospective replications and retrospective?

**Donald:** Prospective is for things that are still in press that have not yet appeared. There the appeal is that you can catch people's errors. That's a big benefit. That is something that would be routinized at the point where the paper goes to the copy editor. It also goes to the replicator who creates the metadata.

Once you have a free moment doing prospective work you can start to work backwards and say "okay, the previous issue appeared without archiving, let's go back and look at the previous issue–"

**Holden:** So you're saying that in addition to prospectively archiving your staff was retrospectively archiving. And retrospectively is probably more expensive.

**Donald:** It is. It's much harder, because people have to find stuff. And there weren't clear norms about what you would preserve. When we started out, we were really ham-handed about it. Things have gradually gotten better.

**Stephanie:** When you're talking about finding an organization to do it, maybe a university, do you anticipate a kind of problem with neutrality?

**Donald:** It is an issue insofar as you want to make sure you don't get mixed up in the politics of tenure or promotion. I don't think that that should be what's going on, and that's why having a university per se do it is probably not the right thing. Probably better to put it in the hands of a group that sees its fiduciary role as permanence, preservation, archiving.

**Holden:** I would probably do both. You could start with a totally independent group, that's the norm, that's the standard, that's how it's done, and then you can fund a university and if they're not doing a good job you have a benchmark right there. They're going to do things the right way and people can't yell at them asking, "why are you asking for my data?" because they can say "I'm just doing this thing that this other group is doing." And you can slowly dissolve the original group.

**Donald:** Also, universities charge overhead. This is how they operate. One of the reasons we could do this so efficiently back at ISPS is that this is just a direct transfer of endowment funds into outputs with no tax on it.

**Holden:** They charge a lot of overhead, right?

**Donald:** It depends, as the grantor you set the terms, you can say we're not paying 63%; we'll pay 10%.

**Holden:** True, fair enough.

**Donald:** But going back to this broader question of incentives you raised at the outset, how can you overcome this reticence about sharing? And I think that there are a few different strands to it and each one has a different remedy.

One strand of it is "this is costly and time consuming and I have too much to do." Trying to reduce the costliness is one obstacle. And another is "I'm reticent because I don't want to be embarrassed." And there at some level that person just has to suck it up. I'm sorry, but you're throwing this work out there and you have to stand by it.

I think one of the things it's so important to recognize about the whole credibility revolution is that you can't even understand social science without the idea that almost all of it operates under the assumption that the truth will never be revealed.

What's different about experimental research, especially on a grand scale, is that for the first time you have the answer. And you can ask whether you got the right answer because it's a reproducible process. Some might say that it's a boring or mechanical thing but it's real. It'll take a long time, but increasingly the focus is on the design and on the analysis of the results.

I think that the reason people fear embarrassment is because of overselling. That is not one of those incentives that we're likely to deal with in the short-run, that's a long-term culture change kind of thing. When you go to graduate school nowadays you talk to some professors my age and it's all about the story, it's all about having a good hook and engaging the reader, which is the entertaining side of social science. And then there's the science side of social science, which says, "Don't oversell." It's not about propagandizing your research. It's about being cautious and admitting what you don't know and emphasizing the science in what you do. Those two philosophies are at war. Nothing that you're going to do in the short run is going to influence that. But in the long run…

**Donor:** That brings us to a broader discussion, what you call this "credibility revolution," a term that I hadn't heard before. I'm wondering what other aspects are there to this. We've been talking a lot about registration and about replicability of data. From a very bird's eye perspective is this where you see the biggest gains for people like us who work in philanthropy to get involved in social science? Or do you think there may be other ideas, other beefs that you have with the science, other things that may be unrelated to what we've just been discussing that you think are problems that ought to be addressed and that could be addressed?

**Donald:** Sure, that's a very broad question, a very engaging one. One of the institutional failings of social science is the lack of planning documents. Another is that reviews are often done on the sexiness of the findings, not the persuasiveness of the design. Those are things that funders can work on and so can journal publishers.

If we step back and ask what kinds of things facilitate more credible social science, increasingly the issue comes down to training. You want to train people in a way that emphasizes the science part of social science. That is not what's currently being done. It's not as though I want to eradicate the other part; I'm not at war with it. I'm just saying that there are vast swaths of intellectual inquiry that are regarded as too boring, too pedestrian for academic research.

One of the most interesting examples of that is development economics. You would have thought 20 years ago that it would be too uninteresting to go to schools and do a deworming experiment. Leave that to the public health people. In some sense the genius of the Michael Kremers was to say, "no, we're not going to retreat from these issues just because they're not as interesting as the grand theoretical issues that capture the imagination of academics."

What I would say is if you could create research institutions in the developing world and train a cadre of potential students in that style you could recruit these very people to do indigenous evidence based policy work, administrative work. Their expertise and capacity all comes from the outside. If you train them from the inside and talk to them about the sort of research methods that enable you to facilitate cost-effective interventions, you make the countries better off and you can facilitate a style of social science that is not governed by entertainment.

**Holden:** Why do that in the developing world? To me the concern is that if you tried to build research capacity in the developing world they would converge to the same kind of games that academia in the developed world has converged to and if anything what you want to do is go in the US and take the next tier down, researchers who aren't quite happy with the existing set-up, aren't quite making it and getting them to do work that's really useful and not as sexy.

**Donald:** You could do that; it's just that the status hierarchy in the developed world is already established. The people at the top are the people who don't muddy their hands with data. I remember encountering this in Berkeley. The very best econometricians would announce that "the very best of us, we never work with data." These were brilliant people; these were not idiots. But that's the status hierarchy, where the more grand your theoretical vision, the more it's a nuisance to have to deal with facts. In the developing world you could imagine saying "look, we're training you in a two or three year program and it's going to be like a public administration program, and you're going to be taught methods for evaluating what works and figuring out clever strategies and learn to read the literature out there so that you can think for yourself." Whether it's in Africa, in South America, or South Asia, if you ask me what can be done to produce a sea change, take a lot of people with technical talent and brainpower and put them in positions where they can be in charge of resources.

**Holden:** Another argument for the developed world, I guess I'll keep pulling these out, I kind of have this idealistic vision of a world in which status is usefulness and that the most prestigious thing you can do is the most useful thing, the most useful leveraged thing. To some extent you could argue that there are parts of Silicon Valley culture that feel this way: where you're considered to be good if you do a good job, create value, you're considered to be not good if you run around with grand visions and never get anything done, but you're excellent if you innovate in a way that may seem simple and silly if you describe it but where the actual impact on the world is enormous. Those are the highest status people. Why can't academia work like that? And the more I learn, the more I think that my vision is not so unrealistic because I think that external funders have power to control what is considered high status in academia because academia cares a lot about funding and funding can make a big difference to status.

**Donor:** That kind of raises the question, in academia it's all about publication and tenure, those are the words that I always hear about what drives people and what gets them to publish, and this whole story about making interesting points and telling your stories and laying the hook. But on the question of how the university appreciates your scientific contributions and how you make a career as a consequence, I keep getting reminded of what you said about how the fact that journals these days determine what you're worth is kind of silly. There are many ways in which an academic could contribute to good knowledge in the world. And yet it's all about how highly ranked are your publications and how many you have. How entrenched is this?

**Donald:** Academics is a funny business. For the people who are at the top of the top echelon, it's pretty clear what animates them, and that's the desire for immortality. They're immortal through their works. Not everybody has that grand ambition, but that's a lot of it.

I think that it is true that junior faculty who are in the hunt for tenure are very concerned about the instrumentality of what will get them tenure, but if you sit back and ask them "what are you doing?" it has to do with making an enduring mark on students and readers. They think of their students as the people through whom they propagate their ideas, and their readers are in the same category. When their book goes out of print they are deeply pained. It is true that some people are just counting their Google citations or some such silly thing. But the reason why big ideas are often such a high status thing is because we as academics often tell the story of our intellectual achievements by reference to the big ideas.

One of the philosophical disputes that I continually have with my colleagues is that they will often put tremendous stock on asking interesting questions whereas I feel that it's easy to ask interesting questions and it's hard to answer them.

**Holden:** It seems like a funder could have an influence there.

One of the things that I've been thinking about is– you'd have to do it right, you can't just pay people – but could you have an award that was given out in such a way that people would consider it an intellectual achievement to get the award, which means that the right people have to be giving the award and it has to be for the right reasons. The Nobel Prize is not about the money, there's a list of Nobel Prize winners, everyone knows who's on the list; that is immortality. Could you create another award like that? And I think that in some ways everything works in our favor, because what do we want the award to be for, we want it to be for the greatest contribution to human knowledge that helps humans flourish. And you could call it that, and that's a pretty good place to start.

**Donald:** The awards that are currently given, at least in my field of political science, tend to be for the big ideas. If you look at the awards that are given in Japan or in Sweden, they're lifetime achievement awards for people who are working with big ideas. It's a matter of crafting it right. I would guess that there are some people who are going to be in line for both awards, both for the big idea and the greatest contribution to knowledge. But then there are going to be people who are going to be in one category and excluded from the other.

**Donor:** We had a list of thoughts and ideas that we had heard from different people; we could get your thoughts on these ideas.

**Donald:** That sounds like a good idea.

**Holden:** I had a couple of other questions first if that's okay. One thing that I wonder, you say that there's a credibility revolution, is it happening and how fast is it happening is an important question.

If there were nothing at all going on we would probably throw up our hands and give up. That's not the case; there are things that are happening. If I thought that five years from now or ten years from now people are all going to be using best practices I would also not be so interested. My impression is that it's somewhere in between.

I feel like for a long, long time you've been doing the kind of work that we wish everyone was doing. And it hasn't been ignored and it hasn't totally not caught on, but it hasn't caught on at the pace that we would need it to catch on for it to be universal in a decade. And I think that with funding, you could make things like data sharing and registration universal in perhaps a decade. If the conceptual control group is that it's 30 years then I'm happy to do it, if the control group is that it's 12 years then I don't want to do it. So how can we answer this question?

**Donald:** There's clearly a change afoot, you can see it any time you go into seminars nowadays. When I was a graduate student if the speaker would come and you were to raise your hand and say "I think there might be some kind of problem; I'm not sure I can safely infer causality from the data that you were showing me" everyone would turn around and say "Okay Mr. Smarty-Pants, what reason to you have to think that the speaker is wrong?" Whereas nowadays, if the speaker comes in without a well thought out identification strategy, within five minutes the first hand that pops up is "what's your identification strategy?" The burden of proof is squarely on the shoulders of the presenter to make the case. That's a sign of how the thinking about causality has changed.

One of the reasons to write a textbook is that even though experimentation is a teachable method, it's not being taught. There are very few classes, even in graduate school, on experimental research methods. So we wrote the field experiment book with the hope that if you build it they will come. We wrote the textbook from which people could teach, and we distributed lecture notes and we had free R code and the whole nine yards so that people all over could teach this class that doesn't exist. I would say even 15 years into this experimental journey, some of the basic ingredients of propagating this way of thinking have yet to take root.

**Stephanie:** When did the book come out?

**Donald:** Three days ago. We've been working on this book for a long, long time.

**Holden:** My question would be if the people at this table all decide to drop this and start delivering bed nets instead, when will registration and open data be near universal?

**Donald:** I'm guessing twenty, twenty-five years. A generation from now. About the time I retire.

In part, they see registration as a kind of trump card that is not easy for an observational researcher to play. Typically you've seen the data, you've seen the time series and now you're going to muck around with it.

**Holden:** There are times when you can play that card, when the data's not available yet, there's going to be this data set.

**Donald:** There are people in medical research, epidemiology that have called for registration of observational trials precisely because they need it the most.

**Holden:** In some ways I feel like the biggest potential change is exactly on that side. The biggest gains from registration are for the pure observational studies.

**Holden:** We're interested in meta-research for all fields, not just development economics. I think [Donor] is interested in meta-research for development economics targeted at the poorest people in the developing world. We want to know what are the problems in all of the different fields of science, anything that could be useful. There may be some fields that are never useful; we don't care about fixing them. But medicine, biology, even – I had a friend who gave me some ideas for computer science. How do you make computer science different? That could make the world better if we did it. Psychology. How much have you thought about these other fields and how would you start… We don't know where to start, and so I'm trying to get a bird's eye view of how much money is in each and where the money is coming from. I can't even find that.

**Donald:** It's a mind-bogglingly large topic. I am, of course, not in the world of development economics. I tend to do research on campaigns and elections, on crime, on education, on the mass media. Many of their problems are versions of the problems we've talked about. In some ways they're worse. And then I think that the whole idea of setting up new institutions that make for openness and sharing, you know, those institutions don't exist in those fields. It's even true in medicine. I think in the New England Journal of Medicine this guy said "I'm going to try to get ten people to give me their data" and I think he got a promise from one that never materialized. That can't be good.

**Holden:** The salaries in your field, they're paid by universities?

**Donald:** Yes.

**Holden:** But universities don't give you any other funding so you need to work with a foundation…

**Donald:** Well, in the case of ISPS it set up a field experiment fund of $75,000 for each experiment and these were all direct costs, nothing for salary, there was no fluff in our budget. That funded a whole lot of experiments including Leonard Wantchekon's experiment in Benin, a lot of famous experiments. Research funding sometimes comes from universities but I think for the most part extramural funding is what drives research.

Taking a step back – and this is something to bring back to the foundation communities, I think – foundations are reticent about funding what I would think of as basic research. They don't want to do basic research. Very often the field experimenters get called in to do randomized evaluations of programs. I think that this is fine in terms of a practical exercise; I use this to train many of my colleagues in how to do experiments. Program evaluation is a good way to cut your teeth on this kind of exercise. But a better model is urging academics to come up with especially telling tests of different theoretical ideas.

**Donor:** Do you have knowledge in the field of development economics, are things better there than in the field of political science?

**Donald:** I think in political science we're a little bit better. Very often in the World Bank world they have to evaluate the things that come from the groups themselves. We've done that of course, but the most interesting experiments often occur when we go to a group and they don't quite know what they want to do. They say, "what do you think we should do?" and we say "very glad you asked." We have lots of hypotheses about different conditions that they could test and those are often watershed events because they merge social psychology theory with political behavior. There are some such studies in microfinance of course.

**Donor:** My impression is that J-PAL doesn't run into difficulty asking the questions that they want to ask.

**Donald:** They're in good shape. They are a very elite organization so I'm sure that's not true for the field, but then the question is how many interesting important fundamental questions are there…

**Holden:** This is what I'd like to ask, what are the studies that you would either like to do or see that you haven't been able to get funding for?

**Donald:** The studies in my area that are out of bounds for me are studies that involve partisan behavior. Those tend to be the province of the Analyst Institute or the internal workings of campaigns on the left and right. That's one of the biggest constraints on me.

**Holden:** You can't study the internal workings of campaigns you're saying.

**Donald:** No, I can, but it's not really in my academic guise, not in my professor role, because these studies don't involve 501(c)3 money, they're just done in the world. They're fascinating studies but they're in that murky world … essentially university quality research is being done outside of universities. In my own world, I don't have a lot of constraints; I'm in pretty good shape.

The studies that I would like to do in the future are things that just don't seem to be getting done; the studies on the mass media. There are all of these ideas about how the mass media could be used to diffuse very important and beneficial health information or educational information or information about the prospects that you could pursue to achieve your human potential. But these PSAs, these messaging interventions have as far as I know never been evaluated using randomized designs, real-world interventions, and unobtrusive outcome measures.

**Stephanie:** I was excited to see your study in Rwanda, *Using Mass Media to Change Norms and Behaviors*. Are there other studies like this?

**Donald:** You know Betsy has done a few relatively small scale ones, one in the Congo and one in Sudan, kind of lab-like field experiments. What I want to do is more akin to the randomized experiments that they did with Spanish language radio where you randomize the media market, or the ones that I did with the Texas gubernatorial campaigns in 2006 where we randomize the propagation zone. Or there's one that we're doing with inserting pro-social messages into scripts on Telemundo. So the nice thing about that is that we're measuring behaviors in time as the pro-social messages roll out. So we have to do car seats and scholarships and cholesterol, all kinds of messages, they're woven into the scripts.

**Holden:** How do you normally go about getting your funding, I'm just curious when you say you get to do what you want for the most part. I wouldn't have guessed that because most of the foundations that I've talked to have pretty definite agendas and you say that most of your money comes from foundations, so how does that work?

**Donald:** I've been able to secure my own research funds. It's not a vast amount but it's pretty good. Then when I need to supplement it project by project I call people up and try to get funding for them.

But the reason I'm so interested in the mass media thing is that I sense that that's where I'm headed. As I mentioned, I've done the $20,000 experiments, and the $200,000 and the $2 million experiments and now I want to do the $20 million experiment. Mass media is a great unknown, so much is hanging in the balance and we don't know the first thing about its effectiveness.

The stakes are huge. And when you think about the amount of money that goes into this… when we started doing this, we always thought that the folks in the economics market, they know the answer to these questions …

**Donor:** Which questions?

**Donald:** The question of how cost-effective these ads are. We asked, "What is the evidence that your ad campaigns works" and it was the most feeble form of research design. And we would say "You are getting $500 million for this campaign and you don't know if it works or not?"

**Holden:** So here are some ideas, I'll run through them now and would be happy to talk another time. We have one idea for a special issue of a journal where people just publish a dataset as a publication and then if people use that data for a future study they're supposed to cite it. The idea here is to make the creation of useful data sets more prestigious, have it contribute to researchers' impact factor.

Similarly there was the idea of a special issue of a journal devoted to replications or critiques of publications.

**Donald:** That I'm very enthusiastic about. The first idea that comes to mind – I'm a little less taken with ideas that just seem like monkeying with dubious statistics.

**Holden:** How is that this? It's just publishing a dataset.

**Donald:** I'm not opposed to that, I'm just worried that if the pretext for doing that is ramping up someone's impact factor…

**Holden:** I see.

**Donald:** But I'm all for publishing the data. And the idea of having a publication – that's important. I think the irony of how we operate our journals now is that it seems that we're increasingly clamping down in a way that is so unconnected with technological advance. We should be making it easier and easier to publish with lower and lower barriers and almost no barrier in terms of length. What difference does it make, say as much as you want. It's all online. Whenever people get bored they can stop reading. So having a journal of replication, I'd even volunteer to edit that journal because I think it's so important. The only thing that I don't want to do is make the existence of that journal a pretext for other journals to reject replications. That's why my suggestion would be to have the journal be a *sui generis* format where it doesn't look like any other kind of journal article. So someone cannot say "You can't publish your journal article because you've already published there." Make it so that it's not written in prose or something.

**Donor:** So that it's not a substitute you mean.

**Donald:** Right.

**Holden:** Another idea for how this could be presented is that journal of X has a special issue devoted to replication and discussion of journal of X studies.

**Donor:** A journal of good ideas, a journal of good designs where the data isn't in yet. A journal that would commit to saying "This is a good enough question and a good enough design that provided that you actually pull this off and publish this data we commit to publishing it."

**Donald:** The only proviso that I would add is that there should be an escape clause where if your analysis is totally incompetent…But I fully believe that there should be like a version of articles sent to reviewers that does not cloud people's intuitions by showing them the results.

**Holden:** Then we've got this idea for what someone called a social observatory. The idea is to have a much more top down centralized attempt to collect general-purpose data and make it publicly available. So what happens is you go into a country and you say "We're going to follow 20,000 people for 20 years. And we're going to convene academics to figure out what data to collect and how and when. And half of it will be a control group, and no one will intervene. The other half will be sectioned into ten treatment groups." We'll auction those treatment spots to academics who want to try different treatments. It's a little hard to articulate what's different about this from the way academia is done. But I think the idea is "data first" – a project to collect a huge amount of data at a large enough scale. Then people come in and start doing experiments rather than having a data set built around one question.

**Donald:** It's an intriguing idea. And I like the idea of economies of scale. The only question is whether there's something obtrusive about this study because it's so big it's on people's radar. If you're in Guinea and all of the sudden you see a whole lot of people with survey clipboards it could cause people to change the way they answer the questions. You have to be careful when you do things in scale.

**Holden:** I see, okay. A similar idea: what if there were a dedicated organization – the "shocks" organization for whenever there's a shock, just anything that happens that's unusual and seems like a causal inference opportunity. What you do is you go around and poll all of the relevant academics and ask "What interesting questions could this shock help us answer, what kind of data do we have to collect to answer them and what do you predict that the data will say?" and then you fund the collection of them.

**Donald:** It's kind of like a SWAT team for observational research.

**Holden:** Finally, you know the Cochrane Collaboration? I like the idea of seeing that kind of work in more fields. I guess in your field, I don't know if the Campbell Collaboration is active.

**Donald:** It's really not live in political science. There's nothing in political science. It's mostly criminology, a little bit of education. I think the idea is okay, but I'm not so sure that we're at the point where meta-analysis is an option … but there are some areas, like my area, campaign selections, where there definitely should be.

**Holden:**So maybe let's not be quite as literal about Cochrane and instead of meta-analysis talk about systematic reviews.

**Donald:** I like the idea. Particularly one of the reasons I like the idea is because that is a messed up enterprise, meta-analysis. Chapter 11 in my book doesn't try to poke holes in it but it tries to say "look, before we go down the path of meta-analysis and put everything in a blender let's ask whether the estimands are the same." Because there's this idea that you standardize everything and put it in the soup and you get out a general conclusion and I think it all depends on what you're trying to estimate. So I would like to see meta-analysis done more, but the meta-analyses that I can think of are in political science tend to violate all of the rules, even the famous ones in psychology tend to be pretty bad.

**Holden:** Well, the goal would be to have someone specialize so that their job was to do it.

**Donald:** The only thing is that if you have people specializing in it, there's a little bit of an incentive to have them do it as opposed to saying "Yes, I'm specializing in it and my recommendation is not to do it."

**Holden:** Well, they would specialize in systematic reviews. And doing meta-analysis when appropriate.

**Donald:** That sounds good.

**Holden:** And then the final one, this is a much bigger picture thing, but if we step back and ask "We're a customer of research, how do we wish research looked?" It involves no journals and rather ArXiv instead, you know ArXiv in the mathematical community, something like that with some sort of reputation system that's based on real world credentials so that your real position impacts your impact on reputation scores and people can respond to each other and responses are linked and they affect reputation. That's on one side of it. And on the other side of it you create incentives to get a good reputation score and you try to get the system to work such that tenure correlates better with your reputation score than traditional measures. There's kind of a long path to get there, I have a couple of ideas.

**Donald:** I like that. In a way that's what people try to do but they do it in a very old fashioned and unsystematic way as part of tenure reviews.

**Holden:** I'd like to hear more about that so maybe we'll meet again.

**Donor:** I have a last one, a small one. How interesting and important are priors? There's very little discussion around expectations, "What does the field think?" What does the field think that this experiment will yield?" Not just because it's fun to think about it, but because this has all sorts of implications about whether the tool that you're collecting data with is any good. This is also one of my favorite topics; it's also in Chapter 11. Basically the institution that comes out of this is a betting market. One of the reasons that we had such a strangely designed social pressure study in 2006, the one that was in the 2008 American Political Science review article, basically exerting social pressure on people to vote, is that we totally underestimated the size of the effect; we had an internal betting pool amongst ourselves and did our power calculations based on the upper most bet in the pool. That was just a case where it didn't work out, but you could imagine doing that in a wide array of different cases, the nice thing about that is that it throws overboard naïve classical significance testing because you're not always testing against the zero, if people really do believe a particularly large claim, "I think this is a 19" it's supposed to give us the 19. You also have a very clear way of saying "here's the prior based on this betting pool and now here's my posterior based on the data."

**Donor:** Could that be linked to a registry? That requires in some sense that the design is public. That creates this whole concern about fishing and all that.

**Donald:** I imagine this is the kind of thing that could be cobbled together literally two days before the results are going to be run. There isn't going to be much of an insider trading problem.