

2 Abhijit Banerjee and Esther Duflo

Esther and Abhijit probably need the least introduction of any of the economists featured in this volume. They are founders of the Jameel Poverty Action Lab at MIT, arguably the most prominent center promoting the use of randomized control trials in development research—and the application of findings from that research to policy. Their book *Poor Economics* is a bestseller and winner of the 2011 Financial Times Business book of the year prize. Esther is a winner of the John Bates Clark Medal (2010) and a MacArthur Foundation Fellowship (popularly known as the “genius award”) (2009). Abhijit has been both a Guggenheim and Sloan Fellow, and has served as president of BREAD, a consortium of researchers focused on development. Currently, she is the Jameel Professor of Poverty Alleviation and Economic Development, and he is the Ford Foundation International Professor of Economics, both at MIT.

Born in France, Esther attended some of the country’s most selective educational institutions before moving to MIT for her PhD in Economics. In a highly unusual move, she was appointed Assistant Professor in Economics at MIT after completing her degree.

Her thesis exploited a natural experiment in the expansion of public schools in Indonesia to assess the returns to investment in schools and education in that country and was published in the *American Economic Review*, one of the profession’s most prestigious journals. She has gone on to publish dozens of papers in top journals as well as be recognized as one of the leading thinkers in experimental economics in development.

Abhijit, meanwhile was born in Calcutta—his parents were both academic economists. His early work was primarily in microeconomic theory, a background that continues to heavily influence his work in more applied economics. He describes his shift from theory to applied economics as a result of “a sense that I know all this economics, so why isn’t it helping me understand stuff that I see next to me?”

TO: What was the first experiment you were involved in?

AB: An experiment in a rural Udaipur district in 1996 I think. It was on changing the teacher-to-student ratio and seeing if that changed test scores. It was the first experiment of many that found the same thing, which is that you can add a lot more teachers per student,¹ you double the teacher to student ratio in this way and there's a precise zero effect on test scores. I think that we've found many times and eventually provoked us to understand a little more about the need for different curricula.

ED: My first RCT was the Balsakhi experiment.^{2[1]}

TO: It seems to me there's been a missing conversation between people looking at poverty and entrepreneurship in the developing world and people looking at poverty and entrepreneurship in the developed world. I find it particularly striking in terms of what we know about small-scale entrepreneurs—the local small business owner, the pizza shop, the strip mall gift store—in the US: they are low growth, low profitability, and high-failure-rate enterprises. It seems little of that knowledge or research was applied when it came to microenterprises fueled by microcredit.

AB: I think there is an important commonality and a very important difference. The biggest difference is that in places like the US where labor markets are strong, the incentive to go into entrepreneurship is very different from a setting where the low-end jobs are missing. The reason why there are so many more people in the low-end entrepreneurial sector in developing countries has a lot to do with the labor markets in those countries. The people who become entrepreneurs in the US actually do it more often out of a desire either to be an entrepreneur or that for some particular reason they don't want to be in the conventional labor market.

That's the difference.

The commonality comes in exactly in the sector you described. The area of US entrepreneurship that has that flavor of a developing country is in these recent-immigrant-run industries, undifferentiated and relatively low-tech businesses.

1. Note as discussed later in the interview, there is a difference between adding more teachers using the standard certification and hiring processes versus adding tutors who offer remedial education.

2. An evaluation of the value of adding non-union teachers to classrooms for remedial education, discussed later in the interview.

To use an example I know, imagine you come to the US because you are sponsored by your brother who is an engineer but you are 45 by the time your brother becomes a citizen and he sponsors you to come over. What are you going to do? You are not going to enter the labor market very easily. You probably also have teenage children who will have an equally hard time entering the labor market. That generates a classic labor availability-driven entry into business.

That particular end of the entrepreneurship market is very similar to that in developing countries, which is very family labor-driven. But I think the key in the conversation is the difference in the labor market. These are the people who are excluded from the traditional labor market.

ED: I don't think it's true that there's not much conversation. I think there is a lot of conversation and I think it's a lot of the same people doing the research like Sendhil Mullainathan, Dean Karlan; I work in France. We work in Chile, which is more like the US than India in many respects. There are a number of people who work in both types of environments with the same kind of method and the same type of approach to the problems.

I think there are two reasons why they have not talked about US entrepreneurship much. One is a good reason; the other is a bad reason. The good reason is what Abhijit said: at the end of the day the people who run a business here are not really poor.

When we are interested in the poor, which in France doesn't mean running a business, it means having no job whatsoever, it means dropping out of school at 16 with no skills. The research is focused on things like payday loans, not on entrepreneurship.

The bad reason is more what you are talking about. I think a lot of the discourse on entrepreneurship in the US is not centered on these guys. It's centered on Facebook and Bill Gates who are extremely interesting people but aren't representative.

I think this is changing. I see more graduate students who are interested in this idea of whether people are running businesses as a result of losing a job and those sort of questions. I don't know whether it's a real phenomenon or just looks like it, but it does seem to be something that people are starting to get interested in. Seeing graduate students taking note of these questions is a good sign for research to come.

TO: One of the things that got me thinking in this direction is the Field and Pande³ experiment delaying the first repayment in a standard microcredit contract.^[2] In the US no one thinks that banks are a good source of entrepreneurial capital. Why did people buy into the idea that a banking model was a good model for entrepreneurial capital?

AB: Let me disagree with you. I think that banks have always been a decent source for paying salaries. Essentially the only funds you get from a bank are very short term loans that you use to deal with your cash flow, like a line of credit. If you think of what a developing country entrepreneur is, she often doesn't have a lot of fixed capital.

For a lot of these microbusinesses, cash flow really is the core issue, so in that sense it's less implausible that banks are good sources of capital for microenterprises. Once you've bought the weights for your fruit stand, and you might already have that, all you have to do is replace one form of cash flow financing for the day with another form. In that sense microcredit is much more like paying salaries or wages than like setting up a business.

TO: But the capital that small businesses get from banks in the US isn't growth capital. No one really expects it to finance growth. Lines of credit have always been conceived as cash management devices, not as a solution to launch growing businesses.

AB: Even in the US, your credit line expands if you do well. And if you think that 90 percent of the capital needs of these businesses is cash flow, the fit isn't that bad. There are lots of reasons why microcredit doesn't work as well as it should, but I do feel that there's some similarity—the bank model is not entirely wrong.

ED: Two points. To some extent the fact that the businesses are like that is endogenous to the source of the money that is being offered. That is the point that the Field paper makes. That's the point that you were making. I am going part of the way to what you said.

I think the Field/Pande experiment—giving people a little bit more time before repayment is such a small thing that allows people to consider

3. In traditional microfinance, the first payment on a loan is due a week after the loan is made. Field and Pande randomly offered some borrowers a two-month grace period before repayments began. They found that those who did not have to repay the first week invested more in equipment and ultimately had higher profits, though also a higher rate of default on the loan. See video interview with Rohini Pande here: <https://www.youtube.com/watch?v=rCYSSYf3laM>.

something like buying a sewing machine—really shows that the reasons the businesses are like that is a product of the financing. Maybe something else would be better.

However, the question you have to ask is: what's the alternative?

Maybe some form of equity is better. This is what people believe to be the right way to give capital to entrepreneurs here. But how do you do equity easily in a system where very small enterprises makes it extremely costly. So my point is slightly different, which is maybe it is not the best way to finance entrepreneurship in the best of all possible worlds, but we are not in the best of all worlds.

Maybe not. Maybe someone could still come up with a way of providing equity in an intelligent and cost-effective way. Abhijit has been talking for a while about that the fact there are many chartered accountants in India. You could use them to provide a little bit of verification. Maybe it is possible to do equity to microenterprises, but we are yet to find someone to bite, who would be willing to give their own equity.

AB: I think there is a lack of alternatives. Even in the US, most small businesses' fixed capital is financed by second mortgages on a house, borrowing from relatives, something like that.

ED: In the US, when you are big enough, you start getting access to formal finance.

AB: That's where in the US there is a huge difference. Once you've gotten to a certain size there are lots of people who give you lots of money, take equity in your business, but also even debt capital. But I think that, at the beginning, it is the case that essentially even in the US, people go to the bank to pay salaries, and then they go to suppliers who give them some trade credit and then they use their second mortgage to buy a machine. So in some ways it's not that different in developing countries. It's just that the levels of income are so low, so that you can't buy a machine with your second mortgage.

TO: But I think it is an example of something you wrote about in your book^[3]: an oversimplistic understanding yields a solution addressing the wrong problem, which yields inertia to keep doing the same thing. I think a huge part of the overindebtedness crises in microfinance is that people really believed that they were providing growth capital. If you really believe that you're financing growing businesses, why would you stop? Why would you rein that in?

ED: It's been obvious for a long time that these business are not growing. The Grameen Bank has been around for many many years and their loans are still very very small. Just forget about subtle impact evaluation or whatever; it's been staring us in the face that these businesses are not growing, and the vast majority of people are not growing out of poverty or anything like that. If we had not been obsessed by the romantic idea of microcredit, then maybe there would have been an earlier realizing of what microcredit does and what it doesn't do. I think people are coming to that, to a small extent maybe because of our work stirring the pot. To be honest, it maybe would have happened anyway. But there's been a lot of delay given that the facts were pretty obvious.

AB: The crisis in microfinance in India⁴ was a result of the 3 C's: credulity, cupidity, and corruption. The politicians were corrupt, we were all credulous, and the microfinance people were greedy. Put them together and you get the crisis. Our credulity was significant. Somehow we believed that all repayment happens in microfinance due to some magic, which made no economic sense. We knew it didn't make economic sense. And then suddenly one day we wake up to the fact that the actual loan officers would come to a borrower's house; maybe they don't beat the borrower up, but they do harass them.

You don't need to do an evaluation to start asking questions. You just need to think about it for 10 minutes. These are desperate people with lots of financial demands. People in the family are sick, people lose jobs, the daughter needs to get married.

ED: But they repay anyway. Someone must be very convincing.

AB: But 90 percent repay. What is going on? How could we believe this was because of some tweaking of economic incentives? As economists, I think we were basically inept in thinking about it or we would not have believed it. The core fact was staring us in the face. And we didn't look at this, we evaded the gaze of these facts that were looking back at us.

TO: Speaking of facts staring us in the face, there are some facts that I feel like are staring me in the face, but I don't know what to make of them. Where do you feel the balance of evidence is on giving people money and letting them choose, or knowing that there are some things they are under-investing in and so we should give them, say, a bed net?

4. The reference is to 2010 when microfinance institutions in Andhra Pradesh were required to stop collecting repayments by the government as a result of accusations of abusive collection practices and people being forced into debt traps.

ED: The problem is that no one has done that comparison. We know they are not buying bed nets with their GiveDirectly⁵ money.^[4] I guess I don't know that for a fact, but it's knowable. I think it's the type of question that will be frankly difficult to address because you're required to value benefits across sectors. So it depends what your objective is. If your objective is to control malaria, you can do an experiment where you give money and you give people bed nets or you do something else and you can see how many people sleep under a bed net at the end of the day, and what's the number of malaria infections you've averted. So that's a question that's well defined. Whether it's better to give cash or to give bed nets would require you to make a judgment about what is the importance of making people healthy versus having them buy a roof. And that's one I'm not prepared to make.

TO: Shifting to another topic, let me ask about women and girls. In almost all advertisements for aid and charity these days you see some version of the idea that the focus of aid should be on women and girls. There's indication in your book of some disagreement between the two of you on the evidence for that claim and of the need to focus on women and girls.

ED: That's mostly in jest

AB: We both agree that the idea of focusing on women and girls is very intriguing but slightly overblown. We don't actually know. Men versus women seems to make a huge difference in some places and not in others.

ED: It's not overblown necessarily. Women do spend money differently than men. The issue is whether women are inherently better people or the differences are the outcome of the social structure.

TO: You write about your fathers. Your fathers weren't drunkards who cared nothing about their children. I think about the generations of my family and we're not so far removed from rural farmers or the lower rungs of the working class. My grandparents and parents didn't go to school only because their mothers invested in them. I'm troubled about the implicit idea that mostly black and brown men in poor countries don't care about their families and white men in rich countries do.

5. GiveDirectly is an NGO that makes direct, unconditional cash transfers to poor households in Kenya. A randomized evaluation of the program was conducted finding that households increase food consumption and investment in durable goods do not increase consumption of alcohol or tobacco, and large gains in psychological well-being.

AB: That's the sense in which I think the evidence is overblown. I think that the evidence is clear that the particular forms of social dysfunctionality that emerge when economies are not working well have very different effects on men and women. That's partly because of the social roles assigned to them. So when you are a farmer and you only have a small bit of land and you realize that everybody else is making much more money than you, somehow your assigned role as the income earner of the family means you feel that in a different way than your wife who is also feeling the pinch of not being able to afford things. But in a sense it is definitely true that it is not her assigned role, and I think it makes a difference that it's not her assigned role. She is more able to discharge her assigned role than the man is in difficult economic circumstances. And that's almost surely one of the reasons why in the US, when the urban blue-collar industries went into decline, the traditional working class and particularly African American working class family got under a lot of stress.

I'm sure this phenomenon exists everywhere in the world. When you lose your job or your job is not up to snuff, it's difficult for men in traditional societies. I think it has a lot to do with the fact that their assigned role of income earner is particularly vulnerable to the shocks that the families were subject to. I do think that it's exactly what you said, I don't disagree with that. I think there is an easy essentialization of gender that is clearly dangerous.

ED: There's a factor of what I think is called "benign sexism." For instance, there are surveys where people ask, "Do you feel that women possess an inherent quality of gentleness that men lack?," something like that. And people say "yes." That's also a form of sexism.

TO: There's some hint of this as I understand the results of the Spandana microcredit study.^[5] You found that women didn't spend more in the categories like education or food where people expected that women would spend more of their income than men did. I wonder if there's a discretionary income issue. Men traditionally have the discretionary income, and when they feel they have done the work to earn the income, then they have the right to spend some of it on themselves. But in traditional societies women are earning or receiving a portion of the family income for a specific task—to buy food and take care of the children. So that's what they spend it on. When women become discretionary wage earners, they behave like the men and spend more money on themselves—as is their right.

ED: That's what Chris Udry and I found in Côte d'Ivoire and we talk a bit about it in the book.^[6] It is true that when women do better than men in a particular year that they spend more on food which is their particular job, their social role, but they also spend on themselves. So it's not that women do not want to spend money on themselves. Their traditional role involved making sure that people are fed. And that's less the case for men. Why societies evolved like that makes sense.

TO: There are perennial questions about external validity of any study, what results from one study tell us about what might happen somewhere else. This speaks to your theory of change. It's always been my impression that the idea is not just to transfer a program and assume it will work but to transfer and test it in the new context.

ED: Absolutely, keep testing. But it's not just repeating the same test over and over.

We have six RCTs of the impact of microcredit.^[7] Most of the energy shouldn't be spent on doing 50 more of those tests in different contexts. Although there might be because it's now easy for people to conceptualize how you would do it. So the marginal costs are diminishing at the same time as the marginal returns are diminishing.

You're also seeing these eight ultra-poor studies⁶ that we've put into one paper.^[8] Now that there is more money and more recognition and more partners, et cetera, now that there is a system, you are seeing more replications. I've always thought that they would come, and they are coming.

Now we're saying, for instance, that the results of the six RCTs of microfinance are broadly consistent despite the fact they are in different contexts. You have people like one of my students who are trying to think about how we go from saying they are broadly consistent to being able to pull them apart by site, or by another dimension of heterogeneity, and what that can tell us. To what extent are the results pulling together and to what extent are the pulling apart? What's making them move?

So as we have these replications, as we have more tests, it's not just that we can say, "oh, nice, it replicates," or "this one doesn't replicate." Replications

6. These are a unified evaluation of a particular program targeted at those living on less than \$2 a day that provides asset transfers, livelihood training, and other supports. The programs and the evaluation are discussed in more detail later in the interview.

also generate the ability to exploit that data in a structured way to learn about the importance of the context.

When, say, a government is interested in the result of an experiment, my experience is that they are not usually interested in taking it wholesale and applying it, they are interested in doing their own experiment.

There is this sort of, “experiment begets experiment.” We saw that with the conditional cash transfers [CCT].⁷ After the initial experiment, countries didn’t just adopt the program that Mexico had. They started doing these experiments over and over again.

TO: Then they starting doing unconditional cash transfers instead of CCTs.

ED: Exactly. For a while it was just CCTs, but after some time they were like, “OK fine, can we learn what’s important in the CCT?” So unconditional cash transfers came up. Or there’s stuff I have done on labeled cash transfers.^{8[9]} So there are two ways in which the research builds on itself. One is replication across contexts, and the other is going deeper into the mechanisms and variance between subjects or implementers and other factors that have policy implications and further research implications. So you never stop testing.

But it’s also true that it’s never happened that a program was just copied wholesale, except with the possible exception of deworming,⁹ and with deworming the theory of change is so basic at some level maybe it’s

7. A now widespread social protection program, conditional cash transfers are welfare programs that require specific actions on the part of recipients, such as getting vaccinated or attending school. Conditional cash transfers came to modern prominence in Mexico in the early 1990s.

8. In an experiment in Morocco, some families were enrolled in a CCT contingent on school attendance, others were simply given cash that was explicitly for school expenses but attendance was not monitored. The authors find similar effects of CCTs and the labeled transfer. For more discussion of the study and its interpretation, see Berk Ozler: <http://blogs.worldbank.org/impactevaluations/cash-transfers-sorting-through-hype>.

9. The objective was to provide medicine to rid people of worms or intestinal parasites like soil-transmitted helminths. The study that many point to as the launch of the RCT movement was a randomized evaluation of a deworming program in Kenya by Michael Kremer and Ted Miguel. Eventually the study led to massive expansion of deworming programs, but also controversy. See additional discussion in interviews with Kremer, Elie Hassenfeld, and Angus Deaton.

appropriate, that you have just one study and you try to spread it around. It’s empirically not what’s going on.

There’s a study here and a study there, and then another thing that’s related and another thing that’s related, and you start forming an idea for what might be a solution to a particular problem. For example, the low learning level in schools. That seems to be the case in a lot of countries, but it’s embodied slightly differently wherever you go.

So you try something out in what seems to be the right context based on what you know. And you learn about all the implementation issues that come up along the way. It takes three or four experiments just to iron out the details before you can come up with something that looks more like a turnkey approach. You don’t go from the Balsakhi¹⁰ project to “here is what you should try,” without 10 years of trying things out.^[10] I don’t think the time frame is going to be shortened that much. It’s been 15 years since the first Balsakhi experiment started in 2000, and now its 2015 and we are *maybe* ready to start saying, “you can do this.”

TO: I’ve written about what I think your theory of change is but have never asked you directly about *your* theory of change.

AB: Most of it I just keep hoping [laughter]. There are two things that all of us do, me no more than others. One is the really concrete thing of saying “look, I went and measured.” In the end people have the sense that in the policy world there’s a lot of respect for actually having gone and collected the data and seeing what is happening. I’ve talked to lots of bureaucrats. When they realize you know a fair amount about something, you get a lot of respect. Some of that comes from understanding you did an RCT, some of it comes from the fact that you’ve been and seen a place that they’ve never been.

Part of the theory of change comes from just engaging in the conversation. Policy makers often seem to live in a void. They think, “there’s really not much to know, it’s on me and I have to,” in the equivalent of the white man’s burden, “solve this problem.” But when you’ve gone and measured something and you talk to the bureaucrat, they begin to appreciate that there is a lot of knowhow out there. I think that’s the first thing we have to do. We have to show we have credibility, that we do know something about

10. An experiment to provide “contract,” that is, non-union, teacher assistants in classrooms to tutor children who have fallen behind grade level. The program has since been tested in Kenya, Ghana, and other locations.

this issue. There are a lot of externalities in credibility obviously. When I meet someone and he says, "I met Jonathan Morduch and he said this," Jonathan has already paved the way to this guy listening to me. There's a collective investment in credibility that is very useful, that's why J-PAL and IPA are important. They've created an institutional basis for credibility. A lot of change is coming from the fact that policy makers and civil servants are saying, "well, there are these people who talk about evidence and they seem to know something and other people seem to take the seriously, so let's talk to them." Part of it is just getting the conversation.

In my experience, it's not the case that the policy makers have perverse views on what they want to do. They want to do something good. That's often true of even people who might in other ways not be great human beings. They might be corrupt or all kinds of things. It's not that they don't want to do anything. They have some ambition, maybe some social mindedness, maybe some ambition that if I do something maybe I'll stick around longer. For all those reasons they're responsive to things that you if you go and say to them, "what you're doing makes no sense," for this and this reason, it's not that they don't necessarily believe that.

Part of what we have achieved I think is a bit more credibility. Now, if you are in government in India, you might have heard of J-PAL and therefore if we come and say we really don't think this is the thing to do, at least you get a hearing. Some of them might actually think about it. And then they ask, "what do you think?" You do hear a lot of "what do you suggest?"

That's not just true for us. That's what's true for organizations like Pratham¹¹ because they have been on the ground, they have data, that whole nexus of familiarity and some expertise. All of that together is very useful in getting past that initial hesitation of, "what, do these people know any more than me?" I think that landscape has changed and I think there are enough people who believe that there is expertise out there just as they believe when they want to issue a bond they know that there is expertise out there. Even the corrupt bureaucrat calls in the three bond market experts from Goldman and talks to them.

Sadly, that's where we haven't reached yet, but we might, that when people want to do a program, they will call in people who actually have data. To

11. The education-focused NGO in India is perhaps most famous for initiating an annual report on learning levels in schools—an approach that has been copied in several countries.

some extent it has happened. Even in the previous government you saw that people like Jean Dreze, who may be wrong but at least who had done some research, had some influence. There is some understanding that there's expertise out there and programs don't necessarily work because they sound good.

I think that idea—that programs may not work—has penetrated. Perhaps what has penetrated most is that. Not so much that we know how to make things work but we can at least tell you something useful about how things might go wrong. But that's still progress.

TO: How much have you thought about influencing specifically the people who make the small changes in policy? You do this work with, I presume, the idea that it will change the way certain people behave and they'll stop trying to make the big sweeping changes because they understand how things might go wrong and instead try smaller things and test them.

AB: People talk about how hard it is to change policy or influence policy and that the institutions in developing countries are resistant to change. I feel like a lot of what we see as the culture of government is actually in some ways a very mechanical product of the last 50 years. Most countries in the world have created a whole bunch of institutions very, very recently.

Developing countries inherited a kind of a state that was not particularly devoted to the welfare of the people. So they saw these models of the developmental state elsewhere and decided, "we're going to have this particular set of people implement this particular set of things." So the institutions are not the product of a long accretion of incremental changes that made the bureaucracy particularly well suited to a particular kind of policy. This bureaucratic culture we have in so many places is just happenstance.

Most developing countries have only been independent since the late 1960s. So we're talking about a period of 40 years. It's hard for me to imagine that what we are observing in developing countries now is a steady state result of some complicated process.

Most things are done with such amazing casualness that I feel like it can't possibly be anybody's reflective thought on how or what a process should be. People were given lots of power, lots of decision rights, little training, lots of ideology. I don't feel that the bureaucracy in most places has reached an approach where it's snugly sitting and we'll never change it. Lots of

bureaucrats and especially their political masters are very sensitive to the fact that somehow they lose elections. My sense is that they're quite bewildered by that. I think, in principle, there's a lot of demand for rethinking government.

When we talk to bureaucrats in India, obviously there are some who will tell you that they know everything and don't need any help. But there are a lot who say, "yes, I think we should do that; can you help us with that?" The reaction is often "why can't I get many more people like you to come in and help us with how to redesign this whole thing?" They know things are broken, too many programs don't work, but it's just too difficult in the middle of the job to fix it.

ED: I think we are reasonably internally consistent. If we want to be internally consistent, then we can't be calling for a big revolution when one day everybody will start looking only at the evidence before making decisions. That's not going to happen, let's face it. There are a lot of reasons why things happen the way they do. So if we were trying to advocate to replace the world the way it is with some fully efficient technocracy that experiments with things before trying them out and then launches them with no error in the process, that would be inconsistent with what we are saying, and *crazy*. That makes our life easier. If a big change, a revolution, is not what you are targeting, if you are targeting improvement at the margin, there are lots of margins where you can start doing things. Maybe it's a somewhat opportunistic thing to say. We don't have a huge reason to fight people who are really resistant because we can always try with someone else. Like Abhijit was saying, there are a lot of people who are willing to try things out. In the beginning of J-PAL we worked a lot with NGOs because they were more nimble. Even there you have some who are never going to change their ways, and some who are more flexible and we worked with the ones who were flexible and eventually some who weren't initially willing came to talk to us.

With the government, and with international institutions, it's a bit like that as well. You can start having conversations and trying things out and trying to improve something somewhere. So we're really talking about: can we make your school committees work better or something "unambitious" like that. And from there you can demonstrate that it can be done and eventually it can become bigger and bigger.

So, for example, we started working with the police in Rajasthan. We describe this in the book. It happens to be that at least some people in the

police in Rajasthan were interested in improving performance. They tried some things and it had some effect.^{12[11]} Then we could use that as an example when we went to talk to the government of Gujarat about improving their pollution inspections¹³ and in the process of that got the attention of the central government who were thinking about a cap and trade scheme for carbon.^[12] And so Rohini Pande and Michael Greenstone, who were working with me on the pollution issue, then began to help design a cap and trade test. But it didn't happen by us knocking on the door of the minister and asking, "can we design your cap and trade system?" It happened by taking the path of least resistance.

TO: One of the things that you say toward the end of your book that leapt off the page at me is, to quote directly, "The poor bear responsibility for too much of their lives." I can imagine a lot of people reading that sentence and being taken aback. It seems to be a stark contrast with the Sen idea that escaping poverty means having more freedom, more control of your life.^[13]

AB: I think we choose our words carefully. Control is not responsibility. In a sense I think having a lot of responsibility undermines control ...

ED: ... and freedom.

AB: And freedom. When you want to exercise control, you need to have the psychological freedom to actually exercise it actively, rather than passively reacting to many many things. Control is not passively reacting to many many things. It's agency. I think responsibility, lots of responsibility for very difficult things, undermines agency. It's not that you choose to do those things. It's that you have to do those things. So it's not choice. Responsibility is not choice. It's things that are dumped on you that you have to struggle with, and in the process I think your agency is undermined because you can't reflectively decide that this is the life choice that I want

12. Banerjee et al. tested a variety of approaches to improve police performance in Rajasthan. They found that reforms that ran through middle managers were not effective, but the announcement that "secret shoppers" would be visiting stations pretending to be civilians trying to report crimes and direct on-duty training improved performance.

13. This was an experiment to change how pollution auditors were paid. Auditors were paid by the firms they audited (as with financial audits). By changing how the market for auditors worked and how they were paid, auditors reported pollution results more truthfully, and audited firms were more likely to reduce pollution levels.

and this is the life choice that I don't want. I wouldn't call that being in control or being free. I think taking responsibility off people's hands and giving them a domain of freedom that is uncluttered where they can make choices without being constantly frightened by all the things that can happen, all the risks around, just the intellectual challenge of balancing 500 things, I think it is giving them freedom. Maybe it's a disagreement about the nature of freedom. *Prima facie* I would not say that responsibility is control.

ED: I don't think that's against Sen. It's going further in the Sen path. The whole argument of Sen is that freedom means nothing unless you have the capability to exercise it. There are many examples of that, but one very striking example of that which he cites is the family in Bengal that is completely free to buy grain but they have no money, so they can't do it. The point he is making is that freedom is meaningless unless there is the possibility to exercise it, the capability to exercise it. We are not saying anything different. I mean, we are not free in the US to drink water that is contaminated with *e. coli* because it comes to us clean. Is our freedom reduced in comparison to the person in Kenya who is free to do that because, if they don't want to drink the water with *e. coli*, they have to put chlorine in it. I don't think we are less free, I think we are more free.

I'm not saying that we are not making a political statement on some level on the nature of freedom, but I think it's an argument that's pretty naturally in the Sen line.

TO: Of all the things you've worked on, if you were to predict which are likely to have the biggest policy impact, what do you think those will be?

ED: It's very hard to tell because you can count the number of people touched when things work. But I do think the RCT studies of microcredit already had an impact. I do think they have already changed the industry in ways that I don't know exactly how to quantify. I can't say what would have happened without us, but I do think the conversation is different than it was in the past. When you hear the practitioners, you can compare what they are saying now to, for instance, the IPA conference at Moody's.¹⁴ The

14. Duflo is referring to a conference on microfinance research hosted by IPA, Moody's, and others in October 2010 that included representatives from microfinance institutions. The conference was held when many of the randomized impact studies of microcredit were just beginning, and the general stance of those in the microfinance industry seemed to be that such studies were unnecessary because the positive outcomes from microcredit were well known and proven.

statement that these same guys just put out^{15[14]}—they've acknowledged that they lost in some sense relative to what their claims were and now they're just changing the claim—which is totally fine by me.

TO: It's amazing how hard it is now to find anyone who ever thought microcredit was going to be transformative.

ED: Exactly. But that's fine by me. I'm not in this business to score points. Just hearing the way in which they disagree with us, it has changed in such a fundamental way. I think we've had an impact there.

I think the education work can have influence because I think it's fundamentally right and not difficult, but the politics in the education world is very different. I don't know if we'll be able to manage it, but I hope that it can work. We have to try. I think we have to have a couple of successes. We have had successes, but I think they're still seen as small projects. We have to get to state-wide adoption.

Those are two areas. It's not as important, but my smokeless cookers failure of impact work is important. That paper came a bit earlier in the cycle of how these things get rolled out compared to other areas we've worked on. Smokeless cookers were going to become the new microcredit, but that work put an end to that.

TO: It's a curious case, because the improved cook stove has been around for 40 years, and you continue to hear, "we didn't get the design quite right but now we know." And then the next one doesn't get taken up, and the cycle repeats.

ED: It's still going on, but now you have more discussions about what users want, and I think more recognition that getting it right is so difficult.

Another area is the pricing of health products, something that I have nothing to do with, but I think will continue to make a big impact.

More important, I think, is the general recognition that experiments can be a part of the learning in the organization. There's a pulling back from

15. Duflo here refers to a statement from the Microfinance CEO Working Group, a group comprised of the CEOs of six large international microfinance networks, in response to the publication of the six randomized evaluations of microcredit. See Microfinance CEO Working Group, "Measuring the Impact of Microcredit—Six New Studies," 2015. The statement, and a prior statement of the Microfinance CEO Working Group disputing the findings of one of the first randomized impact evaluations of microcredit, is discussed in chapter 7, in the interview with Alex Counts who was a member of the working group at the time of both statements.

evaluation as an accountability tool, moving it away from evaluation offices and professional evaluators back to “the doers” and getting governments to accept it, more than that, excited about using it as a tool.¹⁶ That is probably the biggest impact. It’s a change of the conversation.

TO: I find it a little ironic that the most well-known set of RCT results is the “no effect” of microcredit. It feels to me philosophically that at the heart of randomized evaluations is the belief that relatively small actions ...

ED: can make a difference. I don’t know if the microcredit studies are the best-known set of results. Hopefully the ultra-poor evaluations will also become famous. And the GiveDirectly evaluation got a lot of play.

I think the reason why the microcredit evaluations have gotten so much attention is because microcredit is so big. So having something to say about it is important. If it had been positive, it would have gotten a lot of play too. It’s not just the counterintuitive results. Many things do work so that’s good too, and those things will get attention.

TO: I’m often discouraged by the reaction to programs that show improvements but small ones. It seems that there’s some threshold that’s quite high that people expect programs to meet. Regardless of cost effectiveness the programs have to be “transformative” or they are a failure. One example is the CCT program that was tried in NYC,^{17[18]} had positive and significant results, but they were small, so the program was essentially canceled.

ED: I think that’s a bit of problem that we encounter as well. We have that problem with the immunization incentive program.¹⁸ That’s

16. What Duflo describes here is remarkably similar to Lant Pritchett’s (chapter 9) critique of the RCT movement—that evaluation is coming from outside and therefore likely to be resisted, rather than evaluation coming from inside as part of an organization’s efforts to learn and improve.

17. The Opportunity NYC program was a privately funded CCT experiment in New York evaluated by MDRC. Several of the program components were found to have small, but statistically significant, effects and to be cost effective, but the program was largely discontinued.

18. One of the early randomized evaluations to gain attention was a program in which parents were offered free lentils and a metal plate in return for immunizing their children. It had a much larger effect than improving the availability of vaccines. The study is frequently cited in behavioral economics literature as an example of people acting “irrationally” at least in terms of economic models of fully rational actors.

something I care deeply about just from a human perspective. I want to get that expanded. But even though we’ve never seen anything better in terms of lives saved per dollar spent, there’s a lot of resistance. There are many reasons why people are reluctant about things like giving people an incentive for immunization. Some are ideological: “you shouldn’t pay people to do something that’s good for them.” But some people say, “oh, that’s good, you increased the immunization rate from 6 percent to 38 percent but there are still 60 percent who are not immunized.” My response is, “well yeah, do you have something better?” We published that work in the *British Medical Journal*.^[16] One of the referee responses was, “that’s not enough to guarantee herd immunity, so why do we care?” Well we care because it’s still good even if it’s not herd immunity. We care because the 30 percent of people aren’t going to die. And even if you don’t have herd immunity, you still get contagion effects. All of the evidence on vaccinations is that you get effects [of lessening epidemics] at all levels [of vaccination coverage].

Of course, if you get to 100 percent, you get the big prize of eradicating a disease, like Gates is trying with polio, which is a completely legitimate exercise. But it’s not like you shouldn’t do anything if you can’t fully wipe out a disease. There’s no reason not to improve measles immunization from 6 percent to 38 percent. But people don’t get all that excited about it.

By measuring well and showing what the effects are, you get these kinds of realistic answers. When you don’t measure, you can always claim something big.¹⁹ Starting from a context where most things are at least a little bit overblown, then the sobriety of measurement puts any program at a bit of a disadvantage. But I think that’s a culture that could change, that people could get used to using a different lens. The honest but modest number that has been established from a trial or in another rigorous way could replace the rosy description of what you wish was happening, but that change might take a little longer.

TO: Have you seen progress in the number of people willing to take on evidence or the ability of evidence to make change?

ED: It’s a revolution. It’s not progress. It’s like a sea change really in the last few years. Just look at the number of projects. J-PAL has 600 ongoing or completed projects. And that’s just J-PAL researchers. That’s the tip of the iceberg. Then there’s the whole of IPA, the World Bank, who are doing their own thing, and there are many people who are not affiliated with any

19. Lant Pritchett points out that this is a reason why many organizations avoid RCTs.

of these organizations who are doing their own projects. Even the fiercest critics of randomized evaluations have tried it.

So that's in terms of the supply. Of course, there wouldn't be a supply of projects if there weren't partners to do these projects. They require partners.

My experience with working with governments is now we're able to work much more closely with them. At the moment, they are more willing to experiment than to take up new things based on other experiments. That's fine by me. You can see that there is a new culture of trying things out, and acting on the results, that is spreading much faster and farther than I would have predicted a few years ago. A lot of things have accelerated in the last two or three years.

AB: I guess, yes. Now, at least in India, the media is more conscious of the quality of evidence question. There's now an online magazine called *India Spends*, which is essentially focused on presenting evidence. The public does participate in the evidence debate. Now the next step is to persuade them that some of those people who are talking about evidence are corrupt or wrong, and sometimes both but mostly wrong. I think they are just misguided, though I wouldn't say that about everybody.

I think there is a need to get back the idea that all evidence is not equal.

You often still hear someone saying, "oh, some anthropologists found something else." But you go look at those papers and they had no control group. The battle for evidence is not a trivial one. It's easy to get confused about the fact that there are different ways of collecting data, and then there's the fact of whether there's a control group. With no control group, you have no idea what you're comparing to, but that's a totally separate issue from the ways you collect data. I think people often get confused that somehow through anthropological methods you can avoid the need for a control group. That's just a logical fallacy. You could collect better data maybe or different data, but the fact that you are interviewing people who are participants and some other people you might meet who are not participants to use as a comparison is not a matter of economics or anthropology, it's a matter of logic.

There is a lot more conversation about evidence in the world now, everywhere. Therefore there are lots of legitimate and illegitimate claims going on. But I think it's better to have that conversation than to be in the world 20 years ago when there was no conversation whatsoever. People ran cross-country regressions and made claims. That particular version of analysis, I think, has been successfully pushed off the table.

People at the table are making claims at roughly the same level of evidence, but it's still difficult for nonspecialists to understand because there are so many claims. You hear things about mixed methods, or whatever, and a lot of those claims are about not understanding what the words mean. Some of them are legitimate. A lot of them are not understanding what the words mean. They're just noise. I think there's a fair amount of that.

TO: Do you feel like there's been a significant shift in the academic community about both the quality of evidence and replication issues?

AB: Absolutely. To a first approximation, with some very specific exceptions of high-quality work on specific US programs, the level at which people understood identification was weaker 25 years ago than I think the median MIT undergraduate today. People ran regressions any which way, called whatever they wanted instruments, and made any claims they wanted. The standards in the academic community are very high, perhaps too high, but certainly not static. They have improved enormously. That's a change since I was a graduate, which was a long time ago but not the last millennium, just the past century.

TO: Is the war over RCTs in economics being won? Are you having to spend less time explaining, justifying, fighting, the internecine battles within the economics profession about the limits of RCTs?

ED: Abhijit and I disagree on that. I think it's been completely won in that I think it's just happening. A lot of people are doing it without us. It's being used. I think it is now understood to be one of the tools. The argument within the economics profession had two main consequences, both good. First, it raised the profile. If something was debated, people began to believe it must be significant.

Second, it did force us to answer the challenges. There were a lot of valid points that were raised and it forced us to react. We've become more intelligent as a result. I think Abhijit somewhat disagrees. I think he sees very prominent people, like Angus Deaton who came out so strongly against RCTs, have provided cover for people who were against evaluation to have a very quick answer, "rigorous evaluation is not all it has cracked up to be." My view is that if people were not enthusiastic, they were not going to be enthusiastic one way or the other. It's just a front that argument takes.

I think, on balance, RCTs are a useful tool and people realize that, and therefore it will continue its life.

I don't think you can do something that is important and changes the world without meeting some opposition. I also think that's a good thing. We wouldn't want to replicate the microcredit arguments, but our whole view is that things should be questioned and I think that applies to us well.

AB: I am less certain that it has been won. The acid test of whether an idea has come to stay is that it becomes something that no one needs to justify using. This has happened first to game theory and then to behavioral economics during my years in the profession. RCTs aren't there yet: it is true almost everyone is doing them, but many of them are taking the trouble to explain that what they do is better than a "mere RCT." We need to get to the point where people take RCTs to be the obvious tool to use when possible to answer a particular class of empirical questions

TO: You've been involved in at least one of the studies in the few areas where there have been a bunch of RCTs, like microcredit, the ultra-poor programs, the teacher assistants.

With the microcredit evaluations there was no *ex ante* coordination, the education studies have a lot of common authors, but I don't think those were planned as replications, and then the ultra-poor program evaluations that were very explicitly planned as a multi-site replication. Should we think about the evidence that emerges from those varied approaches differently?

ED: I think the gain of coordinating is not tremendous, but it helps ensure that you are evaluating the same thing. With the ultra-poor evaluations, even though the programs are not the same from place to place, we do know that they are a conscious adaptation of a single model. It was a nice thing to be able to say that while it's not the same program, the way in which it is not the same is a completely endogenous response to the variation in the context but the same original model.

In the case of the different microcredit programs, while all those organizations are looking up to Grameen, it's not that they coordinated among themselves. So we were lucky with the microcredit evaluations—maybe not lucky, it's not strange that many different researchers were interested in the same questions when it comes to microcredit. In the ultra-poor program, the outcomes were planned to be the same. With the microcredit evaluations, it could have been more of an issue.

AB: I think the benefits of the coordinated evaluations, it's not that you have to do exactly the same thing. In the graduation projects, you have a

core that is almost identical and then a periphery of different things. The Ghana project in which I've been involved is different in many of the additional treatments than the ones in other countries, and you can learn a lot from those variations. But it's also useful to have a core that's really comparable because we're not going to have 30 high-quality studies very quickly. When you have five or six, it's still useful for them to be pretty similar in one core. So I do feel that the coordination was very valuable.

ED: On the learning level it's different because the studies are building upon each other in incremental ways. They're not like a replication of each other. They couldn't have been done in parallel. There is so much learning that is going on from one to the next. The next wave of ultra-poor work is going to be building on these first eight studies. There's going to be a bunch of ultra-poor work coming up over the next few years. It's going to be merging, I think, with the unconditional cash transfer work. A lot of the questions are going to be around just giving cash to the poor, and what things around it make a difference.

TO: Should one approach to those replications be more convincing? Does the fact that while the microcredit evaluations were all so different, and done by different teams without any *ex ante* coordination, but have such similar results convince us that we've narrowed in on the truth?

AB: What you said is exactly right. Imagine that we found, which we couldn't have guessed, that the results were different, then at that point we would have had a hard time knowing what to make of it. If there was an essentially identical design and we had different results, I think we would be much more willing to attribute that to the fact that the program itself had different impacts.

In other words you can think of it as being variation in the impact of the program or you can think of it as being variation due to the fact that the programs are not the same. In principle, if you said "it sometimes works and sometimes doesn't," that's different than saying that it works in some variant and not in others. Those are different statements. I think the interpretation of the variation in the results would have been somewhat different. I think you're right that we get more confidence because the results are consistent.

But one could come back and say that, "because there were six variants, none of them has been tried more than once, how much confidence do we have?" It's true that a few outcomes were tantalizingly close to being significant in one direction or another. Some were negative and close to

significant, some were positive and close to significant, some were negative and significant but small, and on and on. If I really replicated each of those, maybe the positives would blow up more, but maybe it's the negatives that would blow up more.

The optimal design of a set of experiments depends very much on how strong your priors are. I'm writing a paper about this, so I can go on and on. If your priors are pretty weak, so you want to be satisfied over a range of priors, then you want to design a bunch of different experiments. If your priors are pretty strong, so you know what you're looking for, then in some sense, you might as well vary the treatment a lot, so you see what you get. If I know what I'm looking for and I know if I'm supposed to find something specific, and then I don't find it, I'm willing to say this whole agenda is nonsense, then I should focus right around that; I should vary right around where I think the truth is. Whereas if I don't really know where the truth is, then it might make sense to be a bit more exploratory. To suggest that even if the truth was *this*, the evidence doesn't stand up, and even if the truth were *that* the evidence doesn't stand up. Let's try a few variants so we are not too close to one view.

Depending on how tight your priors are, your experimental design could be either mostly driven by your theory or mostly driven by, "I have no idea, let me try a bunch of different things."

Now in the case of microcredit, I think our priors are much tighter, so we'll be much more interested in seeing if we can get an impact on the top group, the top 10 percentage. And we have theories on what they need, and maybe the product to try out would be quite different.

I'm not sure that's what inspired the different approaches. These are just accidental differences right now. The differences between the set of microcredit evaluations and the set of graduation evaluations are just accidental. If we had had the choice, we would have done microcredit evaluations the same way as the graduation evaluations. To a first approximation, it's easier if people didn't say "three of your studies had collateral and 3 didn't." For someone who doesn't want to believe the results, for that person the variation is useful. To convince that person, the person who has a strong prior, you want to put the studies where the prior is. If you don't, he'll say, "well 5 out of 6 of these studies don't correspond to what I think is relevant." That's the danger.

That's why it would have been better if we had been able to say, "this is your prior."

On the ultra-poor evaluations, people had the prior that poor people will just eat the assets you gave them. So the fact that all of them basically focused on that was great. But if you thought people had lots of different ideas about what might happen, then you might want to design a different set of experiments. People had very well-articulated priors, so you want to push on and test those priors.

Now we are in a very different space. Everybody has a different theory of why the ultra-poor programs work, so now we have to do much more. We're at a "let a 100 flowers bloom" stage. So the design of the next experiment has to be based on how strong and how narrow people's priors are.

TO: At the margin, where are we in terms of the value of replication or moving on to the next questions?

AB: At this point I would say replication isn't a huge priority because a bunch of replications on the things that are close to policy have been done. Important replications have happened, in education, in microcredit, in ultra-poor programs, in cash grants. And there are a bunch more in process. And now we are in a place where we know those things, but we know only small pieces on some very important ones—like what to make of microcredit—but there's a bunch of stuff where we really have no clue—like if not microcredit, what?

There's a bunch of stuff that was very tantalizing five years ago, that in three more years will be known. I feel like right now there are a lot of questions implied by the research that's happened. "If not microcredit, what?" is one of them. I think there are lots of pointers toward "what." Another is, "why is it that we haven't succeeded using computer technology to teach better?" I think computer technology in education is one where there are so many claims, so much money, so much energy, but zero effect. There is one successful RCT we did 15 years ago and that's it.^{20[17]} That's the extent of all successful RCTs in that space. That's a shocking fact.

Probably the biggest benefits right now would come from research on a set of things that are close to green field research. There are a bunch of things that seem to have now become very sharp questions that need to be answered.

20. This is a study on using a math practice computer program, very close to using flashcards. Large gains were found in basic math scores in the short term, with some gains persisting over the longer term. The paper is the same as cited in note 1 above, in relation to placing tutors in schools.