Three years ago I gave a talk at DFID called Impact Evaluation 2.0

I didn't think the talk would get around much, which is why I gave it such an arrogant title. To my enduring surprise and dread, some people actually read it.

My point then: to talk about how impact evaluations could better serve the needs of policymakers, and accelerate learning.

Frankly, the benefits of the simple randomized control trial have been (in my opinion) overestimated. But with the right design and approach, they hold even more potential than has been promised or realized.

I've learned this the hard way.

Many of my colleagues have more experience than I, and have learned these lessons already. What I have to say is not new to them. But I don't think the lessons are widely recognized just yet.

This talk is an attempt to draw out what more I've learned in the three years since that first talk. An impact evaluation 3.0? You be the judge.

These have been early days of RCTs

We kind of take opportunities where we see them

And we have learned a lot. I think the summaries sent around illustrate a lot of interesting findings.

But the evidence so far is a grab bag of different interventions

Take elections work. I look at the different results from RCTs, and it's really hard to draw out the general lessons. Trying to influence people with information seems to influence them a little.

But if I were a program manager in Liberia right now, thinking have I learned that affects the way I program around the 2011 election, I'm not sure I have many answers.

Another way of asking this question: What did I know about human behavior now that I didn't know before seeing this work?

In fact, the answers raise more questions. People seem to be influenced by small cues or info campaigns. Does this mean that the gains are easily reversed? Are we such fickle creatures?

That, in some sense, might be the important question

Question I would have for DFID: What are the strategic lessons you need to learn as an organization?

You are one of the biggest donors in the world: you can set strategic objectives and answer questions

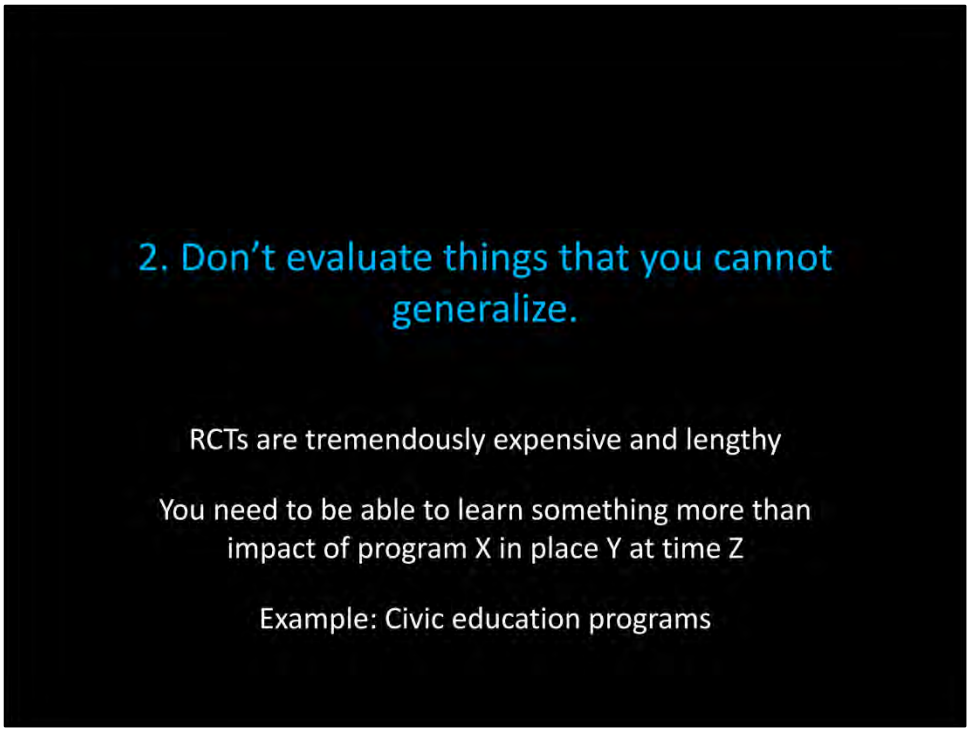
More importantly: if not you, who else?

If you leave it to the academics like me, you'll get lots of answers but not necessarily to questions that help you do your job better. A lot of the questions we ask are directly relevant to your work, and making aid better. Others answer more arcane questions. Maybe these are important as basic research, but there's a good chance they are questions that are immediately relevant.

I think you should support both—basic research in aid is important. But to tip the balance to relevant work you are going to have to drive the strategy yourself.

So: Fund agendas and big questions, not project evaluations

Where is this happening? Microfinance. Savings. Cash transfers.



Don’t evaluate things that have very little generalizability

Evaluations can take years. Will cost hundreds of thousands if not more

Need to be able to learn something more than impact of program X in place Y at time Z

Need to learn something fundamental about human behavior, or a quantity or relationship that applies broadly

As an example, let me use civic education programs

This is a photo of a civic education campaign I have been working on in Liberia

What does a civic education program or information campaign tell you?

The initial design ideas had us capturing the impact on attitudes knowledge and participation in the community. Compare outcomes in T & C

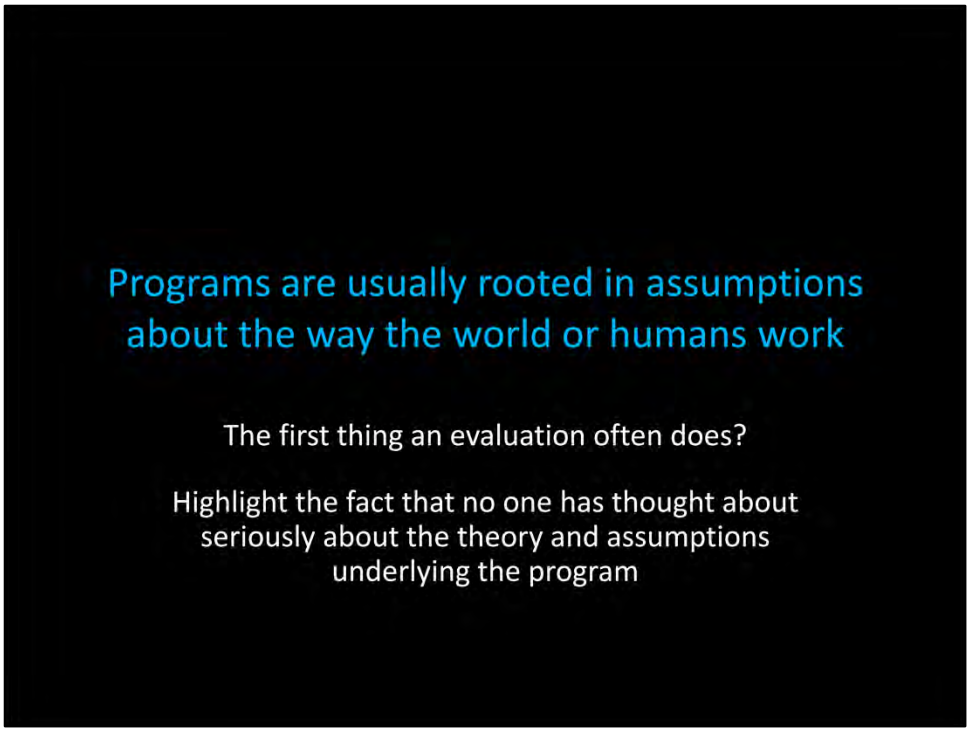
It's not clear what we learn from this. We see how well the program worked in this context. What does this tell me about whether I should push this agenda more broadly?

The discussions we had at the beginning turned to "What's your theory of change?"

Why does information seem to change people's behavior? Are people so info starved or easily influenced?

Personally I'm quite suspicious of this theory of change. It's common to a lot of governance programs: just give people information. As if information is the limiting constraint.

Let me come back to this example in a moment



First I want to make a general point

It really helps to articulate the assumptions on which the program is based

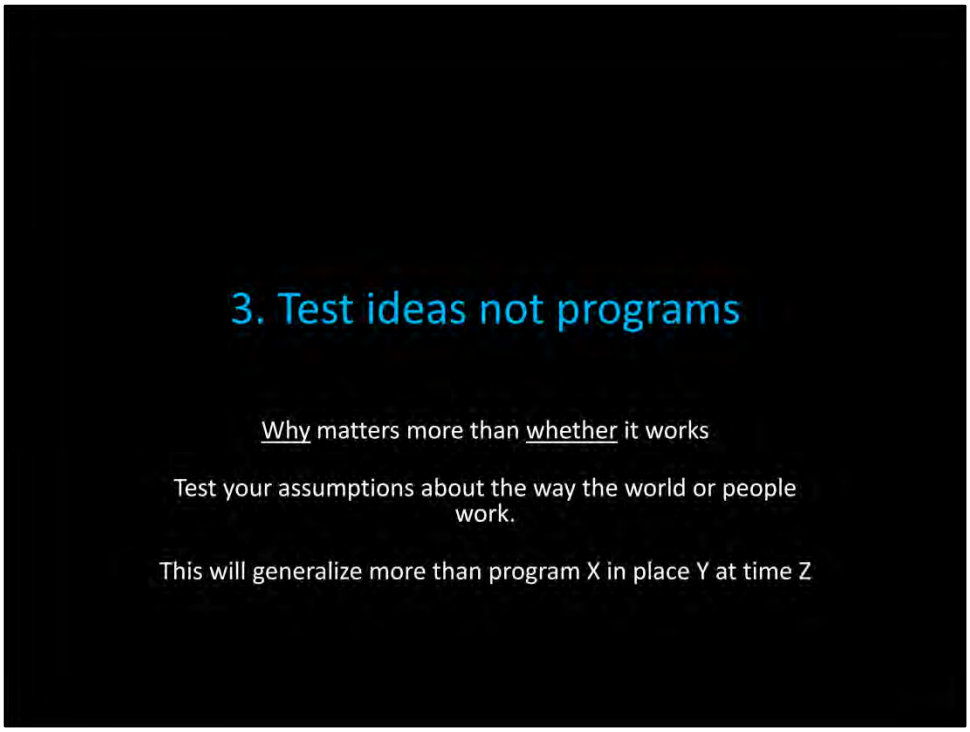
Whether it's implicit or explicit, every program is rooted in a theory of change

Assumptions about the way humans work, or organizations work, or the political systems work

Some are good, some are bad

Sometimes, when you start to articulate that in a detailed or formal way, you realize that the theory of change is not as clear as you thought

Many NGOs and programs are very clear about this, but I would say more are not. It's often superficial.



Test the idea not the program

The theory of change or the assumptions that underlie your program. Chances are this generalizes much more than the impact of program X in place Y at time Z

This means developing and testing theories

It might even mean putting theory and idea testing ahead of program objectives, at least in the short term

Let me come back to the civic education example

As we articulated the theory of change, it evolved from a “elites telling poor people what to think” to something much more nuanced

To put it briefly, the theory of change is that, in the absence of formal justice or admin systems, the rules and practices followed in the village derive from accepted norms

These norms tend to reflect the interests of older males of the dominant ethnic group

The civic education program aimed to create a new set of norms, a reference point for administration and justice that people could point to

To do so, it needed to create common knowledge

And it would have to stick.

Ideally I would have liked to experiment with different content and curricula

But what we were able to do is randomize both intensity of treatment to assess impact on common knowledge

We looked for a difference in spillovers to non-participants

We also randomized timing, so that we could look at growth or decay

Let me give another example, from a post-conflict stabilization program still underway: The Sustainable Transformation of Youth in Liberia (STYL) program.

My basic objective: Work with high risk street youth—street hawkers, eggars, homeless, petty criminals, drug dealers, junkies--in Monrovia to mitigate poverty and violence and potential for political instability

Want to test role of poverty in criminal activity, aggression and political violence

We also think there may be cognitive and behavioral roots to aggression and crime and other anti-social behavior

So we are overlaying two treatments. One is a grant for starting an income-generating activity (if they want). One is a short program of cognitive behavior therapy designed to essentially instill self-discipline and a future focus—to see if preferences and personality traits can be changed, especially the ones that can make people a harm to themselves and others

Some people talk about “nudges”. Think of this as a shove.

We randomly evaluate the impact of the two interventions alone as well as in combination.

We are varying the size of each treatment in order to understand the responsiveness of aggression and crime to these forces, and to text different models of poverty alleviation and assumptions about what drives anti-social behavior

What we learn, we hope is generalizable beyond this specific group.

Idea is that these tendencies may be generalizable to high risk youth across the region

This last program is run in a very tightly controlled way. It's not M&E, it's R&D.

We have been using the wrong letters.

I used to say to people: look, Microsoft does not go big on the market before it has really tested its product. They are the biggest player in the industry, that would be crazy.

Then I realized – actually that's EXACTLY what Microsoft does, and we suffer for it every time we boot up our computer, or get the blue screen of death.

So now I use Google as an example. They are constantly coming up with new ideas, they empower their employees to innovate and experiment, they constantly try new things in beta. They are tinkering, perfecting.

All the while, they are learning something fundamental about how human beings search and get and process info. They know more than anyone how their users tick.

They scale gradually.

DFID: You are big, your money will get spent, the programs will implement, most will be pretty good. You, DFID, can be the Microsoft of aid.

The question is: do you want to be better?

For an example of R&D, let me go back to the street youth example.

STYL is a tightly controlled project

We are self-consciously running this whole program as a pilot for scale up in the region

We started with 100 people. That's all we needed for proof of concept. We showed no adverse effect from giving cash to drug dealers and petty criminals and beggars.

In fact we saw huge drops in homelessness, drug use and crime

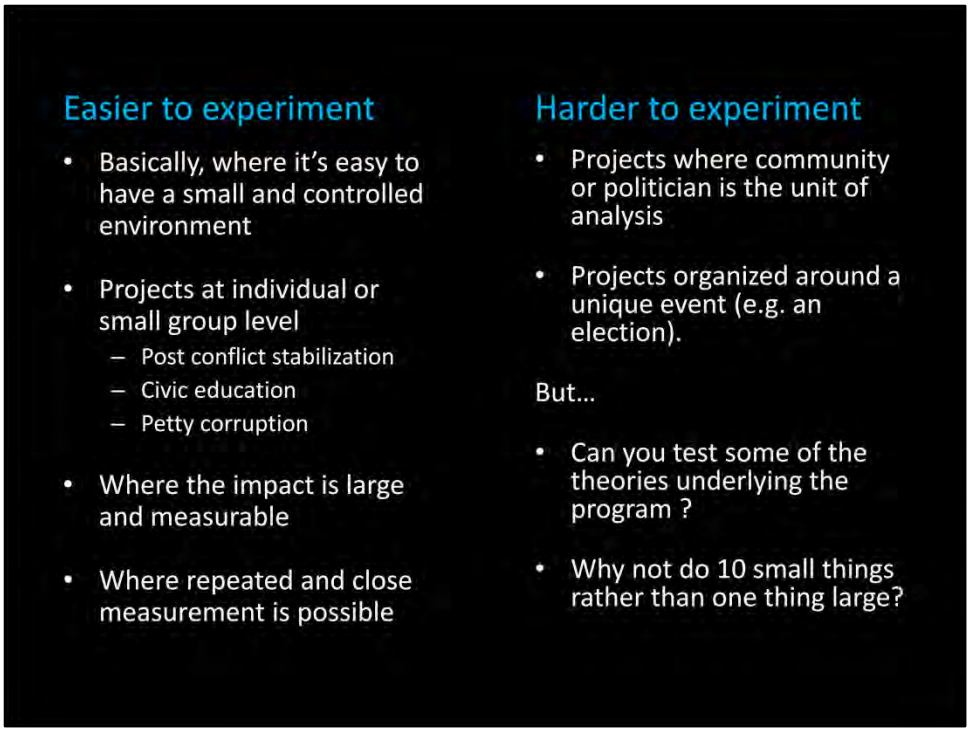
We've just scaled to 400 more. We will be able to get very good precision on impacts because we measure people at 2-week intervals three times a year. We have full time ethnographers.

We are now planning a third and fourth phase, where we will test out different sizes of economic assistance, see what happens if you add in skills training, see what happens if you provide long term behavioral reinforcement.

When we are finished, what we hope to have is a cheap, scalable, proven intervention that is ready to go to scale in Liberia or outside it.

Will it make sense to evaluate the scale up? Why not? Replication is important.

But, if it meant crowding out the time or money or intellectual energy to do R&D on the next project, on the next idea, then I would say: Do the minimum you need to test replication and then focus your scarce resources elsewhere



Some people are thinking "that won't work for my kind of project"

And you are partly right. But probably only partly.

It will work where you can get precise answers without going to huge scale – so anything that works with individuals or groups

If the impact is large enough, it's easy to measure in a small group as well.

This makes it harder to do with a community level experiment, or something where timing around an event like an election is key

But I guarantee you: in every one of those cases, if you wrote out your assumptions about the way that humans or organizations or groups or systems work, I could find CRUCIAL assumptions that could be tested on a small scale

That is, important aspects of your idea and program could be honed and perfected until we know something more about how aid works, and also have a terrific program

My fifth and last message is that you must embrace failure

I read some of the reviews that were circulated. Every one said this program had this impact and that program had that impact

Not one said "and this program had no effect at all" or even "this program affected this and not that"

I exaggerate a little, but not much

What this probably means is that
a) People are not finishing failing studies
b) People are not publishing failed studies
c) People selectively report what works
d) We are all attuned to remember what works, not what fails

I think the answer is e) all of the above. This is getting better.

But frankly I can't imagine anything more damaging to a program designer than not knowing what hasn't worked elsewhere

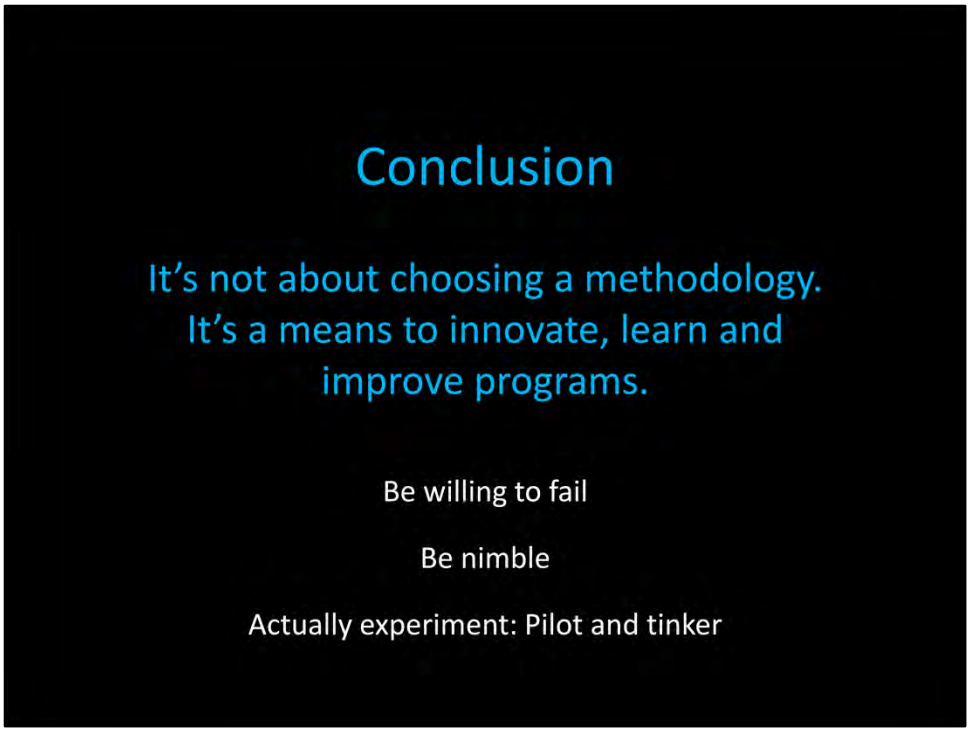
The implication is an organizational shift that will be very, very difficult to overcome

Small things you can do to make it better

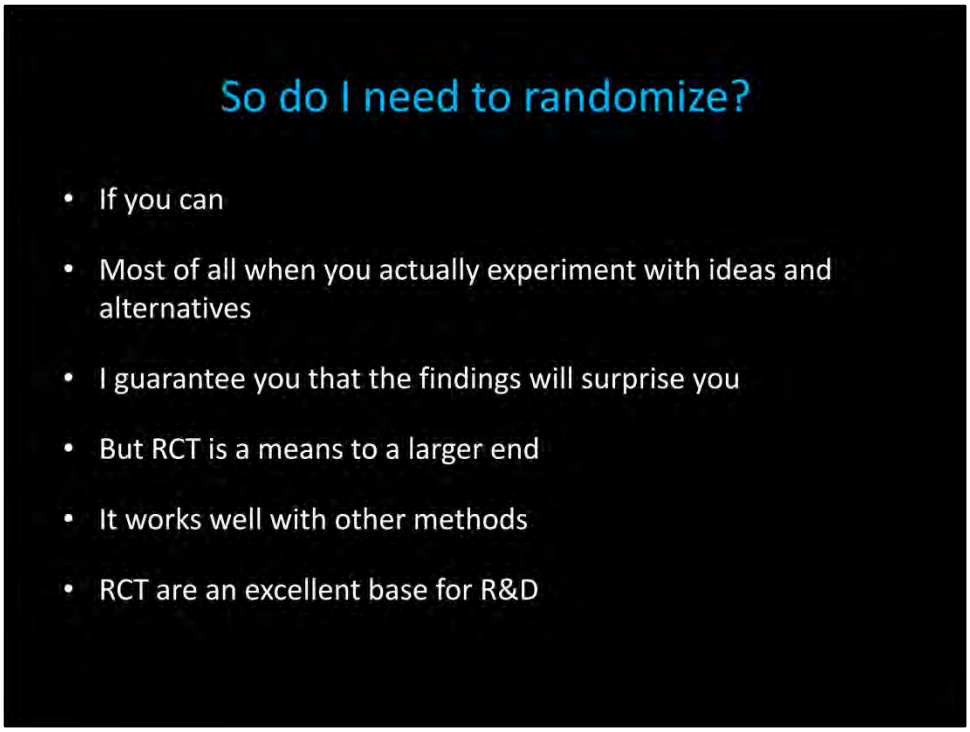
a) Prespecify your theory, tests and outcomes
b) Actually report which ones had impact and which did not
c) Celebrate and communicate failures

Do remember that absence of evidence is not evidence of absence. But take absence of evidence seriously.

Resist the natural human urge to focus on the positive impacts. That is the path of Microsoft.



Ultimately, do not focus on RCTs the methodology. They are a means to an end.



My parting shot:

You are a big organization. You have a lot of money to spend. Most of your programs are fine. You are going to put aid and governance programs out there and they are going to be impactful, even if they don't work perfectly.

That is, you have it within your power to be the Microsoft of the aid industry.

I'm suggesting that you could actually do better.

Maybe you won't be the Google of aid, but surely you want to move in that direction?