A Political Economy of Aid Effectiveness

James Morton

1. Introduction

This paper attempts to summarise how current debates on Monitoring and Evaluation and aid effectiveness look to someone whose first M&E assignment was in 1982. It has been prepared for Mokoro's 2009/2010 seminar on Monitoring and Evaluating Complex Development. An overview of the current thinking is derived from the references cited in the background material which Mokoro prepared and used as a basis for what might be called a practitioner's response to that thinking.

Many years ago, quite by accident, I found I liked driving around strange places in landrovers. So I took an MSc in Development Economics. I learned about welfare economics, shadow prices, and the Accounting Rate of Interest needed to create equity, not just between different social classes but also between current and future generations. [1] It was difficult but rigorous; and it was the foundation of what Mokoro describes as the '*previously ubiquitous, project-focused, quantitative measures of impact (including ex-ante cost benefit analyses)*'.[2]

I then went to Sudan as Head of Planning, Monitoring and Evaluation on a large, innovative project designed to prevent a Tragedy of the Commons in the crop and range lands of Darfur. For that, I was armed with books by Price Gittinger, Casley and others, which told me how to carry out randomised sample surveys and set a baseline for subseqent evaluations.[3] And I did them: 700 farm households, 300 pastoralist households and more. However, in my first meeting with the World Bank team responsible I was kindly told that it was time to start planning the next phase of this pilot project; it was already too late to worry about whether the pilot was working or not. Nobody could wait until I had finished the baseline surveys and analysed them properly, let alone do an impact assessment. When, subsequently, I used all the theory I had learnt to point out that the Phase 2 designers had grossly overestimated the likely benefits and rates of return, my colleagues and I were told, quite forcefully, that our attitude was not helpful. We did not win the contract for Phase 2.

An early, brutal lesson in the political economy of monitoring and evaluation. Mokoro's notes suggest that project-focussed, quantitative approaches to impact measurement were abandoned because they embodied '*objectives defined by outsiders.*' My case was otherwise. The quantitative M&E approach was rejected precisely because it conflicted with an outsider objective of justifying a follow-on project. I will argue later that effective M&E almost always conflicts with the immediate objectives, or at least the incentives, of the different stakeholders in the aid process, be they insiders or outsiders.

---

[1] IMD Little & JA Mirlees, Project Appraisal and Planning for Developing Countries (1974); and P Dasgupta, A Sen & S Marglin, UNIDO Guidelines for Project Evaluation (1972) are fundamental references.

[2] Mokoro, 2009 – Monitoring and Evaluating Complex Development, Day 1 Agenda

[3] J Price Gittinger, Economic Analysis of Agricultural Projects (1982); DJ Casley & DA Lury, Monitoring and Evaluation of Agriculture and Rural Development Projects (1982); DJ Casley & K Kumar, The Collection, Analysis and Use of Monitoring and Evaluation Data (1988)

Yet, in the end, perhaps the project management and the World Bank were right. Because it was only some years later, after a lot of thought and research, that I was able to demonstrate, rigorously, that the basic design of the project was flawed. Clearly nobody could wait four years in some kind of limbo, while the M&E and research process gestated its way to an answer.

Since that first assignment, I have headed evaluation units on two more large projects, in Sudan and the Dominican Republic, and helped to set up M&E systems in Bangladesh and Indonesia. I have led end-of-project evaluations in the Gambia, Nepal and Zimbabwe. Much more recently, I have just completed a cycle of evaluations in Pakistan and S. Sudan.

This career followed the trajectory of M&E as it developed. It started with agriculture in the 1980s. Since then it has extended to cover pretty well everything in rural development, from health to water and electricity supplies, and increasingly to local government. It started on projects managed by national government agencies, but it has ended on programmes implemented by international and national NGOs. Like M&E, it started with rigorous Social Cost Benefit Analysis and randomised sample surveys. It has ended with a broader but not necessarily more coherent selection of participatory case studies, quasi-experimental sample surveys, perception/attitude studies, institutional maturity studies etc. Midway through it all, of course, we welcomed that delightful invention, the logical framework. Now taken to new peaks of structure and logic in the latest DFID prescribed formats.[4]

My paper tries to do two things. Section Two looks at the technical side of M&E and the latest thinking on complexity. Sections 3 and 4 review discussions of the 'evaluation gap' that has been indentified in the last few years and issues of political and institutional economy around the new Aid Effectiveness agenda. An attempt is made to analyse the underlying incentive structures which have created the need for development M&E, a somewhat odd institution; without paradoxically creating any sincere demand for the rigorous impact assessments which M&E is designed to produce. These remarks do not represent striking new insights, but rather the results of one practitioner's reflections over a long career in monitoring and evaluation. How the aid industry handles knowledge, and why, is a topic which I first attempted in 1994.[5] It is one I keep coming back to.

2. Simple is the Language of Truth (*Euripides*)

Development is infinitely complex. That is its challenge and its fascination. In different ways, much of the literature Mokoro presents is based on the proposition that M&E has failed because it cannot handle this complexity.[6] In essence the argument is that complex systems can only be mastered by using more complicated methods. The most basic tools of scientific progress, simplication and abstraction, will not be enough.

> *"The fact that the world is not predictable, linear and orderly means that interventions based on simplifying fictions – such as neoclassical economics – often simply don't work. ..... complexity science offers a set of useful,*

[4] DFID How To Notes: Guidance on Using the Revised Logical Framework (2009) and Standard Indicators (2009)
[5] JF Morton, The Poverty of Nations: The Aid Dilemma at the Heart of Africa, pp 33-46 (1994)
[6] See books/papers by Byrne, Ramalingam & Jones, and Kotvojs & Shrimpton cited in Mokoro's introductory paper for seminar and Mokoro's presentation on Seminar Day 1.

> *challenging 'fictions' about the world, which can enable us to better delineate and understand complexities of the real world."* [7]

This literature draws on ideas from social science in developed countries and advanced work in natural sciences. How it can be applied and made relevant for M&E work in developing countries is not stated. The comprehensive ODI review cited suggests only that the concepts of complexity science may be '*potentially very valuable'*. In other words those concepts remain a long way short of a useable of set of methodologies capable of overcoming the weaknesses of development M&E. The ODI paper is also honest enough to admit that there are some who see complexity science as little more than snake oil.

However, the central argument of this paper is that there is an alternative explanation for M&E's failure. It is not the complexity of developing countries which makes M&E so difficult, but rather the special nature of international aid and the incentive structure which goes with it. The simplifying tools of established economic and social theory have not failed. It is just that they have rarely been effectively applied. And that in turn is because they tend to give results which conflict with the '*objectives defined by "outsiders"'.* In other words, those simple, standard tools have often been shot down not because they were wrong, but because they delivered the wrong message. Conversely, there has been a strong incentive to encourage research into new tools in the hope that they will get the message right.

A look at the stated objectives of international aid helps to make the point. None of the Millennium Development Goals is particularly complicated. But let me take the simplest as my text.

> **Education: Ensure that, by 2015, children everywhere, boys and girls alike, will be able to complete a full course of primary schooling**

This seems like a pretty linear proposition: enough schools plus enough teachers = enough school places. Yes, there are questions about teacher motivation, proper administration and the inability of poor families to send their children to school. But these are straightforward researchable issues, with no power functions or obvious discontinuities to make things difficult. If teachers do not want to teach or pupils do not want to learn, standard economic analysis will almost certainly reveal why.

In a practical world, large, complicated structures and machines can only be built by assuming that the behaviour of component parts is 100% predictable. When designing a car, it has to be assumed the pistons will go up and down. When designing a school programme, it has to be assumed that teachers will teach. Modern society is built on these kinds of simple rules, on an assumption of predictable behaviour in mechanisms which may be complicated but are not in any way complex. And if the initial design fails, then some simple, rule-based solution has to be found. If the solution is not simple and rule based, then it cannot be general. If is not general, then it is not a useable solution. Making this iterative process of design improvement work should be a key M&E function. Why it rarely provides it is discussed below.

[7] Ramalingam, Jones et al, Exploring the science of complexity: Ideas and implications for development and humanitarian efforts. ODI Working Paper 285, 2008

It will be argued that this says nothing about outcomes.[8] If teachers teach, even if children study, how do we prove that it has improved their lives tangibly and sustainably. How, in other words, do we measure impact? The answer has three parts: two negative and one positive. First, there is no point trying to measure outcomes if you cannot measure outputs. I shall argue below that this is not being done effectively, although many assume it is. Secondly, measuring outcomes accurately, and attributing them reliably to the project under evaluation, is endlessly problematic. On the positive side, outputs are a robust, easily-measured indicator of what the primary stakeholders believe the outcome will be. Teachers who believe in what they are doing will teach. Children whose parents know it is worthwhile will go to school.

M&E debates are dominated by the question of measuring outcomes. Having carried out some seven major surveys, and having had the luck to work long-term in the areas where I did those surveys, it is easy to see why. I have become highly sceptical about the possibility of measuring even simple parameters such as yield accurately enough to draw reliable conclusions. This view has been reinforced by my recent work reviewing evaluation surveys in Pakistan. The following is by no means a complete list of the problems:

- Problems of selection, placement and timing bias are noted in the theoretical literature. In practice, they are widespread, but largely unacknowledged.
- The multiplicity of rural livelihoods, the near impossibility of estimating net incomes from own-production, and the uncertainties of respondent recall make estimates of the critical household income parameter highly inaccurate. Many, perhaps most, surveys skate over this problem.
- Year-on-year variability makes it improbable that a particular survey year is anywhere near the trend line.

So much, so negative. But this does not mean that I do not believe in surveys. Far from it. Surveys are a highly effective way of understanding the structure of peoples' livelihoods, the nature of the opportunities they face and their strategies for dealing with them. Without a baseline survey I would never have found out, for example, that in the Dominican Republic training fighting cocks is a viable livelihood for old men; not to mention making spurs for fighting cocks out of plastic toothbrushes. In other words, as a way of understanding complexity, and simplifying it in useful ways, surveys are number one. Better, I believe than case studies because of the difficulty of structuring case studies so as to be truly representative. The Pakistani programme I am currently working with has invested in substantial baseline surveys. They contain a great deal of information about the communities they surveyed. Originally, however, there was no intention to mine that data for useful understanding. The baseline was going to be a baseline and nothing more. An indication of how M&E tends to be taken as a mechanical operation, separate from and not useful to the day-to-day operations of the programme.

Whether baseline surveys are useful for a statistically rigorous test of programme impact is a different question. But I shall never know the answer to that because, so far as I know, no one has ever done a follow up survey against any of the baselines I have carried out. I believe it is rare to find a full set of baseline and follow-up

[8] I hope to avoid semantic discussions about whether outcomes = impacts or not. For this paper, outputs are what a programme or project delivers – education say – and outcomes/impacts are any improvements, or disimprovements, in peoples' lives and welfare that result.

surveys. Which is the other problem with baselines. Few development programmes maintain the level of donor interest needed to fund big surveys for long enough to complete the series needed to measure impact against a baseline.

3. When Will We Ever Learn? And What?

Problems of attribution lie at the heart of much of the literature discussed here. For example, Contribution Analysis is presented as a technique which seeks to provide '*plausible evidence that can reduce uncertainty regarding the "difference" a program is making to observed outcomes'.* This cautious, unambitious wording reveals a belief that absolute proof of attributable impact may never be achievable. '*Plausible evidence'* is the best that can be hoped for.[9] On the other hand, it seems to be implied that this will be enough; that a '*credible picture of attribution*' will be sufficient to justify the programme to its tax-payer funders. Another paper on Current Debates in Impact Evaluation describes how the questions of impact measurement and attribution have become bogged down in a range of methodological discussions.[10]

In 2005 the international development partners signed the Paris Declaration to signal a new drive for aid effectiveness. This was followed, in 2006 by a paper When Will We Ever Learn (WWWEL), which prompted the establishment of the International Initiative on Impact Evaluation (3ie) two years later.[11] The paper's basic thesis was that '*persistent shortcomings in our knowledge of the effects of social policies and programs reflect a gap in both the quantity and quality of impact evaluations.'*

WHEN WILL WE EVER LEARN

"An "evaluation gap"has emerged because governments, official donors, and other funders do not demand or produce enough impact evaluations and because those that are conducted are often methodologically flawed."

"Governments and agencies regularly seek ideas and guidance to develop new programs or to improve existing ones, but on time frames and budgets that do not allow rigorous evidence to be developed."

"These institutions may do well in their normal data collection and evaluation tasks related to monitoring inputs, improving operations, and assessing performance, but largely fail in building knowledge, which requires studies that fall outside normal budget and planning cycles and for which incentives are sorely lacking."

"The knowledge gained from rigorous impact studies is in part a public good."

WWWEL described four parts to monitoring and evaluation:

- Monitoring, ie of outputs
- Processes, ie learning operational lessons from the monitoring results
- Accountability, to funders etc
- Achievement of Basic Aims, ie "What difference does this programme make?"

The paper states: '*that governments and development agencies are better at monitoring and process evaluations than at accountability or measuring impact.*' It

[9] F Kotvojs & B. Shrimpton, Contribution Analysis: A new approach to evaluation in international development. Evaluation Journal of Australia, Vol 7, No 1, 2007
[10] H White, Some Reflections on Current Debates in Impact Evaluation. 3ie, Working Paper 1, 2009
[11] WD Savedoff, R Levine & N Birdsall, et al. When Will We Ever Learn: Improving Lives Through Impact Evaluation. Centre for Global Development, 2006

ascribes the evaluation gap to the public good nature of impact measurement. I would challenge both these statements. Governments and development agencies may be better at monitoring and process evaluations than they are at evaluation. They are nevertheless, extremely bad at it, in a number of cases I have seen frankly negligent.

The Public Good concept is a favourite resort of academics making the case for public funding of their research. It has the politically useful characteristic of avoiding blame. No one is at fault for the 'evaluation gap' if evaluation is, by very its nature, something that will be underfunded. Comfortable as this is, there are immediate problems. First, it is difficult to argue that accountability is a public good. Why does the funding agency concerned not have a direct, private-good interest in accountability? Second, the argument only stands if you accept, as I do not, that development agencies are good at the simpler monitoring tasks.

Having effectively sidelined Monitoring and Processes, WWWEL goes on to focus, almost entirely, on the last item, measuring impact. I am deliberately avoiding any discussion of methodologies: randomised control trials, quasi experimental double-difference, etc. However, the box taken from the Current Debates paper[12] tries to capture the rather abstruse, even semantic nature of the technical debates which dominate discussion about impact measurement. As indeed, they are a central topic in Mokoro's seminar series.

CURRENT DEBATES

"There are two definitions of impact.
- …. outcomes and long term effects
- … attribution …" (ie impact = Y1 - Y0)

"Attribution implies a counterfactual .. This does not mean that there has to be a comparison group"

"A theory-based approach will usually combine methods: quantitative and qualitative …"

"the whole business of policy advice falls firmly in the positivist realm"

I believe the whole discussion can be summed up in one proposition and two corollaries to that proposition:

Proposition: That measured outcomes cannot, without attribution, prove anything about impact.

Corollary 1: That some form of programme logic, causal chain, theory of change – call it what you will – is essential for any attribution.

Corollary 3: That context and of the way it influences the causal chain must be fully understood when assessing attribution.

These propositions seem almost too trite to discuss. (Do I hear 'bleeding obvious'?) But that is the point. This is all that can be found underneath the extensive, even acrimonious discussion of quantitative versus qualitative; of the importance, or unimportance of Randomised Control Trials; of contribution versus attribution; of the evils of positivist and linear thinking; etc.

The paper Current Debates on Impact Evaluation seeks to build a middle ground in these discussions, arguing that a programme theory can provide the framework for

[12] H White, op. cit.

both qualitative and quantitive work. As I will try to show in my next section the last statement in the box, that we are firmly in the realm of positivism, is crucial.

However, this begs the question Why do we have to have such a discussion? The rest of this paper will argue that it is because both the public good argument and the methodological debates are useful red herrings; topics which attract the most attention because they help stakeholders to avoid matters which are simpler and more fundamental to effective M&E. Those simpler topics – monitoring, process and accountability - are politically difficult and, in the worst case, likely to threaten the justification for aid itself. The in-built incentive structures of development aid make it impossible to address these areas effectively.

4. Two Case Studies

A TYPICAL EVALUEE RESPONSE

"Measuring Impact and Effectiveness in an Uncertain World: Interpretive Accountability … the development process, which is chaotic and uncertain …., has been treated as a linear process, with too much emphasis on a 'model'."

"There are few agreed performance standards for NGOs ... beyond probity and some quantifiable impact indicators ..., such as service provision and economic development. No bottom line, the bottom line shifts with the situation on the ground. The non linear nature of NGO work and the need to adapt performance and accountability accordingly, is critical."

"… accountability must be a process of negotiation among stakeholders rather than the imposition of one definition or interpretation of effectiveness over another."

".. the need for evaluation must come both from the NGO and the donors and there must be a willingness to fund it … This has not happened."

The box shows some quotes from a sophisticated and very large national NGO's response to an evaluation report. This NGO was thoughtful and very open to the evaluators. Its own internal debate was lively and a signficant minority of the staff were much more supportive of the impact assessment than this. It was a pleasure to do the evaluation. However, the quotes illustrate a common line of argument. It is almost explicitly stated that rural development is so 'contingent' and strategies are so 'emergent' that no real impact assessment can be made. The line about 'too much emphasis on a model' reflects strong resistance to the theory-based approach advocated in the Current Debates and other papers. And the whole argument shows how what is apparently a technical discussion about 'positivism', 'linearity' and over-simplification can spring from a development agency's wish to soften, or even brush off, the potentially adverse result of an evaluation. The default setting is lesson suppression, not lesson learning.

The last quote is also important. This was a very large national movement spending $100 million and more of aid funds. Yet the major donor had not sought any significant impact assessment for nearly a decade.

As I have discussed earlier, the idea that simpler monitoring work was being done well was just plain wrong. The organisation had major credit and infrastructure operations, its management and accounting functions worked and good records were kept. Yet, the monitoring data was fundamentally uninformative. Results were presented as cumulative over a 20 year span, saying nothing about the current situation. Relatively simple steps would have made it possible to measure unit costs

and other aspects of efficiency. Yet neither management or donors, including Government, had seen this as necessary or useful.

A number of large, complex double-difference baseline surveys were being carried out. Yet there was a glaring M&E gap. Not the 'Evaluation Gap' identified in the When Will We Ever Learn paper but a much simpler 'Monitoring Gap': measuring outputs such as beneficiary numbers and utilisation. Despite building many thousands of infrastructure projects, little attempt had been made to survey the numbers of users, project sustainability and other simple measures of programme outputs. In the old fashioned world of cost-benefit analysis, the next step would have been to estimate a value for those outputs at social prices. Increased net farm incomes, cost savings on water collection, reduced morbidity would be examples of the kind of 'Immediate' or 'Direct Impacts' which can be measured, relatively reliably and cheaply, in this way.

My other case study is a country which has recently emerged from a prolonged war and remains one of the poorest countries in the world. Donors have committed very large funds to a number of trust funds most of which are funding international and national NGOs to deliver services, principally basic health, education and water and sanitation. I will limit my comments to saying that after four years, the standard of monitoring is a long way below the standard needed, let alone impact assessment. In this case, the funds are still principally managed by international agencies and implemented by NGOs. The lack of monitoring cannot be laid at Government's door.

5. Why Does Anybody Need M&E? – A New Institutional Approach

At bottom, the methodological debate described above is centred on a division between those who see development as too complex, 'contingent' and non-linear to be captured by the standard tools of simplification and structured causal chains and 'positivists', like myself, who believe those tools be essential to the design of workable policies and effective development programmes. Another paper provided by Mokoro takes this division into the domain of Power, Mutual Accountability and Responsibility in the Practice of International Aid.[13]

In that paper, Eyben identifies a distinction between '*substantialist'* and '*relationalist'* interpretations of mutual accountability. The former sees stakeholders holding each other to account for performance against pre-established objectives, through '*mechanisms for regulating behaviour between autonomous parties'*. The latter '*understands entities as mutable, shaped by their position in relation to others. Relational notions, married to ideas of process and complexity illuminate the messy and contradictory quality of aid relations that substantialism finds difficult to cope with.'*

I share a number of Eyben's starting points:

- That of trying to get to grips '*with the problems of power that impede effective aid'*; in particular the problems which impede effective M&E.
- The concept of a *'multiple accountability disorder'.*
- The way the '*owned process of the Paris Declaration appears to be generating self-sustaining, un-owned processual effects, such as far too many international*

[13] R. Eyben, Power, mutual accountability and responsibility in the practice of international aid: a relational approach. IDS Working Paper 305, 2008

conferences and documents which are becoming a collective embarrassment' and a suspicion that this may just be reinforcing the status quo.

> THE INADEQUACY OF 'SUBSTANTIALISM'
> The risk of managing for results .... is that it handles messes as if they were difficulties and processes, such as poverty, as if they were material substances. It ... seeks to secure a limited set of agreed, time-bound and measurable results in circumstances of self-generating, non-attributable processes of complexity in which any one of us can only have a partial view because of our relative subject position, made worse by power that privileges some viewpoints to crowd out others.

Where I part company is over causes. As the statement in the box illustrates, the relationist analysis uses similar language to those who reject 'positivist' methodologies. Aid relations are too complex, contingent and non-linear – too '*messy*' - to be managed by straightforward accountability mechanisms. My alternative hypothesis is that aid relations are, in fact, rather simple and clearly understandable in terms of the underlying incentives faced by the different parties. The widespread '*accountability disorders'* do not reflect any special complexity or imbalance in power relations. Instead they are the natural result of the way the incentives in favour of the status quo, of weak, even non-existent accountability mechanisms, are far stronger than the incentives to make a reality of the Paris Declaration's objectives. The rest of this section seeks to explain why.

The question at the head of this section seems naive. Why does anybody need M&E? But it helps to underline how development aid is a unique activity, not directly comparable with any other. Do BP or Unilever have M&E departments? Obviously not. Their objectives – to deliver product at a profit – are measurable by standard management and accounting procedures. Well what about government? Do first world ministries have M&E units? Not really I think, although moves towards evidence-based policies do parallel much of what M&E does. Why, then, does development aid need this unique institution?

A second question springs from the previous discussion. If, as argued above, the 'Evaluation Gap' is not just because impact assessment is a public good; and if there are Monitoring and Accountability gaps as well: Why is that?

Part of my answer to these questions is, as it has been since the early 1990s, that aid is a unique economic good: a first world consumption good from which the benefit, the economic utility, comes in the form of the good feeling which results from the act of giving and at the time of giving. That good feeling would be lost if it were ever learnt that the money was wasted, but the cost of checking is high and the value to the giver is low.

The result is that aid is perhaps the least accountable operation the world has ever seen. As I put it in 1994, "*the spiralling levels of academic debate on aid testify to the difficulties of definitive measurement*." However, "*provided that debate continues without a clear outcome the giver – the first world taxpayer and donor to NGOs – is willing to give aid the benefit of the doubt. .. it costs him so little and even if it is only marginally useful it is still worthwhile*."[14] The imprecise language used in the recent Contribution Analysis paper seems to confirm this point. Attribution is elusive and the best that can be hoped for is measurement with '*the aim of reducing uncertainty about the contribution made, not proving the contribution made'.*[15] Hardly a strong

[14] Morton, op. cit.
[15] Kotvojs & Shrimpton, op. cit.

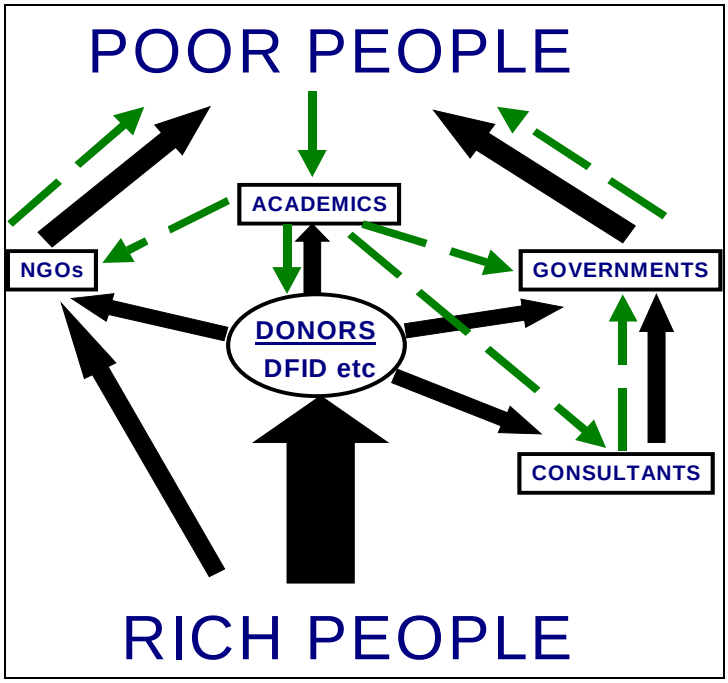
statement, but it seems that 'reduced uncertainty' is sufficient for aid donor purposes. Proof is not required. The question of how much uncertainty is too much, is left hanging.

The second part of my answer draws on two concepts from new institutional economics: the Principal : Agent problem and Time Inconsistency. A separate concept, that of goal inconsistency, is also important.

The Principal : Agent problem is fundamental. In commercial companies shareholders struggle, as Principals, to make sure their Agents, the managers of the company, look after their interests. In developing country agriculture, landowners face the same problem getting their sharecroppers do a good job. Asymetries of information mean that managers or sharecroppers know more about what is going on than shareholders and landowners, allowing them to serve their own own interests, not those of their Principal.

And that is just between one set of principals and agents. Figure 1 tries to map the Principal : Agent problem for the first world taxpayer (RICH PEOPLE - The Principal) who supports aid or contributes to an NGO in order to help POOR PEOPLE. The black arrows represent the flow of funds and the green arrows the flow of knowledge about how poor people live and what they need. The different organisations boxed in the middle of the figure are the various Agents who are expected to implement the Principal's wishes.

Figure 1



Even on a simple, bilateral aid programme there can be as many as three, or even four agents between the principal – the first world taxpayer – and the point at which aid effectiveness can be measured in terms of a better life for poor people. And none of those agents' interests are strongly aligned with the Principal's desire that his aid be used wisely. Aid agencies, consultants and NGOs all justify their existence in terms of results, but like any organisation they have to balance this against the

private interests of their political masters, their management and staff, and the organisation's interest in its own survival. And of course, things become much more complex when it comes to the recipient country's government, which itself comprises a complex set of Principal : Agent relationships.

It might seem that this proves Eyben's point, that aid relations are too complex to be managed through simple Principal : Agent contracts. She draws a distinction between three situations: complicated '*(cause-effect knowable with expert input)';* complex *'(cause-effect coherent in retrospect but not repeated)'*; and chaotic *'(no clear cause-effect pattern)'*. She suggests that substantialist approach is designed for complication, but not for complexity. There is no obvious reason to believe that the network of aid relations charted in the figure should be any more difficult than complicated.

The answer to the question - Why does anyone need M&E? - falls out quite simply from this. M&E is designed to give the aid donor Principal assurance that his Agents are working for him as they should; the same service as audited accounts provide to a company's shareholders. But there are crucial differences between BP's accounts and M&E reports on an aid project. First, a shareholder in BP still expects to make dividends and capital gains from his shares. So he has a real, personal interest in reading the audited accounts. If the company fails he will suffer directly. An aid donor, by contrast, expects nothing more from his money once he has given it.[16] Second, the share price provides a real-time, every-day measure of the company's performance and a strong incentive for large numbers of journalists and analysts to search for information about what the company is doing. Crucially, bad news is just as valuable to those searchers as good news, possibly more valuable. Lastly, the share price, and investor interest in that price, puts real-time, daily pressure on management to perform.

None of this is true for an aid programme. And this gives the answer to the second question: Why do the evaluation, monitoring and accountability gaps persist some thirty years after aid agencies started to put emphasis on impact assessment? At bottom there is no real control on Agent performance and no strong alignment between their interests and the Principal's wish, as the original giver, to see a better life for poor people. The pressure to prefer good news over bad is particularly strong. The result is a strong consensus for a nicely judged level of M&E, sufficient to give the minimum assurance that aid really does work without raising hard questions. To investigate those too closely might make aid work better, but it might also show that it is not working and should be stopped. This is not something any of the agents are prepared to risk.

The time inconsistency issue for M&E relates to the fact pointed out in the When Will We Learn paper that everybody's interest in good data is highest at the moment when a new project is being designed, the point when funds are being committed. As my experience in Darfur in 1982 showed, if M&E analysis is not complete at that point, then life moves on without it.

[16] Many years ago I suggested that the World Bank's work would improve immeasurably if it acted like a real bank and took the risk of not being re-paid for any of the projects it had sponsored which failed. Instead, the WB is, with the IMF, the only institution in the world which has its loans guaranteed even in the event of a sovereign debt failure.

A quote sums up the last point: Goal Inconsistency. "A wide range of inconsistencies can arise during requirements engineering as goals and requirements are elicited from multiple stakeholders. Resolving such inconsistencies sooner or later in the process is a necessary condition for successful development ...." This quote comes from literature on software engineering but it is highly relevant to development aid, which is riddled with inconsistencies between the requirements and expectations of different stakeholders. The Paris Declaration emphasises that effective aid reflects an alignment between the policies of the donor and recipient partners. In practice, however, most aid programmes are built on an unstable, lowest-common denominator agreement between the major stakeholders.

And this brings us back to time-inconsistency. On both sides of the development aid agreement, time-inconsistent pledges are sincerely made during negotiations. However, once the financing agreement is signed, both parties face strong incentives to break those promises.

6. One Last Inconsistency

There is one last inconsistency, one that is largely unique to aid, which can be called Role Inconsistency. I use this to describe how most of the intermediaries working between first world givers and poor people in developing countries do not, in fact see themselves or each other as agents. Almost all of them combine multiple roles: as political partner, beneficiary and agent all at the same time. Developing country governments and civil society, in particular, see themselves as independent actors, with their own entitlements and constituencies. They resist, with considerable success, attempts to hold them accountable for their performance. Even first world NGOs have a strong political position, and lobby forcefully to maintain that position.

Which brings us to one last difference between the management of BP or Shell and the Agents who work to deliver development aid. The Principals of a commercial company, the shareholders, hold and quite frequently use a final sanction over the managers: dismissal. In aid, there is only one equivalent sanction, which is to stop the funding. But that is a nuclear option which leads to mutual destruction. Whether it is the World Bank, DFID or an NGO, stopping funding is a threat to the organisation itself. And even for the first world giver, the Principal in my little model, stopping funding is to deprive himself of the pleasure he gets from giving.

Aid goes in waves. There was a downswing in the 1990s, sharply reversed in the decade since 2000. Although economic circumstances have an impact, these swings also reflect waves of relative optimism and pessimism about development aid. In the end, the first world giver does begin to suffer from donor fatigue and a nagging feeling that if the evidence about impact is still uncertain after a number of years, something must be wrong. There are signs that a new wave of aid scepticism is nearly upon us. Books with titles like

Ensuring DFID has robust qualitative and quantitative information to be able to report to the UK public the results achieved with taxpayers' funds, to strengthen project management capacity among partners (demonstrating what success looks like), and to provide evidence of progress to stakeholders.

Ensuring the logframe contains all the necessary detail against which DFID and its partners can monitor project progress as well as measuring and evaluate impact.

Dead Aid (Dambisa Moyo) and Aid and Other Dirty Business (Giles Bolton) are being published. Older aid sceptics such as William Easterly are getting more of a hearing. The box quote from DFID's latest logframe guidance, and indeed the extraordinary complexity of the new logframe format, shows the degree of concern, at the highest level, that results can be demonstrated.[17]

My own recent experience at much lower levels leaves me doubtful whether that concern will be translated into effective impact assessment, still less into the kind of monitoring and accountability that is needed to have any impact in the first place. Until the inconsistencies and misaligned incentives I have tried to describe are faced head on, the temptation to concentrate on the easy 'red herring' topics of theory and methodology, and on long-term impact assessment rather than immediate delivery, will be too powerful.

[17] DFID How To Notes: Guidance on Using the Revised Logical Framework (2009)