# Making predictions of programme success more reliable

Nancy Cartwright (Durham University), Lucy Charlton, Matt Juden (University of London), Tamlyn Munslow (Durham University) and Richard Beadon Williams (Durham University)

# Making predictions of programme success more reliable

## Nancy Cartwright (Durham University), Lucy Charlton, Matt Juden (University of London), Tamlyn Munslow (Durham University) and Richard Beadon Williams (Durham University)

Nancy Cartwright
Centre for Humanities Engaging Science and Society (CHESS), Department of Philosophy, Durham University, 50 Old Elvet, Durham, DH1 3HN, UK
nancy.cartwright@durham.ac.uk

Lucy Charlton
c/o Centre for Humanities Engaging Science and Society (CHESS), Department of Philosophy, Durham University, 50 Old Elvet, Durham, DH1 3HN, UK

Matt Juden
SOAS, University of London, Thornhaugh Street, Russell Square, London WC1H 0XG, UK
mj28@soas.ac.uk

Tamlyn Munslow
Centre for Humanities Engaging Science and Society (CHESS), Department of Philosophy, Durham University, 50 Old Elvet, Durham, DH1 3HN, UK
tamlyn.j.munslow@durham.ac.uk

Richard Beadon Williams
Centre for Humanities Engaging Science and Society (CHESS), Department of Philosophy, Durham University, 50 Old Elvet, Durham, DH1 3HN, UK
richard.b.williams@durham.ac.uk

**About CEDIL**
The Centre of Excellence for Development Impact and Learning (CEDIL) is an academic consortium supported by the UK government through UK Aid. The mission of the Centre is to test innovative methodologies in evaluation and evidence synthesis and promote evidence-informed development. CEDIL-supported projects fall into three programmes of work: evaluating complex interventions, enhancing evidence transferability, and increasing evidence use.

**CEDIL Methods Working Papers**
The CEDIL Methods Working Paper Series offers innovative research methods to develop impact evaluation and evidence synthesis work in low- and middle-income countries.

**About this working paper**
This paper, 'Making predictions of programme success more reliable', provides an account of a type of detailed theory of change called a 'causal–process–tracing theory of change' that can be very helpful for programme prediction, planning and evaluation.
The content of this paper is the sole responsibility of the authors and does not represent the opinions of CEDIL or the Department for International Development (DFID). Any errors and omissions are also the sole responsibility of the authors. Please direct any comments or queries to the corresponding author, Nancy Cartwright, at nancy.cartwright@durham.ac.uk

## Acronyms and abbreviations

| | |
|---|---|
| AEM | action effect method |
| APPP | African Power and Politics Programme |
| BPSR | Bureau of Public Service Reforms |
| CEDIL | Centre of Excellence for Development Impact and Learning |
| CCT | conditional cash transfers |
| DFID | Department for International Development |
| EFCC | Economic and Financial Crimes Commission |
| NAFDAC | National Agency for Food and Drug Administration and Control |
| NAPTIP | National Agency for the Prohibition of Traffic in Persons |
| PERL | Partnership to Engage, Reform and Learn |
| SAcc | social accountability strategies |
| ToC | theory of change |
| WDR | *World Development Report* |

## Contents

## 1 Aims

This paper aims to fill a gap we see in policy and programme planning, not just in development but across policy domains. It involves the question of effectiveness, for both *ex ante* prediction and *ex post* evaluation: is your proposed programme likely to produce the kind of results you are hoping for if implemented in the way you intend in this location at this time? Did the programme that was implemented do what it was supposed to?

There is not much systematic, well-grounded advice available about how to answer effectiveness questions, and especially how to answer them in a way that is recognisably well justified. There are a great many useful specific suggestions about things that should be considered, such as ensuring that local stakeholders are on board. What is missing is a systematic account of the information you should have to back up claims, whether *ex ante* or *ex post*, and how that information fits together. This paper offers an account of a type of detailed theory of change (ToC) called 'causal–process–tracing theory of change' – process ToC (pToC) for short. It can provide significant help for programme prediction, planning and evaluation. A good pToC should make both the prediction and the evaluation of success more reliable and pertinent:

- It helps predict if a programme can be expected to work in a specific setting.

- It offers insights into what design features are needed for success.

- It provides invaluable information for monitoring whether the programme is on track and for fixing problems that arise.

- It reveals the causal processes and related assumptions to be tested in an evaluation.

- It helps in identifying evaluation questions.

- It helps in interpreting evaluation findings, assessing their relevance, and locating a description of them that is useful for programme design and evaluation in other settings.

Some of the points we make may seem obvious. But we note that in many agencies, including the Department for International Development (DFID), the programmes' ToCs, are not generally – if ever – constructed in a form that takes note of it all. Programmes fail over and over again due to neglecting to account for the kind of information we describe, information that often could have been assembled. Some of these oversights are so big as to be startling, for instance: conditional cash transfer programmes set up to pay parents to take children to clinics that didn't exist. We aim to provide the conceptual tools necessary to understand what should go into a good ToC.

We will discuss how to produce pToCs for a programme at two different levels. First there is the pToC that represents how the programme should work across the range of settings for which it is recommended. This wide-ranging theory can then be thickened to produce a

more detailed and concrete pToC for each specific setting in which it is to be implemented. We call this setting-specific theory the 'local-level pToC', and the more wide-ranging theory a 'middle-level theory' since no programme design will hold universally.[1]

**Who is the audience for this paper?**

First, we start with those in many international development agencies, including DFID, who design or endorse development programmes that are expected to work across a range of settings. It is widely recognised that no programme will work everywhere. How are rational decisions to be made about whether the programme you endorse will succeed in a particular local setting? To make these predictions as relevant as possible, decision makers need to know what must be in place and what guarded against in the local setting, step by step, if the programme is to carry through successfully. They need a good local-level ToC.

This is where we think there is often a significant gap in DFID planning. Those making decisions about what is best to do locally are generally not in a good position to figure out just what is supposed to happen, one step after another, and what must be in place or guarded against if the programme is to deliver the desired results. This is your programme that you are designing or endorsing and that you know most about. So you should be trying to provide local users with the maximal information possible to enable them to answer these questions in a reliable and pertinent way. That is what the middle-level pToC for the programme is supposed to do. However, a good middle-level pToC is not usually easy to construct. It can take hard thought as well as more research and theorising. Our account of what makes for a good middle-level pToC is intended to provide a framework that lays out what kind of information you need to supply to make decision-making about particular local settings more reliable and more solidly justified.

Some of the information that pToCs bring together is generally known but only implicitly noted, exists only in people's heads, or is documented in ways not linked to the ToC, such as risk mitigation plans or evaluation checklists. The problem is that the information is either not made explicit or not articulated and documented early enough, or not sufficiently incorporated into the thought process for building a meaningful ToC. So long as this information remains scattered or in the heads of designers, it cannot be of help.

For other information, developers and researchers will need to do the hard work and the additional research required for figuring it out. We urge DFID programme designers to undertake the research necessary to produce good middle-level pToCs and those commissioning programmes from elsewhere to demand them. As Nobel-prize-winning economist Angus Deaton remarked: 'There is a great attraction of being able to make policy recommendations without having to construct models. I understand the appeal..., but I believe that attempts to do so are bound to fail'.[2]

---

[1] Beyond the lack of a clear set of tools for conceptualising what good ToCs for prediction and evaluation consist of, there are surely other reasons that these are so seldom on offer. Just what these are, and how to manage them, is beyond our scope.

[2] Deaton 2019

Second, we address those who are making, or participating in, decisions about whether or not – and how – to implement a recommended programme in a specific local setting. You will need to make a judgement – an informed judgement – about whether the programme is likely to deliver the desired results in that setting. For this you should have a local pToC detailing just what is needed, step by step, for programme success in your local environment. An invaluable starting point for this will be the middle-level pToC for the programme. But because this ToC describes a process that is meant to work across a range of settings, it will inevitably use language that is general and abstract. You need to know what these general terms amount to in the local setting. For instance, from our case study 1, the general programme pToC for the use of innovative mobile phone applications to support nutrition outcomes talks about 'health workers'. In the Indonesian setting we discuss, it is important to realise that in Indonesia, the health workers involved are, more concretely, specific volunteer community health staff. In case study 3 on conditional cash transfers (CCTs), the middle-level pToC calls for the transfers to be low cost to recipients. In Northern Macedonia that was achievable by using cheques cashable at local banks and post offices. But that might not be the way to make transfers low cost to recipients elsewhere. There can also be specific factors that will be relevant in a particular setting that do not generally matter elsewhere. So you, or those who advise you, will have to 'thicken' the middle-level programme pToC to make it fit your local setting. We describe in Section 3 what such a thickened local pToC should contain.[3]

Third are those who evaluate programmes that have been implemented to judge whether the programme did achieve the intended results. Most evaluators already recognise the importance of a good ToC as a source of indicators of whether the programme genuinely contributed to improved results or not. The kinds of ToCs we urge – pToCs – provide a very strong source of indicators to look out for.[4]

Finally, we address anyone interested in designing, researching or evaluating programmes. Besides the accounts we offer of what middle-level and local-level programme pToCs should look like, the additional take-home message for this audience is: middle-level theory matters. This paper will explain where and why.

We shall use two case studies as running examples, one in each of the two areas DFID has asked us to consider: health and democratisation. The first, labelled 'mHealth', involves the use of mobile phones (specifically smartphones) and associated technology to improve children's growth status in settings where children are measured and treated in community health clinics but data is reviewed and aggregated, and resources are allocated, in locations distant from the clinics. The other is a general anti-corruption programme that harnesses dedicated and visible, powerful leadership to effect change. We also discuss a third case

---

[3] There are other ways to justify a prediction about programme effectiveness *in situ* than via the use of a good local pToC. For instance, sometimes there may be strong reasons to support the claim that *this* setting is suffi-ciently like settings in which this kind of programme has worked well in all significant respects. But beware. To be justified, this kind of claim needs to be supported by good reasons – a kind of loose 'induction by simple enumeration' will not do. These kinds of reasons are hard to come by, especially in the absence of a good pToC that shows what the significant factors are that affect success.

[4] There is a call for constructing specific evaluations of locally implemented programmes so that they can help teach general lessons about the programme, which would allow the generation of a better general pToC for the programme.

study, an example of a CCT to illustrate thickening a middle-level theory to adapt it to a specific setting. You can read about these cases in Section 5, where we have constructed exemplars of a pToC for each. But the use we make of them in the main text should not require any detailed understanding of them.

Most readers will be familiar with some of what we say, and different readers with different backgrounds will probably be familiar with different things. A few readers will recognise much. That is all to the good. We do not cut our proposals from whole cloth but build them on different bodies of work from different domains that have already been well researched.[5] What we stress is the need for programme designers to articulate the long sequences of intermediate cause–effect relationships necessary to achieve the targeted goal of the programme and the importance of the middle-level causal principles, which underwrite each cause–effect pair in the sequence. These principles are important for identifying what factors are necessary for the production of the effect at each stage and what factors will dilute or prevent it.

We should underline that we do not try to reiterate general lessons that hold for almost all programmes and that are widely and well discussed elsewhere.[6] Rather we focus on **how to evaluate the effectiveness (or not) of a proposed programme for a set of specified results in a specific local setting in a systematic and informed way** – information that is essential for a reasoned decision about whether to try to implement that programme in that setting to achieve those results.


*2 Theories of change*


**2.1 Introduction to ToCs and middle-level principles**

To get started, this section will explain what we mean here by 'theory of change' and 'middle-level theory'. We begin with middle-level programme ToCs: pToCs for programmes that are expected to be applicable across different settings (though seldom universally). Our proposals are intended to improve their design, since they generally do not have all the ingredients they need to make it easier to thicken them with information about the local setting to build more-reliable local pToCs.

---

[5] Our proposals draw heavily on Wesley Salmon's process theory of causation (Salmon 1998) and J. L. Mackie's INUS ('Insufficient but Necessary parts of an Unnecessary but Sufficient condition) account (Mackie 1980); on Jon Elster's work on psychological and social middle-level tendency principles ('mechanisms' in his vocabulary) (Elster 2007); on realist evaluation's stress on activities and on its work on processes ('mechanisms' in their vocabulary) (Pawson et al. 2004) and of EBM+ (http://ebmplus.org/). We also draw on the more concrete proposals of Cartwright and Hardie 2012, Munro et al. 2016, Layne et al. 2014, Howard White (eg White 2009; 2018), on similar ideas from the International Rescue Committee's outcome and evidence framework and many others.

[6] Such as the need for stakeholder involvement, widespread barriers to organisational change, the general importance of the larger environment in which programmes are set, the need to address differences in understandings of what the problems are and what intended results should be.

As Rick Davies reports in his CEDIL inception paper 'Representing Theories of Change: Technical Challenges with Evaluation Consequences',[7] 'Stein and Valters (2012) have explored various interpretations and concluded that despite the variety of views "Theory of Change is most often defined in terms of the connection between activities and outcomes, with the **articulation of this connection the key component** of the Theory of Change process" (emphasis added)'. Our pToCs build from this idea. The pToC for a programme is a model that presents the step-by-step process by which the programme is to produce its intended results, including specific information about the principles of change for each step and their implications.[8]

Recall that we shall discuss theories of change at two levels. First at the middle level – the pToC design for programmes that are expected to be of use not just in one place but more generally across some specified kinds of settings. The second is at the local level – a pToC specific to a particular setting. These are a key source for predicting if the programme can work in the local setting, and they are also a valuable source of indicators for *post hoc* evaluation of whether a programme did what it was supposed to in the local setting.[9]

ToCs are usually presented using arrows and variables. Our pToCs are meant to trace the causal process by which the results are to be achieved. The arrows represent causal relations: each step is meant to cause the next. A central feature of the pToCs we recommend is the inclusion of the causal principles by which each step is to produce the next. These will usually be middle-level tendency principles. Here we explain briefly in what sense they are middle level. Later in Section 2.3, when discussing assumption type 2, we will explain what we mean by calling them 'tendency principles'. You can find a more thorough account in the Appendix.

'Middle-level theory'[10] is a loose umbrella term. In social science, middle can be with respect to the level of abstraction or to the level of generality and breadth or both. On one side, it covers anything between down-to-earth local predictions in language that is readily applied and operationalised. On the other, it covers high social theories, including those of Karl Marx, John Maynard Keynes, Pierre Bourdieu, Anthony Giddens and Judith Butler, with their abstract, general concepts such as materialist theory of history, monetary versus fiscal policy, habitus, structuration and gender performativity respectively. We shall discuss two

---

[7] Davies 2018

[8] Theories of change are richer than logic models since they go beyond the activities to 'define the assumptions and necessary and sufficient preconditions for the sequence of outcomes needed to reach a goal' (Dhillon and Vaca 2018, 65). Generic causal links refer to 'the simplistic relationships among elements using arrows. They provide the reader an indicator of the causal order of and relationship between different outcomes that are caused by the strategies and activities' (Dhillon and Vaca 2018, 68). In addition, specific causal links offer a 'nuanced depiction of relationships among elements using arrows and lines' (Dhillon and Vaca 2018, 69). These relationships are not necessarily linear, even though the arrow is clearly meant to influence a particular output.

[9] Note that these are not meant to be useful for generalising. They are to help us get right the prediction about whether this programme will work here. Whether – and where – it will work elsewhere requires a great deal of theorising and research. How to go about the construction of a well-constructed theory of where a general programme will work is not our topic here, beyond the suggestions that follow from identifying a good middle-level pToC for the programme.

[10] The term originates in the work of Robert Merton (Merton 1968). For more about it see Boudon 1991. For application specifically in the domain of programme evaluation see Pawson 2009.

interrelated types of middle-level theory that can help in predicting a programme's effectiveness in target settings:

- the p*ToC* for the programme.

- middle-level principles or tendency principles – psychological, social, political and economic.

We have already explained why we call the first 'middle' level – they are not expected to apply everywhere but they are expected to apply across a range of settings. They also use language that is more abstract than in local-level theories but not so abstract as to be untestable. The second are between familiar high-level principles such as 'agents act to maximise their expected utility in the face of expectations about others doing the same' and principles very specific to particular groups or settings, such as 'Arsenal supporters tend to wear red and white and, before the match, meet at the Tollington or the Twelve Pins'. We discuss these and their importance in effectiveness prediction more in Section 2.3 when we turn to assumption type 2.

**2.2 Building up a middle-level programme pToC**

This section will show the steps involved in producing a middle-level programme pToC. Our aim with these pToCs is just like that of the action effect method (AEM) developed for medical care interventions. As its developers report, 'the need for programme theories and logic models is well articulated. However, there is little practical guidance available on how to construct good quality diagrams. The AEM adds to this through explicit articulation of the components of programme theory and their relationship to one another in diagrammatic form, something that other models often lack. Articulating complex concepts in a single diagram plays an important cognitive role in supporting readers'.[11] The AEM offers a basic arrows-and-variables model depicting the causal process by which a programme is expected to produce the targeted outcome. We shall enrich this to include further kinds of information needed for predicting whether the programme will work and showing the relations among these various types of information.

Our project is thus wider than others we have seen addressing what should be in a ToC. But it is also narrower in some respects. For instance, we suppose that there are already a clearly articulated aim and an articulated programme. When designing or adjusting a programme to suit a local context, it is (as many reports urge) useful, and often necessary from a practical, ethical and political point of view, to agree on the programme design and intended results locally with all stakeholders involved, including those who must implement the programme and those who may benefit or lose from it.[12] How to negotiate and settle on

---

[11] Reed et al. 2014, 7.

[12] This is seen, for example, with child protection in the UK, where there is often a disconnect between what the government promotes (research-based policy and the 'what works' centres) and the issue discussed by senior managers in child protection services (see Munro 2011). The disconnect is partly in the different views of outcomes for service workers and government and what are considered as factors of success; for example, the language of better outcome for children and uncertainty about what good might look like. The same kind of disconnect can readily occur in development programmes as well.

aims is beyond our scope. We focus instead on predicting *ex ante* and evaluating *ex post* if programmes achieve those aims. We aim to identify information that a pToC should include for programmes designed to produce already articulated outcomes. These middle-level programme pToCs are intended to maximise the help they provide to those making decisions about local programme adoption and implementation for building local models that produce more reliable predictions about the chance of achieving the results selected.

We begin with the simplest version of this kind of pToC: a simple input–output arrows-and-variables model. Many programme ToCs are presented in this way, with the programme input at the head of the arrow and the expected output at the tail, as in Figure 1. We start there and consider a series of steps to follow next.[13]

**Figure 1: input–output arrows-and-variables model**



A programme seldom has the desired outcome as a first immediate effect. Rather, the effect is brought about through a series of intervening steps, each causing the next, eventually producing the outcome. Simple arrow-and-variables ToCs (sometimes called 'impact pathway diagrams') are meant to represent the sequence of significant causal steps in the process that is supposed to produce the final outcome from the programme intervention. They look like Figure 2.

**Figure 2: impact pathway diagram**



The kind of pToC we prescribe is meant to help both with predicting whether and to what extent a programme will produce its intended results in a targeted local setting. It is also meant to help with *post hoc* evaluation of whether it has done so if it has been implemented. That affects what the arrows and the variables represent and how many boxes there should be.

---

[13] There are a variety of ways of trying to represent visually many of the different kinds of information that we urge should be in a good pToC. We use one that we have been developing for a while (see for instance Munro et al. 2016) and that we find especially helpful. There is also a lot of software being developed, some of which can be of aid here too (see, for example, *Monitoring and Evaluation NEWS*).

Arrows: as noted, these are meant to show that the variables represented in the box at the head of the arrow cause those in the box at the tail, in contrast to mere time sequencing. For example, it is not just that soap was delivered and hands were clean at lunch but that the soap helped to cause the hands to become clean. Items appearing later in the diagrams are meant to occur later; those connected by arrows are meant to be seen as cause–effect pairs.

Variables: the variables in the box at the head of the arrow are to be a cause of the effect in the box at the tail. This is not always what we find in an arrows-and-variables chart. Sometimes the boxes in the middle mark high points or indicators of success along the way. This can be useful for some purposes, but it is not what one needs to know to predict whether a programme will work in a target setting. As realist evaluation and causal process theories stress, because the variables in the ToC are to pick out causes, they will generally represent activities of either individuals or of institutions and not just states of affairs.

Sometimes in a ToC only one or two intermediates appear in the middle. This can be all right if the process, from start to finish, does not cover much time and space. But generally that is not so. A good pToC for helping with design, prediction and evaluation should include every significant intermediate step that must be produced along the way if the initial programme intervention is to generate the final result. Otherwise, we risk overlooking significant information that could help to predict success or failure.

When, as is always the case in development contexts, the inputs and outcomes of interest are separated in time and space, the two must be connected by a continuous causal process. If any stages in this process are missing, the desired results cannot be expected. If they are weak, the outcome is diluted or threatened. So it is important to map out all the significant causal steps in between.[14] Figure 3 provides an example of a ToC from a DFID document with a short number of intermediate steps, though it does underline the important point that many of these causal sequences will have feedback loops. Feedback loops are causal circles in which an effect of a process produces effects that can affect the future running of the process. They can be either positive, enhancing the process, or negative, making it weaker. They often occur over longer time frames than the programme itself.

---

[14] See White and Booth 2003

**Figure 3: example of ToC with short set of intermediary steps**
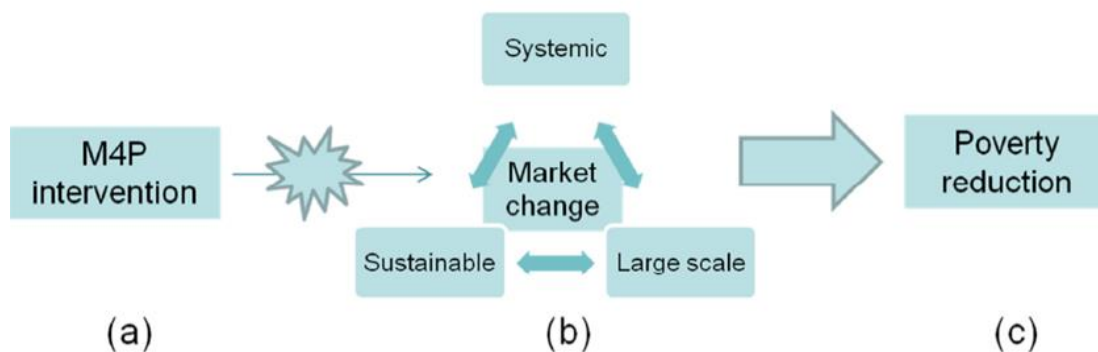


Figure 4, cited in Davies (2018), is a good example of a more detailed arrows-and-variables model with feedback loops.[15]

**Figure 4:[16] example of ToC with feedback loops**



---
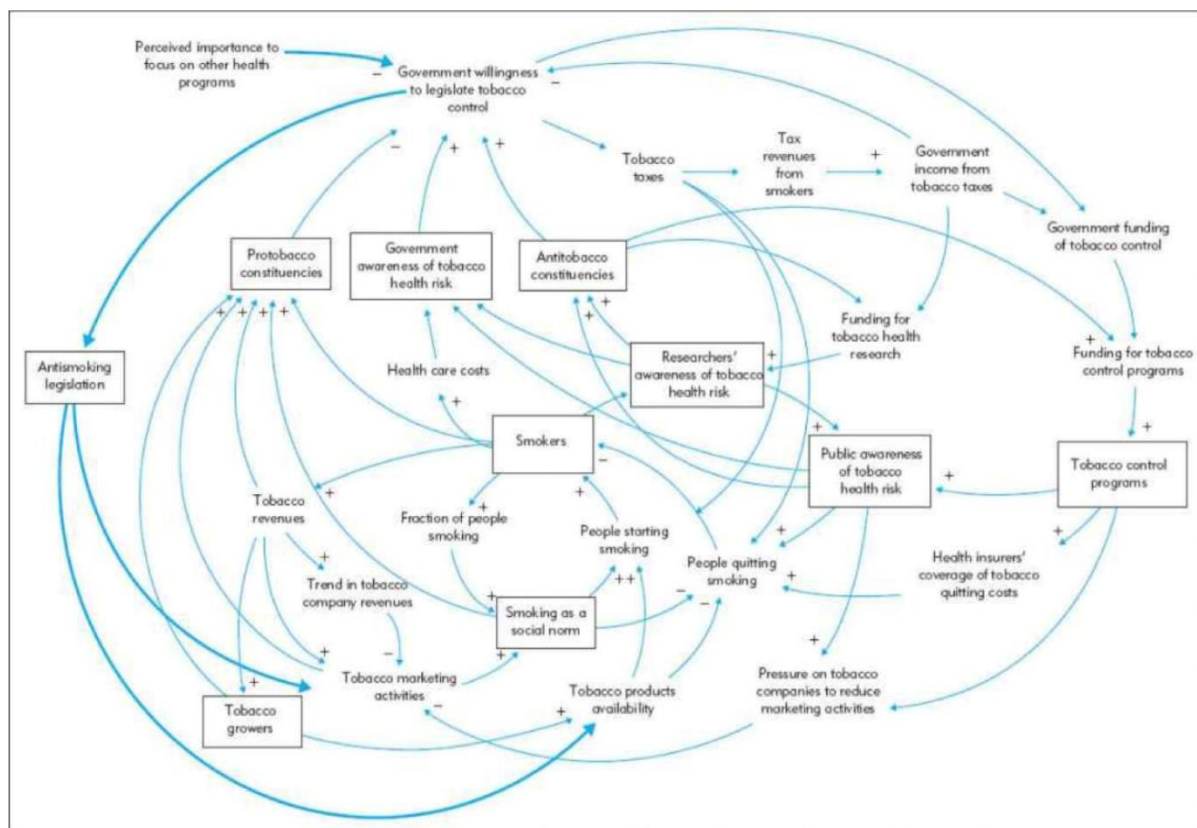
[15] The further claims for mHealth also involve loops. The availability of data taken to be accurate and reasonably comprehensive plus the improved nutrition of children where mHealth is used are expected to promote more government attention and resources to problems of childhood nutrition. We are not graphing these further stages to avoid the diagrams becoming so complicated as to be unreadable. Although loops are not uncommon, our three cases contain only one.

[16] Davies 2018 (originally in Haiku Analytics 2018)

DFID commonly expects that data and evidence for impact, outcomes and outputs be tracked as part of a programme's regular monitoring and quarterly reporting procedures. This leads to the production of 'results chains' in output–outcome–impact charts for programmes. See Figures 5 and 6 for examples.

**Figure 5:[17] output–outcome–impact chart example**

All pillars are concerned with bringing about improvements in changes in social norms leading to behavioural change in government and greater expectations amongst Afghan citizens of improved government performance, and reduced corruption.

The SNG task Force retains the political will to incorporate learning into policy development.

The three pillars of the programme work effectively together.

**Outcomes**

| Pillar 1 –Accountable, Capable and Responsive Government | Pillar 2 – Engaged citizens | Pillar 3 - Learning, Evidence and Influencing |
|---|---|---|
| Improved capacity for delivery of public services through institutionalisation of good governance best practices by SNG institutions in six provinces | Increased community engagement in SNG in rural communities in 6 provinces | Evidence from innovations in citizen-centred public service delivery informs revision of national policy systems |

**Assumptions (output to outcome)**

A worsening security situation or macroeconomic context does not distort government's use of resources
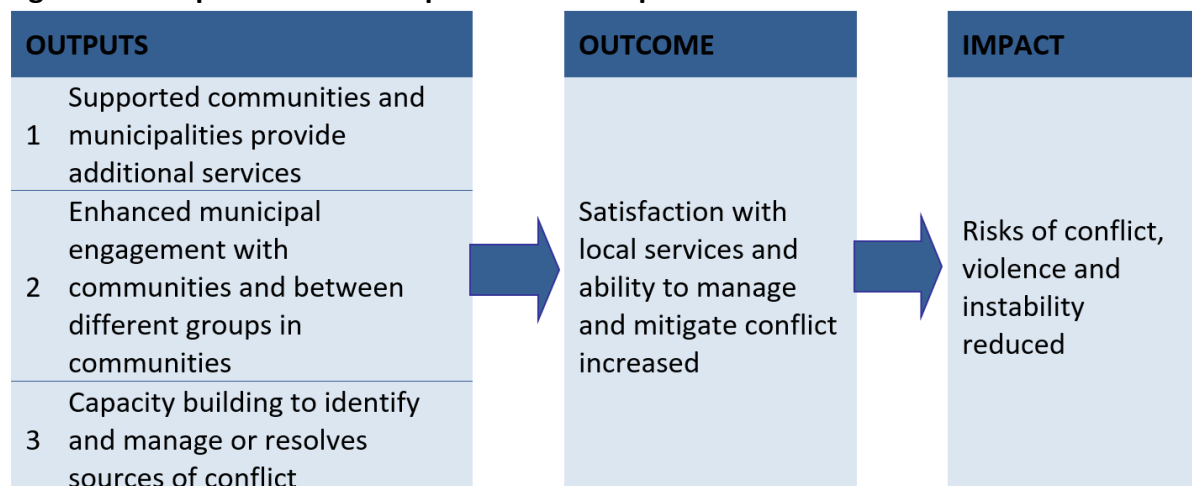
Elites and key constituencies retain an interest in the issues covered by the programme

Successes from learning laboratory can be can be replicated more broadly by the SNG Task Force.

The programme will work in provinces and districts where there is a level of pre-existing reform commitment

The three pillars of the programme are able to form effective partnerships with local constituencies.

*Guiding principles to ensure that outputs deliver outcomes:*

*All pillars focus on issues with political traction where there is local demand and interest in bringing about change.*

*There is a clear understanding of local political economy*

*There is a willingness to experiment with a variety of intervention approaches.*

*Well-developed monitoring and evaluation systems are in place to assess the results and strong management systems adapt the programme design according to lessons learned.*

**Outputs**

| Sub-national government institutions equipped with the skills and tools required to institutionalise good governance best practices in delivery of public services | Strengthened community engagement in citizen-centred good governance practices | Evidence from innovations in citizen-centred public service delivery generated to guide national policy through research and field testing |
|---|---|---|

---

[17] DFID 2017

**Figure 6:**[18] **output–outcome–impact chart example**

| OUTPUTS | | OUTCOME | IMPACT |
|---|---|---|---|
| 1 | Supported communities and municipalities provide additional services | Satisfaction with local services and ability to manage and mitigate conflict increased | Risks of conflict, violence and instability reduced |
| 2 | Enhanced municipal engagement with communities and between different groups in communities | | |
| 3 | Capacity building to identify and manage or resolves sources of conflict | | |

These provide a kind of wish list. They can be helpful in detailing and sorting different desirables that are meant to be achieved and for monitoring that these are or are not being accomplished. There is first what the programme is narrowly designed to produce (the output or product), then the short-term effects on people or social institutions those outputs are supposed to be in aid of (outcomes) and finally the longer term, often indirect effects we hope will be achieved (impact). and then But these diagrams do not give any indication of what the sequence of steps are. By what process is the input supposed to lead to the output? By what process is the outcome to lead to the impact? With respect to effectiveness, we may be interested in whether the input will lead to one or another or all of these in the target setting.

However, mapping out the sequence of significant steps along the way is not going to be enough to help with reliable prediction, as is widely recognised in discussions of ToCs. As Davies (2018) quotes from a recent review of ToCs, 'theories of change are typically made up of boxes and arrows. Boxes are filled with text descriptions of events, and arrows connect them, representing expected causal connections between these events'.[19] But, Davies argues, it is the connections between the events in a ToC that are most problematic, since the arrows connecting the events in a simple arrows-and-variables ToC are without annotation and there is therefore a lack information on the nature of those connections.[20] There are a variety of different suggestions about what more is needed. The ones we urge are chosen specifically for their importance in helping to build local models that will make more-reliable predictions.

## 2.3 Six types of assumption

Here we describe the kind of assumptions that are part of a good theory for *ex ante* prediction and *post hoc* evaluation.

---

[18] DFID 2016
[19] Davies 2018, 4
[20] Davies 2018, 5–6

The first step in constructing a pToC is the one noted already: the sequencing of important causal interactions that must each occur one after the other if the outcomes are to obtain as hoped.[21] Next is to add what Howard White calls the underlying assumptions that the ToC makes, which include both the overall programme theory and many other middle-level principles, as well as assumptions about what specific kinds of fact must obtain if the programme is to deliver the expected results.[22] Calling these 'assumptions' does not imply that we take it for granted that these factors or conditions are present. On the contrary, they are factors or conditions that need to be present (or absent) to enable the causal process to carry through step by step. We organise these assumptions about why and how the programme is supposed to produce the final results according to the role we see them playing in ensuring that the process goes through. We sort these assumptions into six categories:

1. The overall programme theory, which usually includes a number of middle-level principles that together explain why the programme inputs should be able to produce the desired outputs and outcomes.

2. The middle-level principles (generally these will be tendency principles, see below in our discussion of this assumption plus the Appendix to learn what these are) that underwrite each step in the causal sequence.

3. For each step, the support factors that must be in place for that step to help to produce the next (sometimes called 'moderators' or 'interactive variables').

4. For each step, the derailers that may prevent the next step or diminish the chances of it being successfully produced.

5. For each step, what safeguards might protect the causal process from derailers.

6. A general account of the range of application, i.e. the kinds of settings where the programme can be expected to work.

Figure 7 presents a simple example of these assumption types for the mHealth and Nutrition example from case study 1. Assumption type 1, the overall programme theory, explains why the programme should be expected to be able to produce the final targeted results. Assumption type 6, range of application, is again about the overall programme and its results (though the restriction in the chart is only one among all those that will be suggested by looking at every step). Assumption types 2–5 are about each step in the pToC, and here we look at steps 2 and 3 for illustration.

---

[21] This step-by-step sequence is what in Evidence Based Medicine Plus (EBM+) literature is called the 'mechanism' connecting input and outcome. EBM+ is a voluntary network stressing the importance of mechanistic evidence in judging efficacy and effectiveness in medicine.
[22] White 2009

**Figure 7: the six types of assumption: the case of an mHealth intervention**

| Type | Example |
|---|---|
| Overall middle-level theory | Accurate identification of underweight infants in the community health clinics is to lead to more timely submission of data from local to district health centres, especially in rural areas. |
| Mid-level causal principles guiding each step<br><br>Step 2 → step 3 | Automated methods for doing so promote data submission. |
| Support factors<br><br>Step 2 → step 3 | Community health workers are able to submit the data. |
| Derailers<br><br>Step 2 → step 3 | Community health workers could be prevented by external pressures or other priorities from doing so; they could see other tools as dominating the automated methods; the technology could fail. |
| Safeguards<br><br>Step 2 → step 3 | Procedures added to ensure that community health workers believe in the tools and do not have too many other pressures; procedures also added to ensure that the technology works properly. |
| Range of application | mHealth will not contribute to the expected degree of improvement where community health workers are already good at following the formula for calculating nutrition status. Otherwise it should work well where all requisite support factors can be put in place along with safeguards against all derailers. |

*Assumption type 1: overall programme theory*

Overall programme theory explains why a programme could initiate a process that can result in the desired outcome. This is what realist evaluation usually means when they call for the mechanism of the programme to be described.

These theories may be simple in structure or they may be fairly elaborate. For an example of a relatively simple theory, consider mHealth programmes, where part of the theory refers to data accuracy. The theory is that technology will improve the accuracy of data on children's

nutrition status and this in turn will promote a better response by key nutrition stakeholders, e.g. faster referrals to a midwife or other health services. This assumes that one of the reasons response rates have been historically low is the perceived inaccuracy of data on children's weights.[23] We could call this the 'boy who called wolf' theory: if a signal has proven inaccurate in the past, we will stop using it.

By contrast, the Nigerian Partnership to Engage, Reform and Learn (PERL) programme, which we use as the basis for our anti-corruption pToC, has evolved an overall theory with a large number of related assumptions. In the 2019 programme level report, ten assumptions are presented, in two groups.[24] The first four are about how governance reform and service delivery improvements happen in Nigeria. (They label this 'Theory of Change' but we shall not adopt their terminology since they use the phrase in a different way from us.) Here is one example:

> Reform occurs through a combination of supply-side, demand-side and evidence changes.

The other six constitute a theory of how PERL can most effectively support change, including, for instance:

> PERL can work most effectively by incorporating locally led, problem-driven, context appropriate, politically smart adaptive and partnership based approaches.

These are the basic assumptions PERL uses to identify and underwrite potential change pathways and to show how these lead to improved delivery of public goods and services. The 14 change pathways they describe are each intended to eliminate a bottleneck or obstacles they have identified to improved service delivery in Nigeria. They comprise 14 separate input–outcome pairs, such as greater budget realism, which is supposed to enable timely release of funds for predetermined priorities, as in Figure 8. These are thus the basis for a simple two-step input–outcome ToC, ripe for developing into more detailed multi-step pToCs that can be used for prediction and planning PERL programmes and for their subsequent evaluation.

**Figure 8: example of change pathway to eliminate an obstacle**



*Assumption type 2: middle-level (primarily tendency) principles*

---

[23] Barnett et al. 2016
[24] PERL 2019, 3

Then we will need a myriad of smaller theories. Each step is supposed to play a role in producing the next, and that it can do so should not be an accident. There must be a reason why that particular factor can cause that particular effect, some well-justified principle it is in accord with. These principles may be ones that hold fairly generally (though very seldom universally), or across a specific range, or they may hold over a highly restricted range. But they must be principles that there is good reason to believe can hold across the range of settings in which the programme is expected to work. We clump principles of all these different degrees of generality together under the heading 'middle-level theory'. Most will be tendency principles.

What is a tendency principle? We discuss them in more detail in the Appendix. Here we give a brief account to provide a sense of them. The term comes from John Stuart Mill,[25] who argued that most of the principles to be uncovered in political economy will tell what effect a cause is tending to, not what effect actually occurs. That is because what happens is usually the result of a number of different causes pushing and pulling in different directions at once. These tendency principles often describe familiar psychological or sociological dispositions widespread across individuals or institutions, or widespread in individuals and institutions in specific settings. In this sense, tendency principles are not deterministic.[26]

Tendency principles are often familiar behavioural principles, such as 'parents tend to act in the interests of their children'. Equally, they may be the result of social science research, such as 'people tend harbour implicit biases'. Either way, they generally cannot be relied on to tell what must happen. Instead they describe what the cause tends to. This is means that:

- You may have to do something to call the cause into action.

- Even when the cause acts, the indicated effect may not be the observed outcome because other causes influence that outcome as well.

- Nevertheless, the cause may push the outcome in the direction indicated.

Many middle-level tendency principles can seem too commonplace to warrant the label 'theory', such as the principle we mentioned that parents tend to look after the interests of their children. Others can be so specific as to hardly seem middle level, such as 'the mHealth technology tends to accurately calculate growth status' or 'mHealth technology tends to be well designed'. We shall give more examples and discuss these in some detail later. Nevertheless, it is probably easier first to come to grips with the kind of additional information that these principles can aid in discovering – facts about support factors, derailers, safeguards and range of application.

We represent the principles that cover the causal relation at any step by a curved arrow above, from one box to the next, as in Figure 9.

---

[25] Mill 1848

[26] Even when no other causes interfere, the canonical effect may not occur even when the tendency is properly triggered if the disposition is fundamentally probabilistic.

**Figure 9: illustration of the principles that cover the causal relation at each step**
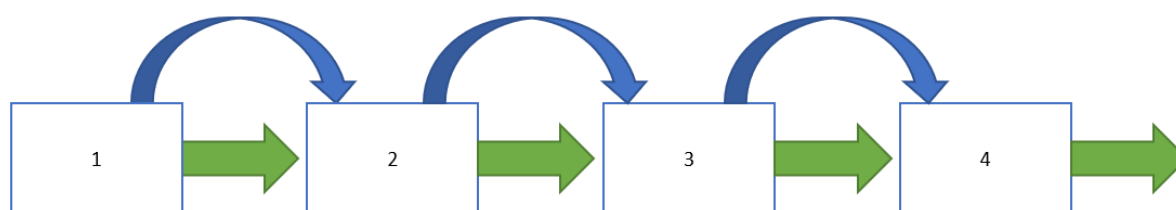


Sometimes the programme design may envisage that the causal relation at one stage or another can occur under more than one principle. In the case of anti-corruption, there are many principles that get bureaucrats with different dispositions to adopt a pro-reform mindset. One tendency, which we might expect among the more active bureaucrats, is to positively adopt a pro-reform mindset once they see that reform is a live option. Another tendency, which may be likely to operate among the more passive bureaucrats, is to comply with whichever mindset is popular. Once a pro-reform mindset becomes popular, the more passive bureaucrats adopt it. Yet another tendency, likely to operate among the more sceptical bureaucrats, is to adopt a pro-reform mindset once they see tangible results. And so on. In this case, more than one curved arrow can be inserted. Registering all these different principles matters since they require different support factors if they are to be triggered, and they can have different derailers.

*Assumption type 3: support factors*

Few of the things we call 'causes' can produce their effects by themselves. They almost all need other factors to cooperate with them to make it likely that the cause will have the kind and/or size of result expected. We call these necessary factors 'support factors'. (Others call them 'helping factors', 'moderators' or 'interactive variables'). Each causal link in a pToC will almost certainly require support factors to enable the feature at the left of the arrow to result in the effect at the end of the arrow. For instance, consider this two-step input–outcome ToC from PERL's document *Public Sector Reform in Nigeria – What Works?*[27]

> 2001 addition of Nigeria to the list of Non-Cooperative Countries or Territories (NCCTs) for non-cooperation in combating money laundering → effective repression of money laundering in Nigeria and 2013 removal of Nigeria from the list.

It is clear that blocklisting will never produce reform on its own. It takes a lot of additional factors to be in place with it before any such results can be expected. PERL lists four support factors that worked with the blocklisting to produce reform:

- Government concern that international blocklisting would harm foreign investment and Nigerians abroad.

- President Obasanjo's background in anti-corruption.

- High-profile arrests that built credibility.

---

[27] Hussein and Sharp 2018

- Effective use of technology that supported public accountability.

For a second simple illustration, consider our example of mHealth, where the first step in the ToC is that mHealth is administered and its use is mandated at the local level; the second step is that infants' weight data is recorded in smartphones at community health clinics. Immediately we see that a number of support factors are required if step 1 is to lead to step 2, including:

- Community health workers have the capacity to use mHealth.

- Community health workers accept that using mHealth is good for their clients.

- Mothers and infants attend the clinics on a regular basis.

Epidemiologists graph the collection of all the factors that must be in place together in what are called 'causal pies'. A causal pie for the PERL blocklisting process might look like Figure 10.

**Figure 10: causal pie for the PERL blocklisting process**



We should also note that there may be more than one set of factors that can cooperate with a given cause to produce the desired effect. For instance, in the blocklisting example, a strong and committed judiciary and police force may have substituted had the president lacked experience with anti-corruption.

The need for support factors is not peculiar to these sweeping two-step input–outcome chains. Rather, everywhere a causal arrow appears between two factors, support factors will be called for. Thus, this is our lesson here:

> A middle-level programme pToC should work to include all the support factors that can be figured out that are essential to the overall theory and for each step. This matters because a causal process is only as strong as its weakest link. Any real implementation of the programme will fail anywhere that a necessary support factor is missing at any stage of the process, from programme inputs to hoped-for outcomes.

It is important to note that support factors may cross levels. Organisations often need and benefit from support factors supplied by the systems – larger units of government or society – of which they are subsystems. Obvious examples that apply widely are funding, realistic (vs unrealistic) expectations, prioritising appropriately, setting the right incentives or instancing an appropriate commitment and culture.

In constructing the diagram for a general programme pToC, to represent a set of causal factors that must all be in place together at a given step if they are to make the contribution required towards the next step, we enclose all the factors necessary to operate together in a circle at the head of the arrow. One might mark out the factors in this circle that programme protocol proposes in some special way, perhaps in a different colour or in a box, as in Figure 11, which is taken from our anti-corruption example (see Figure 10).

**Figure 11: alternative diagram for a middle-level programme pToC**
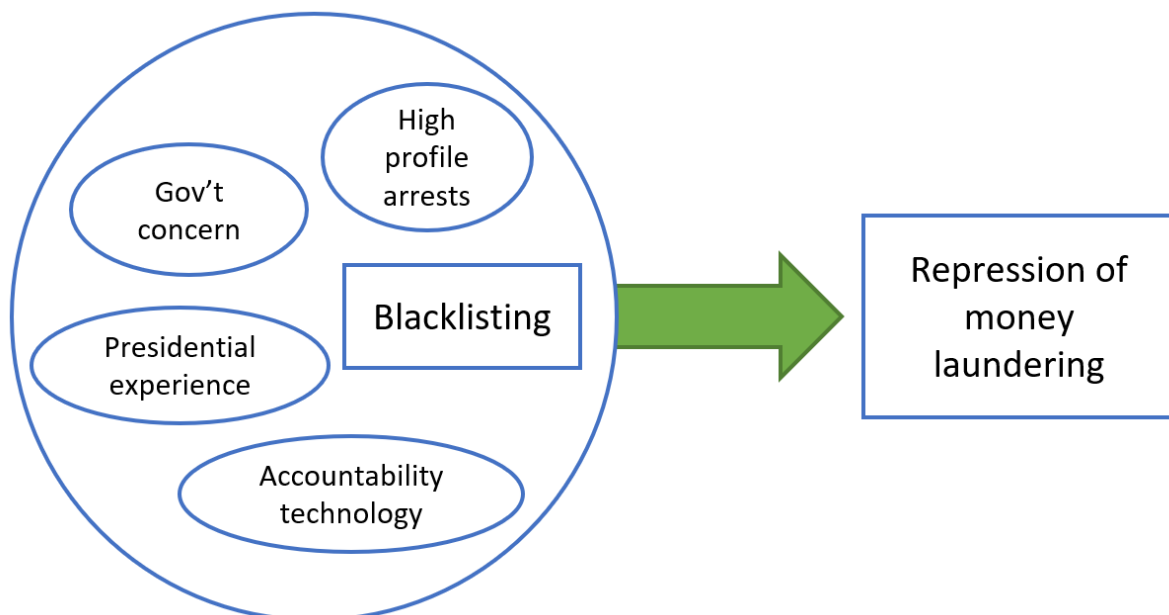


Note that the causal arrow does not originate at the box. The box represents the factor that, through a sequence of previous steps, is meant to be a (partial) effect of the programme

intervention at the start. Rather, the arrow starts from the circle that includes all the factors that must cooperate with that one if the next step is to be produced.

When, as is not unusual, there is more than one set of factors that can cooperate with a cause to produce the desired effect, as in the blocklisting example, we represent this with different circles surrounding different factors at the head of an arrow, as in Figure 12.

**Figure 12: diagram of different sets of support factors that can help the same cause influence the same effect**
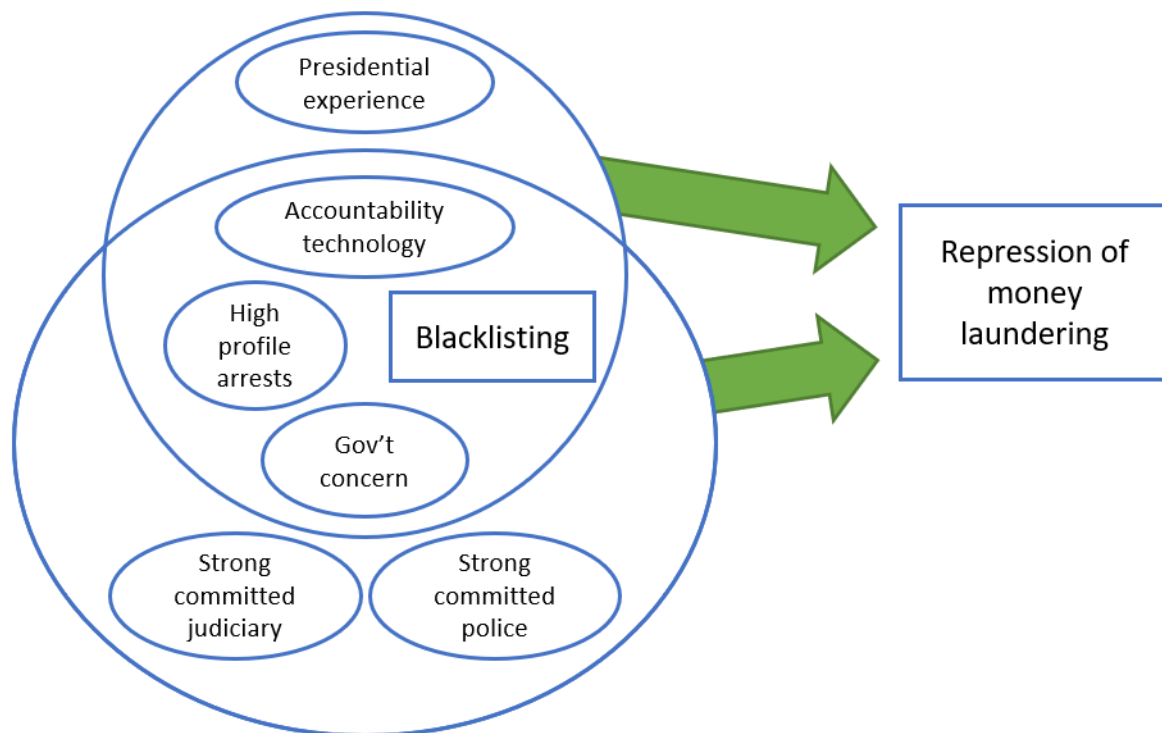


It is important to be careful to avoid excessively high expectations. Even in the most well-researched, well-understood cases (and even ignoring the effects of derailers, which we discuss in the next section), getting a full set of all known support factors in place may not secure the effect required. Our certainty about our predictions for success must be adjusted to reflect the extent that these are unknown. Our plans for implementation, monitoring and stopping the programme in the case that things are not going well must reflect the degree of uncertainty involved, the relative benefits, and the costs of success and failure versus the costs of monitoring, among other things.

Support factors may be unknown for at least three different reasons. The first is because at any stage the highlighted cause and its support factors are seldom the only set of causes that affect what happens next. Multiple other causes will be affecting the outcome independently of the ones implicated in the programme pathway, often pulling in different directions. The most we can expect from even a full set of mutually supporting factors is

that they make the **contribution** expected at that stage: they pull the actual outcome in the right direction but do not dictate what the outcome will be.[28]

We will discuss contributions further in our more detailed discussion in the Appendix of middle-level tendency principles. But note here that the term 'contribution' is not being used to deny causation or to suggest that we cannot with good reason attribute an influence to the cited cause. We are talking throughout about factors causally influencing effects. Consider a school physics example for clarity. Gravity pulls the cup down, the table holds it up, the cup does not move. Gravity undoubtedly causes a downward-directed contribution to the force on the cup even though the actual force the cup experiences is zero, which is why it does not move.

The second reason that the expected outcome may not come about is the well-known one that our knowledge is always fallible and incomplete, especially when it comes to the kind of complex settings of development programmes. No matter how careful we have been and how long we have worked on it, some necessary support factors will likely have been missed and others may have been included that need not be there. This is why you need to monitor carefully and evaluate support factors and update programme planning with emerging evidence about the programme's progress. You might include a wedge containing a question mark in each causal pie as a reminder that it is likely that what is there so far is not all that is required.

In addition to omitting important features from the diagram because we are not aware of them, we will have to omit many we are aware of because we do not have infinite space and those can reasonably be taken to be implicit. The diagram is a modelling abstraction from reality and is therefore necessarily reductive. Constructing a useful diagram requires balancing clarity and completeness. The best that one can do is to be cognisant of this fact and to make reasoned decisions about what to include and what to leave out, as well as to be aware that some elements will be missed entirely. To this end, we also recommend that that a pToC diagram be accompanied by a narrative to disentangle and discuss the aspects that it is impossible to capture in a diagram.

Third is that there may be no general variable that could be filled in. Even if, when we get to the local pToC for a specific concrete setting, we find something that can be added that will secure the required contribution in that setting, it may not be generalisable – there is no description to be given of it that makes it apply more widely and thus makes it appropriate in the middle-level pToC. This worry is exacerbated by the fact that causation in these sorts of context may just be dicey. As G. E. M. Anscombe put it in her inaugural lecture at Cambridge University,[29] there may be in place what is usually enough to produce the

---

[28] We have been describing a pToC for *this* programme to predict whether it will make the desired contribution. Even if it does, as noted in the text, that does not guarantee that you will see the results you expect because other factors may independently make contributions, either positive or negative, so you may see more, or less, than you expected. To predict actual outcomes in a local setting, ideally you should have a full causal model describing all the pathways into the final result. Since those that act independently are neither supporters nor derailers for *this* programme they will not appear in a pToC for this programme. Ideally, the full causal model should also chart out other consequences – side effects – of the programme process in consideration since these matter too for overall welfare outcomes.

[29] Anscombe 1971, 90–91

contribution wished for, yet occasionally it just does not happen or the cause does not make as big a contribution as expected even though a full pie of factors was at hand. We must not think of causation as deterministic. Having enough need not guarantee the outcome. 'Enough' does not mean 'sufficient' in the logical sense, where if A is sufficient for B, then there will be no As without Bs. A full causal pie may not secure a contribution but only change its probability.[30]

These reasons together underline the importance of the usual advice that even in the best of cases we should be cautious, monitor carefully as the process evolves and hedge our bets.

*Assumption types 4 and 5: derailers and safeguards*

That there is many a slip twixt the cup and the lip is a familiar truth in all social policy areas. Everything can be in place for a policy or programme to succeed, then something unanticipated happens to derail the process. In the area of child protection, Eileen Munro tells of the mother who is faithfully going to parenting classes as recommended and whose children are faring better – until the local bus service that takes her there is cancelled or her violent boyfriend moves back in.[31]

'Derailers' are things that can intervene to stop a full set of causes from producing the expected contribution or substantially diminish the effects. In our template for a pToC, we represent derailers in boxes with jagged outlines along the causal arrow joining cause and effect to show the possibility of breaking the causal chain at that point, as illustrated in Figure 13, which is taken from our mHealth example.

As development economist Esther Duflo noted in her address for the 2019 Nobel Prize in economies,[32] many times the problem with introducing change into a pre-existing system is that there are other policies, restrictions or practices that prohibit the change from being effective. For instance, in the Indian study she discusses in the lecture, Duflo discovered that many teachers did not use the techniques taught to them because they needed to conform to the legal curriculum standards and they lacked the time to do both. In the mHealth case, health staff may, for a number of different reasons, have higher priorities than classifying growth status, so they do not get around to doing so, or not in the timely way the programme requires. This would derail the programme near the start.

---

[30] The judgment as to whether causation is deterministic or dicey in a given case can affect what methods are good to employ for causal inference. Qualitative Comparative Analysis (QCA) looks for a set of factors (a causal pie) that is sufficient in the logical sense for the effect. More probabilistic methods can be more appropriate for handling indeterministic causation. For other discussions of this, see Davies 2018 or Mahoney and Goertz 2006. For a serious philosophical discussion see the *Stanford Encyclopaedia of Philosophy* entry on probabilistic causation (Hitchcock 2018).

[31] Personal communication with N. Cartwright

[32] https://www.nobelprize.org/prizes/economic-sciences/2019/duflo/lecture/

**Figure 13: representation of derailers from the mHealth example**
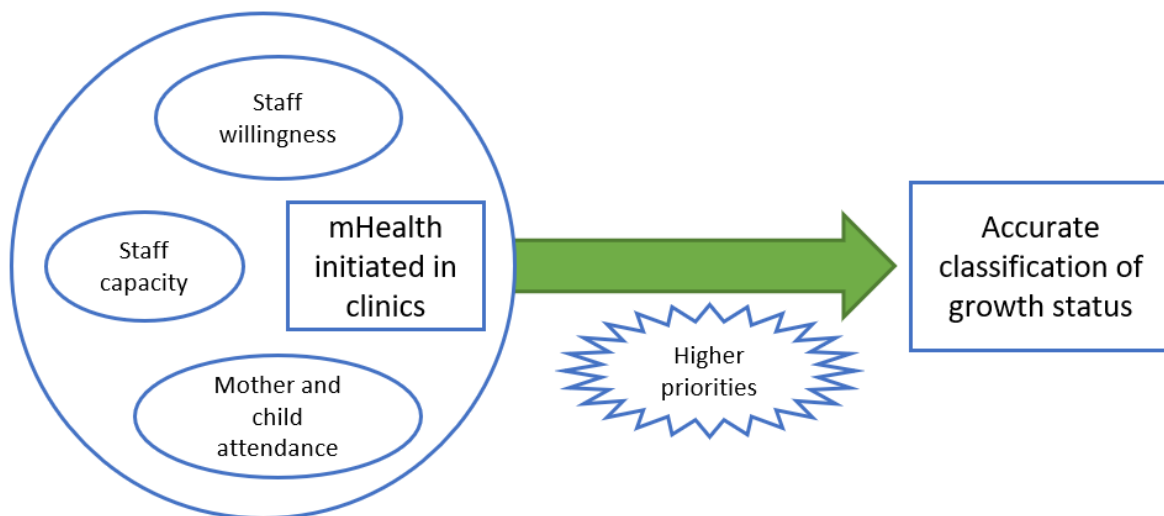


Many of these are unforeseeable – bad luck happens and programmes fail. But many are foreseeable. Programme developers and those implementing programmes at the local level need to work hard to identify what these are so that they can be guarded against or, failing that, the programme can be passed over in favour of one that has a better chance of succeeding. Many of these will be local to a particular target setting and can only be identified from knowledge about that specific setting (such as 'beware violent boyfriend returning'). Many will apply across a given kind of setting. Others will be fairly widespread. In laying out a middle-level pToC, it is important to envisage as many derailers and safeguards as possible.

It is also important to include the information that can be obtained about how to prevent possible derailers or to stop them from harming the process. We call factors that can do this 'safeguards'. These are represented by walls on both sides of the causal arrows that prevent derailers from intruding, as in Figure 14, which follows from our previous Figure 13.

**Figure 14: representation of safeguards from the mHealth example**
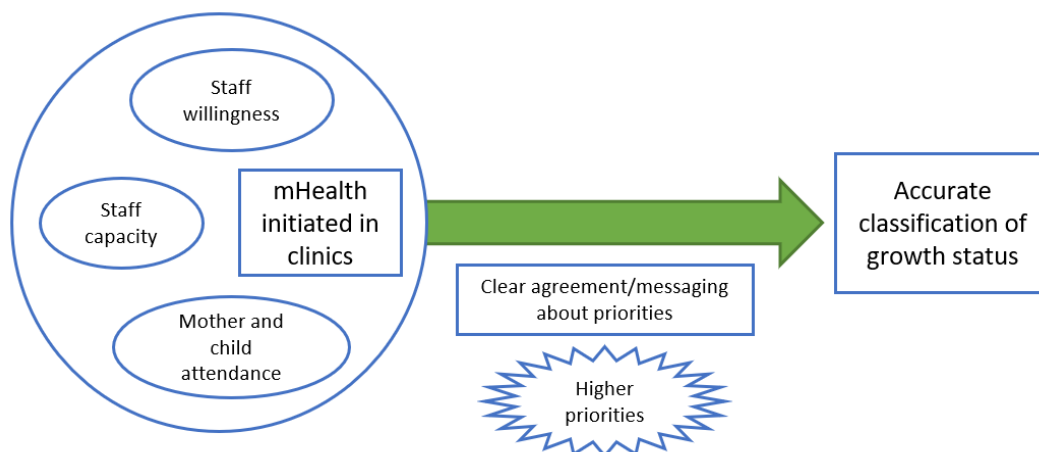
We should note that these are not natural categories but, rather, ways to organise thinking. There will always be a variety of ways of dealing with the same information: a factor that is required may be classed as a support factor, or, alternatively, the lack of that factor may count as a derailer. We keep both categories because many things are easier to take note of – and to justify – when thought of under one label rather than the other. The community health workers in the mHealth programme must be willing and able to input data to phones and to send the data off – support factor. But a sudden electricity outage that prevents the message going out probably occurs more naturally under the heading 'derailer'.[33]

Remember that the purpose of a middle-level programme pToC is to present as much information as possible to help those making decisions about local use build a similar but much thickened pToC for the setting they are dealing with. Even where no general derailers are known, the programme pToC could have some empty boxes with question marks to remind these decision makers and their advisors to consider what these might be at each step in their setting.

Putting in all the types of assumption reviewed so far, a standard general pToC will start off looking something like Figure 15 and then proceed step by step in the same way.

**Figure 15: one step in a standard middle-level pToC**
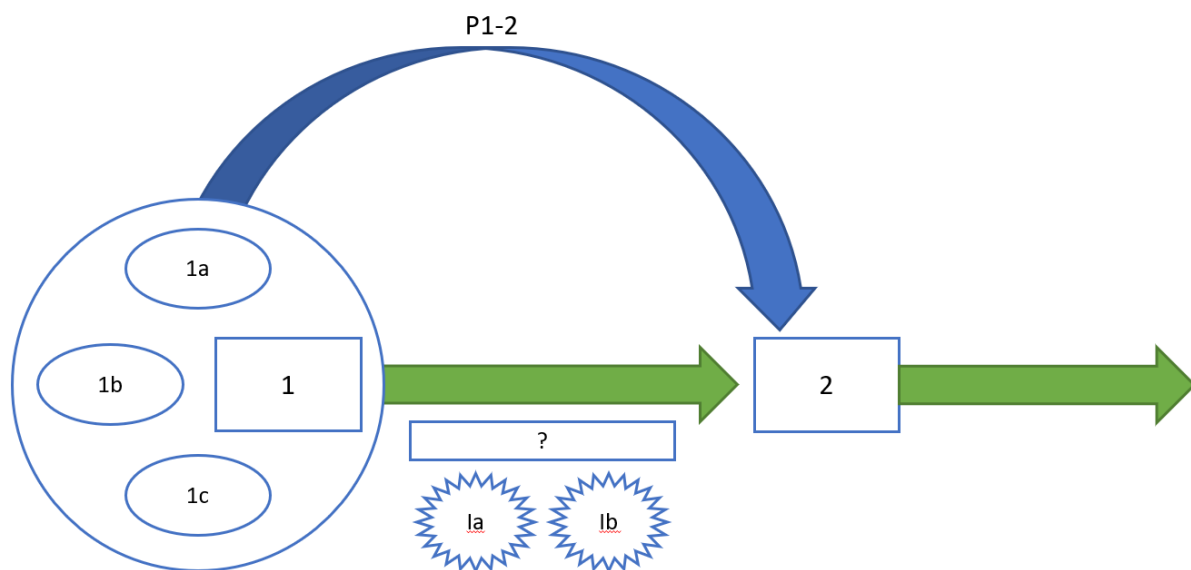


We note again that much of what we recommend here can seem obvious. Yet programme ToCs are not generally – if ever – constructed in a form that takes note of it all. Programmes fail over and over again due to the lack of support factors at various stages or the intrusion of derailers that were foreseeable. As noted in Section 1, some of these oversights are so big as to be startling. Eileen Munro tells of the trainee social worker so intent on filling in the required boxes on the form describing the family situation that they failed to notice the donkey in the kitchen.[34] California passed legislation to reduce school class sizes dramatically by start of next school year even though it was clear to many that not enough

---

[33] Barnett et al. 2016
[34] Personal communication with N. Cartwright

new teachers and classrooms could be made available in time. The Millennium Change Corporation (MCC; a US agency), supported training for farmers on using irrigation although the irrigation scheme would not be completed for some years – and then evaluated the impact of the training! But we do not need a big failure for a programme to be undermined. A causal chain is only as strong as its weakest link. So, a failure of necessary support factors or the appearance of derailers at any stage can prevent the desired results from occurring unless the problems are worked around (or removed) in some way. That is why we urge so strongly the need for good local pToCs and for programme developers to construct general programme pToCs that will be of maximal use for building these badly needed setting-local models.

*Assumption type 6: range of application*

Everyone knows that a programme that has definitely worked somewhere may not work anywhere else and, conversely, that one that has failed even in a number of places may work well in just the right setting. We have no miracle cures for the problem. But detailed pToCs for the programme of the kind we endorse here at both the general and local levels go a long way in probing where and why the programme is likely to work and where and why it is unlikely to do so.

First, the overall programme theory gives broad clues about where a programme may or may not work. mHealth will not contribute the expected degree of improvement where community health workers are already good at following the formula for calculating nutritional status, where the formula itself is not all that accurate and experienced health workers can do better looking at the data and at the children themselves or where the technology in the phones is apt to fail.

When we look at the six kinds of assumption we urge to be used in constructing a pToC, we get even more information about what the settings of successful application will be like – much more – though they do not provide the short, sharp descriptions we might wish for, such as 'degree of democracy', 'good governance', 'women's participation', 'foreign direct investment' or 'gross domestic product'.

Think about all those middle-level, mostly tendency principles under which one step produces the next. These are not to be treated as universal or near universal truths in the way we may treat the law of gravity. They often describe individual or institutional dispositions that are peculiar to certain types of people or institutions. Even in types of settings where they are widely prevalent, they may need effort to trigger or are likely to be quashed by competing dispositions. By considering the types of people, institutions, geographies, cultures and so on that are implied by the different assumptions and principles that constitute middle-level pToC, a list of the types of setting in which that pToC will tend to apply can be derived. An important part of the development of a middle-level programme pToC is to reflect on this list of settings of application and to specify them.

The first stage in deciding whether a given programme might be suitable for implementation in a given setting will be to reflect on whether it satisfies the description of one of the types of setting in which that programme might be effective. For example,

consider the pToC developed for the CCTs aiming to increase school enrolment described in Section 5.4. The overall programme theory is stated as follows:

> Households tend to spend in their children's best interests, but financial barriers (insufficient resources to meet the direct and indirect costs of education) and non-financial barriers (such as incorrect beliefs about the returns from education, excessive future discounting, and intra-household bargaining problems) can cause them to under-invest in schooling. Therefore, alleviating these barriers by providing resources, nudges and incentives to enrol children in school tends to lead to higher levels of enrolment of children in school.[35]

This programme theory tells us something about the settings in which this type of programme might be expected to be effective. For instance, these settings must contain many households that are financially and/or non-financially constrained from investing in children's schooling. The programme aims to act on the demand side of the relationship between households and schools rather than on the supply side. In order for it to be effective, there must therefore be a problem of demand rather than one of supply for school places for children. In a setting in which there are not enough school places available, a CCT will not tend to result in an increase in enrolment.

We can find further clues about the types of setting in which the programme might be effective by examining in more detail the assumptions that make up the middle-level pToC, which is shown in Figure 16. For example, it is clear that the intervention contains three components. Cash is transferred to households (box 1), those transfers are labelled in some way as educational (box 1') and it is communicated that continued transfers are conditional on children's attendance at school (box 1''). The programme is therefore expected to be most cost-effective where all three of those components tend to be effective.

In some settings, the effectiveness of the programme will be undermined by the fact that one of the aspects of the programme is unnecessary. This might be true, for example, if households did not lack resources to send their children to school but rather chose not to send them through distrust of the school. In this type of setting, although a CCT might induce some behaviour change by providing an incentive to send children to school, it would be much less effective than in a setting where the direct income effect expressed by the causal relationship between box 1 and box 2 was activated. Likewise, in a setting in which households do not face strong non-financial barriers to enrolling children in school, such as excessive future discounting, the costs implied by conditionality might not be acceptable, suggesting that an unconditional cash transfer might be more cost-effective.

---

[35] For more detail, and for references for the ideas contained in this sentence, see Section 5.4.

**Figure 16: the assumptions that make up a middle-level pToC**
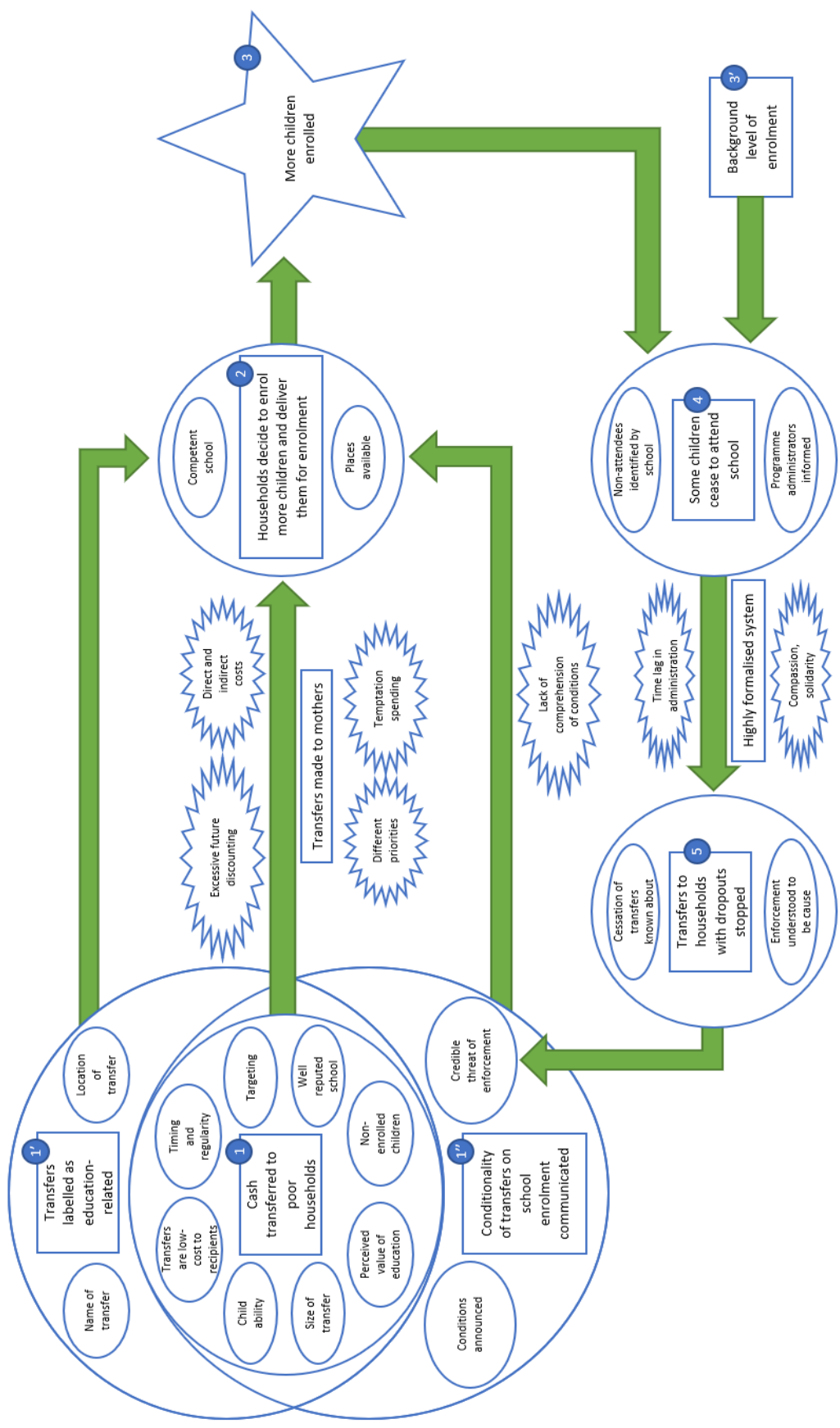
By assessing every component of the middle-level pToC in turn, we can add more information to the description of the types of setting in which the programme might be reasonably effective. For instance, we can identify that schools and the institution administering the transfers must have sufficient capability and capacity to reliably record absence from school and apply the programme conditions, ultimately stopping transfers to some households.

Based on an analysis of the middle-level pToC for CCTs aiming to increase school enrolment, we can then make broad statements about the types of setting in which the programme might be expected to be (cost-)effective. However, when it comes to considering implementing the programme in any specific setting, much more work will be required. In order to assess the feasibility and desirability of implementation for a specific setting prospectively, or in order to evaluate such an implementation retrospectively, it is necessary to thicken the middle-level pToC to create an applied ToC. We discuss why this is necessary and how it can be achieved in Section 3. First, the next subsection describes how thinking about middle-level principles allows us to identify support factors, derailers, settings of application and safeguards.

## 2.4 The importance of middle-level principles: identifying support factors, derailers, safeguards and settings

Here we provide some guidance on how to identify these. This is a hard job, but there are a number of well-known methodologies for doing it. Some are common sense, such as 'if you want stakeholders to be on board, do not humiliate them or belittle their input in planning and negotiation meetings'. Many are suggested by theory. Elinor Ostrom famously argued that free-riding tends to derail collective action.[36] Examining her theory for why this is the case, Ostrom argues that small-scale communities tend to form institutions and norms, which enable them to successfully overcome free-riding but the ways in which this happens are not readily available in larger communities. Others are unearthed by standard scientific methods, such as subgroup analysis in randomised controlled trials and in other statistical studies, qualitative comparative analysis, comparative case studies and causal–process–tracing techniques.

What we have added to this is the role of the middle-level principles that govern each stage in the causal process from start to finish. These are invaluable in figuring out what must be in place for one step to lead to the next and what can get in the way. This is a lesson that realist evaluators have long stressed with respect to the overall programme theory (or 'mechanism' as they label it). For instance, from Pawson and Tilley (1997), CCTV cameras in car parks are supposed to diminish car theft. Do they work under the principle that criminals try to avoid being identified? If so then the cameras should be clearly visible, so that criminals believe they are working properly and will be monitored. Or do they work under the principle that if the police are alerted to a crime in time, they can act to prevent it? In the latter case, the cameras should be hidden, they should be monitored throughout peak crime times and the police should be well positioned to get to the car park quickly.

---

[36] Ostrom 1965; 1990

It is this kind of interplay between support factors, derailers and principles that guided us in adding support factors and derailers to our middle-level programme pToC for mHealth. For instance, consider steps 4 to 5 in the pToC in Section 5.2, in which accurate identification of underweight infants in the community health clinics is to lead to more timely submission of data from local to district health centres, especially in rural areas. We suppose that this happens because of the principle 'automated methods for doing so promote data submission'. This immediately suggests a number of support factors and derailers. The timely submission will only happen under this principle if the community health workers are able to submit the data that way, they are not prevented by external pressures or other priorities from doing so, they do not see other tools as dominating the automated methods and the technology works properly at that point.

Consider some examples from a very different domain. Figure 17 shows a causal process graph offered in the partial reconstruction of a legal case from a book on legal evidence.

**Figure 17:**[37] **a causal process graph in partial reconstruction of a legal case**
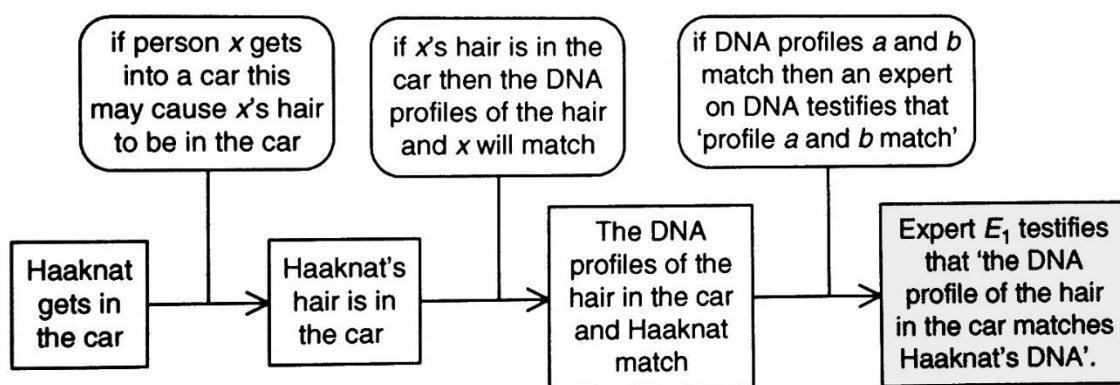


In Figure 17, the principles are indicated by rectangles with round corners, in contrast to our curved arrows. But the idea is the same. It is important for there to be a reason why each of the causal connections should occur. The principles meant to operate at each step provide those reasons. The principles in this legal example are then used, just as ours are meant to be, to guide the selection of support factors, as in Figure 18.

**Figure 18:**[38] **principles guiding the selection of support factors**

---

[37] Bex 2011, 71
[38] Bex 2011, 62
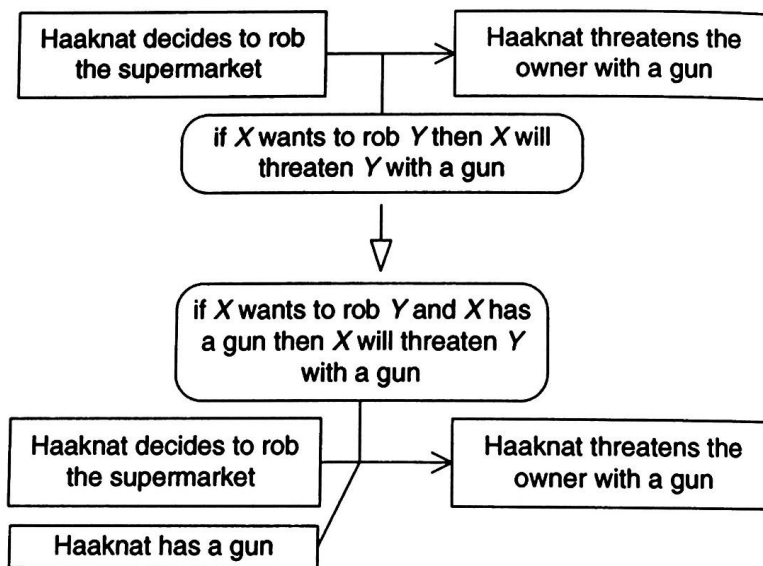
In Figure 18, the support factor for Haaknat's decision to rob the supermarket that results in his threatening the owner with a gun is that Haaknat has a gun. (Note that the support factors are listed below the salient cause with a line up to the causal arrow). Although the need to have a gun may seem obvious from the principle as originally stated, still it is not explicit in the principle, which is the reason for refining the causal principle at work to make it clear, as is done at the end of the downward arrow. Given the importance of the principles for identifying support factors and possible derailers, it is important that the causal principles for each step be stated as fully as possible. There is a very reasonable call in realist evaluation and elsewhere to identify the activities by which one step causes the next. However, we stress principles because the same activity can fall under different principles and it is generally the principles that are most helpful for identifying support factors, derailers, safeguards and plausible contexts of application.

### 3 Thickening the middle-level pToC for local use

This section describes what we mean by thickening a middle-level pToC for a programme for a specific local setting.

Recall our warning from Section 2.3 that a causal chain is only as strong as its weakest link. A failure of necessary support factors or the appearance of derailers at any stage can prevent the desired outcomes from occurring unless the problems are worked around (or removed) in some way. That is why we strongly urge programme developers to construct middle-level programme pToCs that will be of maximal use for building these much-needed setting-local models.[39]

---

[39] For some programmes, either in general  or in some settings, not enough is known to construct a good pToC. Such programmes could be implemented using an adaptive approach that develops programme and pToC in parallel though rapid, iterative evaluation and adaptive programme design.

In the section dealing with the 'assumptions' that go into developing a pToC, we describe how a middle-level pToC should be examined to identify and spell out the types of setting in which the programme might be effective. For any candidate setting in which we are considering implementing the programme, the first question should be: 'Does this setting match the broad description of one of the types of setting in which the programme might be effective?' However, even if we can answer this question in the affirmative, much more work must be done before it can be considered adequate to pursue that programme. This is the work of thickening the middle-level pToC for application in that specific setting. In this section we explain why this thickening is necessary and how it can be achieved.

## 3.1 Matching abstract concepts to specific referents

Here we illustrate how the more abstract concepts that appear in the middle-level pToC should be made more concrete to refer to specific features in the local setting.

We have labelled middle-level programme pToCs 'middle-level theory' to contrast with the local pToCs that are to detail the specific factors that can affect the casual process linking input with outcome in a local setting. Middle-level pToCs will include some information about the settings in which a programme can be expected to be effective, but a reliable, well-justified prediction of effectiveness in a given setting will generally need a setting-local pToC. Given their aim for generality, middle-level programme pToCs will typically use more abstract concepts than the local versions of them, which must use context-specific descriptions if they are to serve as a serious guide for local planning, prediction and implementation. As the Italian political scientist Giovanni Sartori noted, 'the more cases to which we attempt to apply general categories for concepts, the more the latter have to be "stretched" and the less meaningful and useful they may become in identifying the appropriate empirical referents within the specific context of each case'.[40]

So, for local application, it will be necessary to figure out what satisfies the abstract description in that particular context. This will take both local knowledge and a good understanding of what is intended by the more abstract concepts. For instance, we often see discussion of the need for buy-in from local stakeholders as a prerequisite (in our language, 'support factor' for programme success). In mHealth, for example, in the early stages this could be buy-in from community health workers who previously carried out visits with mothers without the use of mobile phones, or buy-in from district health workers who previously conducted their work without the use of mobile phone data.

In the middle-level pToC we construct for mHealth, the support factors are that the community health workers should have the capacity to use mHealth and agree that using mHealth is good for their clients. In some mHealth programme settings, the abstractly described 'health workers' at this stage are in fact volunteer community health staff. Local knowledge supplies the information that community volunteers may not receive adequate training on how to use the technology, so the presence of this support factor cannot be assumed. Likewise, one of the derailers in the middle-level pToC is that technology fails. In some programme settings of mHealth, the relevant technology is mobile phones or

---

[40] See Sartori 1970 and Collier and Mahon 1993. This description of Sartori's claims is taken from Rudra 2000.

smartphones that need electrical charging and an operating mobile phone network. Local knowledge indicates that signal is limited in some of the community health clinics and electricity for charging the mobile phones is problematic at times, so this derailer is a realistic possibility that needs guarding against if the programme is to be expected to succeed.

Adaptation of a pToC to suit the context in which it is intended can be an important step in the thickening process. We suggest that it will almost always be necessary because the middle-level programme pToC is blind to particular local conditions. Similarly, the local conditions that applied in some other setting where the programme worked are unlikely to be the same as those that matter in this setting. Without taking the local context into account, certain theories may not be supported by the local conditions.

There are a number of adaptation frameworks available that provide advice on how to adapt specific recommendations for context-specific processes (for example, the Australian National Health and Medical Research Council's advice in guidelines[41]). The important point is that genuine knowledge of the local setting must be added in at this point if responsible, reliable, well-supported predictions are to be expected about outcomes in the local setting.

When thickening a middle-level pToC for local use, every abstract component must be replaced by a local referent where such a referent is known. If no such referent can be found, then either it must be searched by conducting more extensive research, or it must be accepted that this aspect of the middle-level pToC is not applicable to this setting. That might be positive for a prospective assessment of programme success, as in the case of a possible derailer at the general, abstract level having no local relevance. Alternatively, it might be negative. If no local referent can be found for an essential support factor of programme causation, the possibility that the programme will not be effective must be seriously considered. Recall that pToCs can be useful both for *ex ante* prediction and *post hoc* evaluation. In thickening a middle-level pToC for retrospective evaluation, it might be the case that an essential support factor is missing yet a positive effect is still observable. This would suggest that the middle-level pToC is not a good explanation of success for this setting, prompting the search for a better explanation.

Consider the case of thickening the middle-level pToC for CCTs aiming to increase school enrolment of children described in Section 5.4. Taking the first box and its associated support factors, causal effect and derailers, we can see how the general, abstract terms can be rendered specific and concrete for a setting in North Macedonia. This example is based on the detailed report by Armand and Carneiro of a real CCT in this setting.[42] Figure 19 shows the relevant portion of the middle-level pToC. Figure 20 shows the same portion of a thickened local pToC.

---

[41] NHMRC 2018
[42] Armand and Carneiro 2018

**Figure 19: middle-level pToC for CCTs aiming to increase school enrolment**
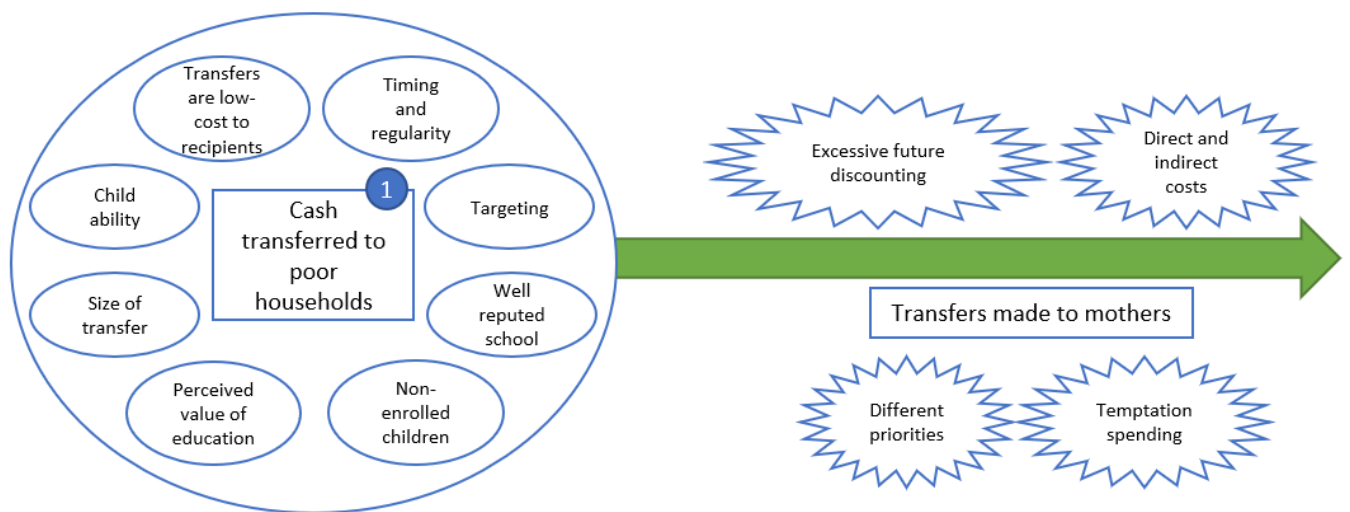


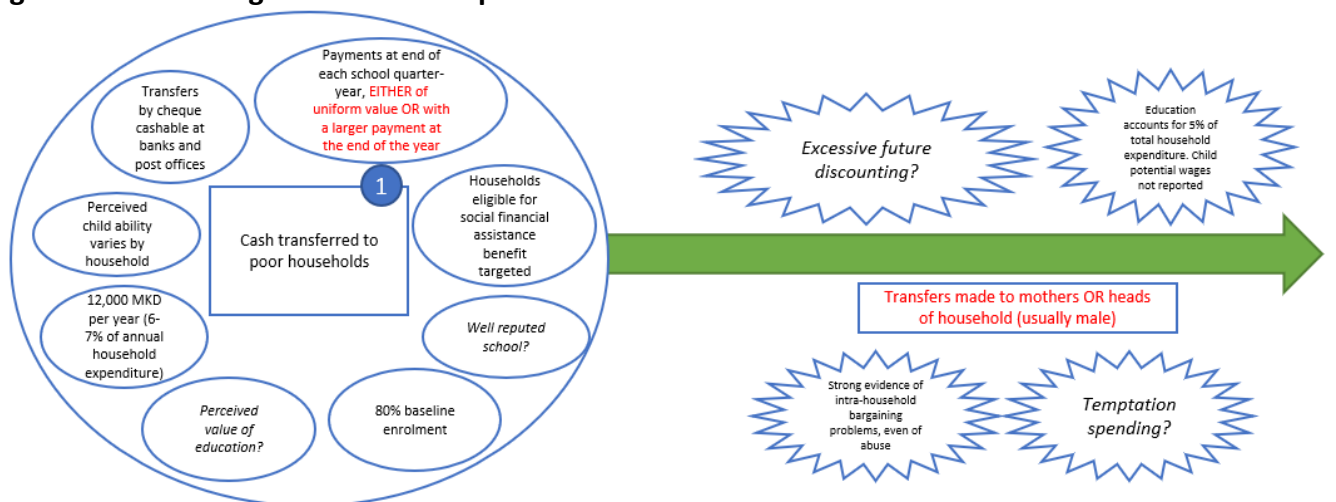**Figure 20: thickening of middle-level pToC for CCTs to increase school enrolment**



Figure 20 is more specific and more concrete than Figure 19. In addition, we see that two derailers have not been successfully thickened. This is represented by a restatement of the general derailer in italics followed by a question mark. In this case, no information on the prevalence of excessive future discounting or temptation spending in the setting were reported by Armand and Carneiro as part of their description of the setting. This information may have been unavailable, or it might have required further research to uncover. Ideally, when thickening a middle-level pToC for use in a specific setting, it should be established whether all the components of the general model have a local referent and if so, the specific nature of that referent. Leaving key support factors or derailers unthickened increases the risk that a programme will not work as well as the risk of unintended consequences.

**3.2 Drawing down and reading up**

Here we explore the relationship between middle-level and local pToCs to see how they inform each other.

An additional feature of Figure 20 is that a support factor and a safeguard have been specified in two different variants, represented by red text and the use of 'OR'. This is because the implementation and evaluation teams behind this programme decided to test their underlying theory of programme causation by deploying four different variants of the programme. Half of recipient households received transfers of equal value each quarter. For the other half, the final payment was larger than the others to increase the value of the conditionality and to measure the effectiveness of this strategy. Independently, half of household transfers were made to mothers of the children on whose attendance further transfers were conditional, whereas for the other half transfers were made to heads of households (usually men). This variation was induced in order to test the assumption that transfers to mothers are more effective. The authors of the study hypothesised that since, in the Macedonian context, fathers are very dominant in the household, they may have been better able than mothers to induce their children to attend school. This may counter the general tendencies for men to spend on adult goods (alcohol, tobacco) and women tend to spend on family.

Armand and Carneiro are far from alone in their desire to use a local application of a middle-level theory to test that middle-level theory. This operation is fraught with difficulty because it is extremely challenging to assess the extent to which local observations (experimental or not) are informative of more general tendencies, as one of us has written elsewhere.[43] However, our local pToC and its relationship to the middle-level pToC can help. By thinking carefully about the extent to which our observations fit with the predictions of our local pToC, we can reinforce or undermine an argument that it is the best explanation of causation in our context.

If there is no clear locally specific explanation of apparent deficiencies in our middle-level model, then the model may be at fault. With the accumulation of such evidence, a strong argument might be made that the middle-level model should be modified in some way. The process of thickening our model for local application is not a one-way process. As well as drawing down from middle-level theory to inform our thinking about a specific setting, we may also have opportunities to read up from local findings to improve our middle-level theory.

**3.3 Adding locally specific assumptions**

This section explains that, besides filling in what the more abstract terms amount to locally, it is also important to add into the setting-specific pToC all those additional factors that may not matter in general but will make a difference in that particular setting. These are often far easier to recognise with hindsight once something has gone amiss, so they often play a role in *ex post* evaluation. In an extreme example, Benhassine et al. report two of their

---

[43] See Deaton and Cartwright 2018, or for a longer treatment, Cartwright 1989

study communities being rendered unreachable by floods.[44] However, it is clearly crucial to identify as many locally specific factors as possible ahead of time. This is what Armand and Carneiro did when they identified an unusually strong level of control of children by their fathers in their Macedonian setting.[45] This prompted them to suggest that a generally employed safeguard ('Try to ensure the mother has control of the resources') might be counter-productive in their setting and to test whether that was in fact the case.

There are a number of strategies that can help with identifying locally specific features of programme causation. For instance, one can attempt to imagine all the ways in which the programme might fail to work or might have unintended consequences in the specific setting. Another is to think through the process backwards, from endpoint to start, as well as forwards. Generally, this requires a good deal of local knowledge. The availability of this knowledge is crucial to making reliable predictions and evaluations about local effectiveness.[46] Assessing this list provides a list of potential derailers that should be specified. Reflecting on potential solutions to those hypothetical problems might provide some additional support factors that should be included.[47]

To do any of the thickening activities described in this section will require concrete knowledge of the local situation. Some of this will be at hand, and some might be uncovered by serious thinking and review of what is known about the local setting on the part of those designing or implementing the local programme. On the other hand, it may only be possible to address knowledge gaps by locally focused research. It may very often be the case that a fully thickened local ToC only emerges after an extensive period of scoping and descriptive research, especially if we are to be appropriately sensitive to unknowns affecting a new type of programme or one that is being tried in a new type of setting. Coming to a reasoned judgement about all the many assumptions that go into a causal model clearly will require effort. How much of this effort is worth it? We have only the standard advice: that depends on the costs (not just financial costs of course) of gathering, negotiating and using this information versus going ahead without it and either implementing a programme that fails or failing to implement one that might have succeeded.

For any actual implementation that is under consideration, although all this thinking ahead is essential, it must also be remembered that in real settings things are alive and fluid and everything tends to affect everything else. The extent to which dynamic changes in a given setting are worth monitoring and reacting to will be implied by a properly thickened local pToC that bears those changes in mind. Perhaps an adaptive, iterative program design is more appropriate than one in which the form of the programme has been designed not to change for several years while detailed data is collected on long-term changes. Or, conversely, this latter programme and evaluation design may be more appropriate. There are no standard answers to such questions, but properly reflecting on programme theory in each specific local setting can help.

---

[44] Benhassine et al. 2015, 92

[45] Armand and Carneiro 2018

[46] This naturally invites the next question: 'How is this knowledge of local facts to be uncovered?' We cannot tackle this question here. Rather our aim is to help by detailing what kind of knowledge will be needed.

[47] See Cartwright and Hardie 2012, especially pp. 95–97 for more detail on this and other strategies.

*4 Help with planning, evaluation and monitoring*

We promised at the start that constructing middle-level pTOCs and then thickening them for local use in the way that we propose will help to support good programme design, good monitoring and good evaluation design. In this section we show how this works, using examples from the three case studies mentioned throughout the paper. We have just explained in Section 3 how to use middle-level pToCs to plan programmes for new settings. So we will address planning here only briefly, then pass on to monitoring and evaluation.

## 4.1 Planning and transferability

This is the only place in this paper where you will see the word 'transferability'. That is because we think the term is misleading. Evaluators are often urged to construct evaluations so as to make their findings as transferable as possible to new settings. This focuses attention on the local settings. But no amount of information about one setting can tell you anything about any other without strong middle-level theory. Talk of transfer makes it sound as if you can make inferences directly from one setting to another without the need for serious theory development and justification. It hides the essential role of theory and the importance of developing good theories, verifying them and justifying the ways they are used. When it comes to helping with planning for new settings, evaluation should be aimed at developing, refining and testing theory that will hold in the targeted new settings and pointers about how to thicken that theory for different settings.

Often evaluation is used to help develop and refine what we have called the 'overall theory' for a programme. We have been urging that it is important to go beyond that, to expand the programme theory to include a good filled-in middle-level pToC of how the programme is supposed to achieve the intended outcomes step by step. Middle-level pToCs are a major tool for using evaluation findings from one setting to inform programme design in another. The underlying logic is to 'go up' from the findings of how a programme worked in specific settings to build a middle-level pToC that accommodates those findings and then 'go down' to inform the local programme pToC for a new setting.
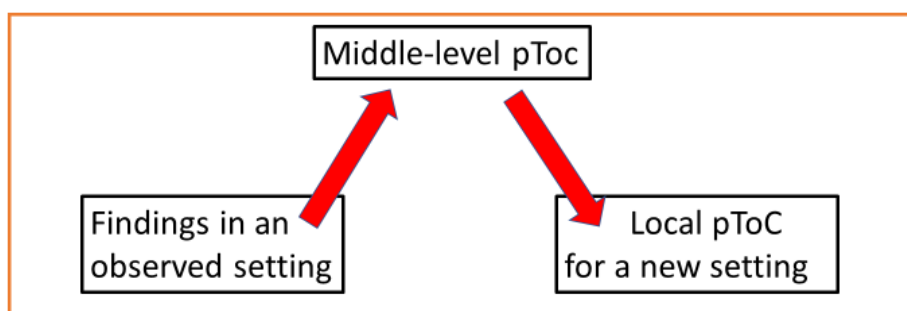


Our focus in this paper has been on the right-hand side of this diagram, on what a good middle-level pToC should include to make the downward inference easier and more reliable. Other CEDIL projects explore how to use findings to develop and refine the middle-level pToC, including the overall programme theory. It is important to note that far more than findings about how the programme works in specific settings will be needed for this. A vast amount of independent knowledge and research goes into understanding, developing and

justifying the causal principles under which the process is supposed to evolve and the understanding of what it takes to call all these principles into play in the same setting at the requisite times. In practice the logic is iterative. Findings from *ex post* studies help identify and refine the middle-level pToC that is then used to help design new programmes that are then monitored and evaluated to further refine the theory and to reassess past successes and failures to understand them better.

For example, Section 3.2 discussed Armand and Carneiro's design of a CCT programme, in which some recipients were given uniform payments throughout the year, whereas others were given a larger payment at the end of the school year if their child was still in school. This variation was driven by an uncertainty about the precise nature of a supporting factor required for the action of the programme. By changing the form of that supporting factor for different sets of recipients, Armand and Carneiro were able to test which form helped the action of the programme most, helping them to refine their local ToC and to inform the middle-level ToC for CCT programmes.

For a second example, consider the mHealth programme. The *ex post* findings from Indonesia suggest for the middle-level pToC that mHealth programmes like this one will work only if several key features are in place, such as 'health workers operate in the client's best interest' and 'mothers' trust'. Several derailers were found to undermine the programme there, including conflicting norms and priorities and external pressure not to perform tasks. In wider mHealth settings, programmes are only likely to work if the setting in which they are to be applied can find contextually relevant safeguards. The safeguards identified in the Indonesian setting may offer useful insights.

The anti-corruption pToC shows that public service reform is more likely to succeed with the right mix of factors, thus offering a number of planning suggestions. In particular, the integrity of leadership and new powers make organisations both willing and able to do what is necessary to implement successful reforms. This was a common thread among the pockets of success within Nigerian public sector reform. Integrity both protected the reform from external corruption and motivated efforts for reform in the face of persistent difficulties. In contrast, the reform efforts were often derailed when leadership lacked the internal motivation to push against the corrupt status quo.

## 4.2 Monitoring and evaluation questions and indicators

It is becoming widely accepted that evaluations of social programmes should not merely seek to evaluate to what extent, on average, the presence of the programme was associated with a change in the outcome(s) of interest. Making good use of the middle-level pToC to identify indicators of interest all along the causal chain allows us to go beyond this simplistic approach to ask to what extent the programme was functioning as intended, both for the purpose of monitoring that the programme is on track in time to catch problems before it is too late and for purposes of evaluation.

When designing a programme, we have suggested generating a middle-level pToC for the type of intervention that is being considered and then thickening it with information specific to the local context to create a local pToC. This process helps to identify the key causal steps

in the operation of a programme, the causal principles underlying the programme, the required supporting factors for it to operate, potential derailers to its proper operation and safeguards that should be implemented to prevent those derailers from acting. These aspects of the programme theory can all be used to generate questions and indicators to be tracked for monitoring and evaluation.

In the mHealth case study, many of the support factors relate to the role of the community health worker and/or the district health staff: for example, 'community/district health staff have capacity' and 'mothers understand the advice'. The importance of these supporting factors raises evaluation questions related to the structure and functioning of key health services and the impact of new technology on an existing health system. If an evaluation were to discover that these essential features were missing, then this should be taken as evidence that the programme was not responsible for the outcome.

Beyond this, the mHealth middle-level pToC suggests several monitoring and evaluation indicators that may be adapted to future programme settings, for instance: correct entry of weight into the mHealth application, timely entry of weight into mHealth application, delay in data sent from community health clinics to district health centres, mothers' feelings towards clinics and reattendance rates at community health clinics.

In the anti-corruption case study, the support factors often relate to the type of leadership in charge of reform. In particular, the integrity of the leader is the lynchpin that enables them to persistently use the available mechanisms for successful reform. This shifts evaluation questions away from the centrality of the external reform mechanisms and towards the centrality of leadership and the positive influence of integrity over and above more orthodox leadership qualities such as qualifications or professional status, for instance: 'Has the potential leader displayed long-term integrity?' 'Is the potential leader internally motivated to achieve reform?' and 'Will the leader gain the powers necessary for organisational reform?'

In turn, the middle-level pToC and the PERL programme findings suggest evaluation indicators that might apply to future programme settings, for instance:

- Integrity-based resilience against corruption.
- Institutionalised authority to reform organisational structures.
- Organisational confidence in the leadership.
- Public visibility of reform.
- Public confidence in the leadership.

For examples from the CCT case study, re-consider Figure 16:

**Figure 16 (repeated): the assumptions that make up a middle-level pToC**
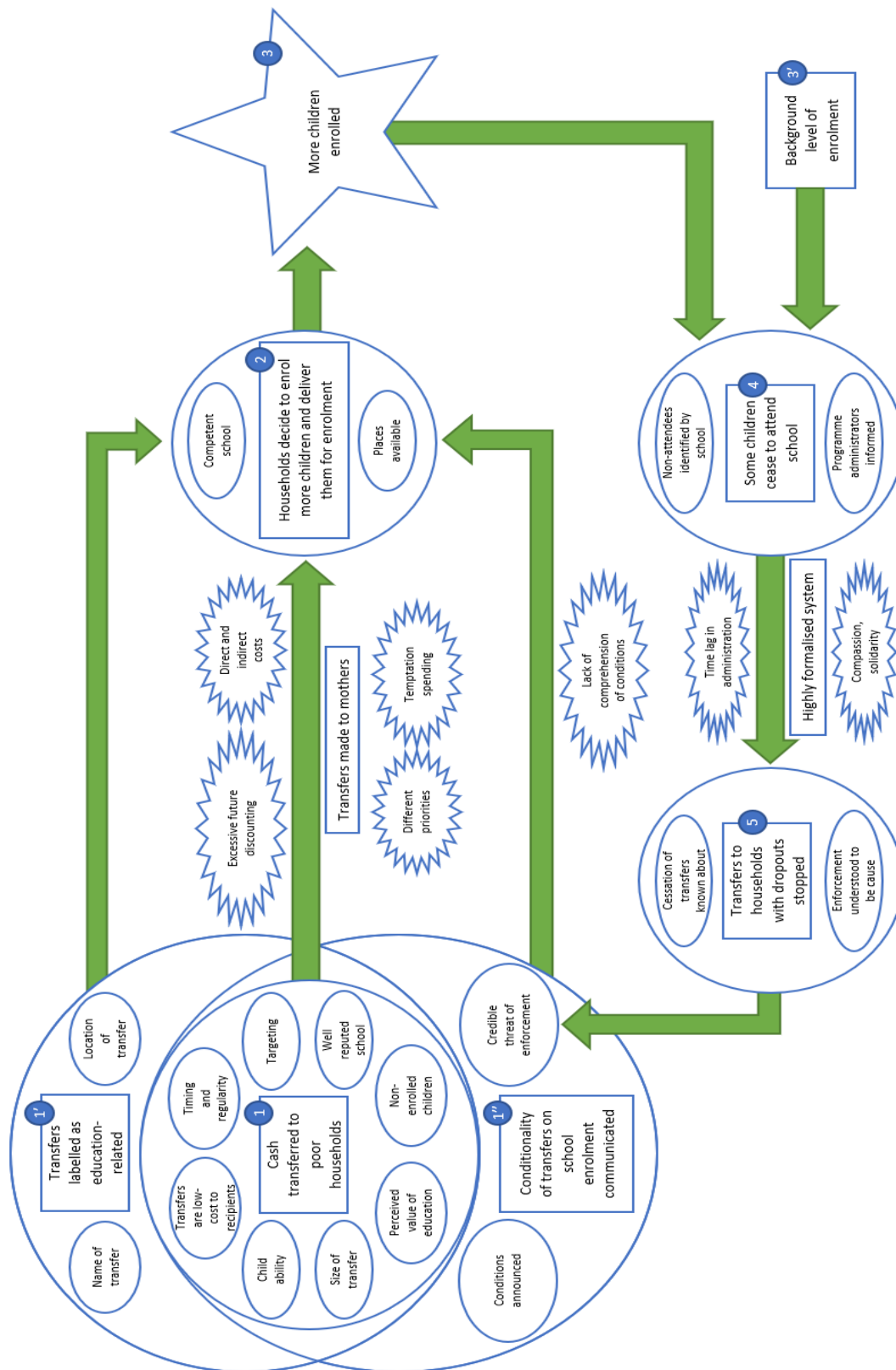
The key causal steps provide indicators that should be tracked as part of a monitoring and evaluation programme. It is obvious that whether cash has been transferred to households should be tracked as part of the evaluation of a CCT programme. Less obvious indicators can be identified by examining the pToC. Tracking the presence of potential derailers and the safeguards against them will help to facilitate insightful evaluation that stands a better chance of being able to explain unexpected failures or unexpectedly large successes in a programme. For example, the pToC contains a derailer labelled 'Time lag in administration'. This reveals that it will be important to track process information about how long the programme administration takes to register a student failing to attend school and to translate this into a warning or a stopped payment to that household. If this period is long, then the support factor 'Credible threat of enforcement' can be expected to have been undermined, reducing the effectiveness of the conditionality of the cash transfer. Likewise, the fact that 'Lack of comprehension of conditions' is a derailer to the action of the conditionality of the transfer means that monitoring households' level of comprehension of the transfer's conditionality will generate a useful indicator for the monitoring and evaluation of a CCT.

Recall that we have introduced pToCs at two levels: the middle-level pToC, which is meant to apply across a range of settings, and the local pToC, designed for a specific setting. The middle-level pToC suggests general types of monitoring and evaluation indicators that may be adapted to each specific programme context through thickening. Take, for example, the middle-level anti-corruption pToC used as an example throughout this paper and elaborated in detail in Section 5.3. The essential first causal step identified by this pToC is that 'leaders and bureaucrats with credentials signalling integrity are hired (over those signalling expertise)'. Any evaluation of an attempt to implement this programme would have to track an indicator for the extent of credentials signalling integrity in new hires. These credentials would be locally specific and would have to be identified as part of the process of thickening the middle-level pToC for local use.

## 5 Three case studies

### 5.1 Introduction

In this section, as promised throughout, we provide completed middle-level pToCs for development programmes of three different kinds: the use of mobile health technology to improve infant nutrition, the role of leaders identified for their integrity in fighting corruption and a CCT programme to improve educational enrolment and attendance. We have tried to make them as realistic as possible given that we have been constrained by the literature available and that these should not become so complicated as to be unusable.

After initial background material, we present the full pToC for each programme and then discuss how we have built it up it in order to give you a real sense of what this involves. Recall the diagram in Section 4 that pictures (via the left-hand upward arrow) starting from specific programmes that have been carried out in specific places and from there building up to the middle-level theory. That is what we have done. As a base we have used real programmes that have been implemented and evaluated, which we describe to you. To arrive at the middle-level pToC for these programmes we have then added our own theorising to what we have found others saying about these and similar programmes,

bringing in straightforward, widely recognised middle-level principles that we expect readers to be familiar with. In Section 3, we have already used the educational CCT case study to illustrate the process, pictured via the right-hand arrow in the diagram in Section 4,of thickening the middle-level theory to produce a pToC for a local setting.

We have done no original research ourselves. These three cases are meant as examples of what realistic completed pToCs will be like. They are not meant to help draw lessons about these particular programmes but rather to help you understand what should go into a good middle-level pToC. DFID chose the three general areas for the case studies: nutrition, democratisation and CCTs. Within these areas, we chose these particular programmes because we had reasonable access to information about them and because they provide good illustrations of the points we want to make. We want to underline that we were not the programme developers, so what we say is not intended as a recommendation.

The completed pToCs may look complicated, and we could have added even more detail, making them look more so. This last is problem you may often face in practice. To keep the pToC simple enough to be intelligible, it is useful to recall that you do not have to do everything in one diagram, nor do you have to use only one mode of presentation. You can, for instance, show different causal pathways in separate diagrams so long as you find a clear way of indicating how they relate. You can also consider presenting support factors for each step in a series of pies like the one in Figure 10, or offer lists of support factors, derailers and safeguards for each step. What matters is that all the categories are covered as fully as possible and that the user will understand where to find what.


**5.2 Case study 1: mHealth and Nutrition**

The use of mHealth to support development outcomes in nutrition is growing. mHealth encompasses a variety of programmes that aim to support governments to transform and modernise community health and nutrition services by introducing innovative mobile phone applications to support nutrition outcomes. Oftentimes, mHealth applications replace manual systems. For example, community health workers across 11 countries in Africa, South Asia and Southeast Asia have been equipped with mobile phone technology to provide mothers with access to basic healthcare information.[48]

The general programme pToC we construct in this case study is heavily based on a pilot application of mHealth for nutrition monitoring in Indonesia[49]. The project name is Posyandu or Posyandu Mobile Health. The project targeted a population of around 11,300 children under the age of five in North and East Jakarta, Pontianak, Surabaya and Sikka.[50] mHealth was integrated into the existing national nutrition service delivery through the *posyandu* programme; *posyandus* offer a monthly service at sub-village level. These are the lowest unit of primary health care infrastructure in Indonesia.[51]

Community health workers in the pilot study sites were provided with a mobile phone to replace a manual system. In 2013, World Vision Indonesia received technical support from

---

[48] World Vision International 2018
[49] Barnett et al. 2016
[50] World Vision International 2016
[51] World Vision International 2015

the MOTECH Suite[52] to design a mobile phone based application to address some of the challenges known to hinder nutrition service delivery in Indonesia; they targeted key issues with growth monitoring and nutrition counselling in the *posyandus*.[53] Following Barnett et al., there are many issues with manual growth monitoring in Indonesia.[54] Some challenges include the incorrect categorisation of children's weight as normal, which leads to cases where parents are not offered nutrition support services. Other challenges are around the slow retrieval of children's details and subsequent delays in offering support services. There are also challenges associated with the limited training and supervision of community health workers for manual growth monitoring status classification and uncertainty around how to calculate and interpret manual growth charts.

mHealth Indonesia serves many purposes. Growth monitoring forms[55] for each child attending the *posyandus* are completed by the community health workers and health facility workers and saved on the mobile phone application via general packet radio service.[56] The growth monitoring forms can then be accessed by health users in different agencies, such as the district health clinics.[57] The mobile application processes growth and nutrition measurements automatically and flags the level of nutritional risk. It does so by classifying a child's nutrition and growth velocities through use of a standard Z-score. In doing so, the application provides tailored nutrition messages that can be provided by community health workers to mothers in real time, or at the time of the visit to the *posyandu*.[58] This process omits the role of community health worker manually working out the growth velocities and is thought to improve (among other things) the accuracy of the data.

The mobile application also serves other purposes, such as the generation of growth trend summaries, detailing areas of extreme prevalence of undernutrition or average weights of children under five in a particular location.[59] The application also provides assessments of underlying illness and the current feeding practices of a child during home-based counselling sessions.[60]

mHealth is considered to offer an innovative solution for community health workers and nutrition stakeholders to improve nutrition outcomes. Accordingly, between 2013 and 2015, a mixed methods evaluation of mHealth was carried out by the Institute of Development Studies in partnership with World Vision Indonesia to better understand exactly how mHealth was supporting these better nutrition outcomes. The evaluation and supporting documentation[61] has been used to further unpack the context of community-based

---

[52] The MOTECH platform is an open source enterprise software package that connects popular mHealth technologies to nutrition outcomes.
[53] World Vision International 2015
[54] Barnett et al. 2016
[55] This involves the collection of anthropometric measurements.
[56] This is called 'GPRS'.
[57] World Vision International 2016
[58] World Vision International 2016
[59] World Vision 2015
[60] Word Vision International 2015
[61] See Munslow et al. 2016

monitoring in Indonesia and the current challenges, as well as how the mobile phone application was integrated into the *posyandus*.[62]

The evaluation offers an excellent insight into the conditions under which the mobile application brought about desired improvements in growth monitoring and nutrition counselling;[63] specifically, improvements in the 'accuracy, timeliness and responsiveness of growth monitoring', all of which are important for effective community-based growth monitoring.[64] The evaluation also provides a rich narrative around the context of nutrition services in Indonesia and motivations of various nutrition stakeholders.[65] Fourteen *posyandus* (in North and East Jakarta and Sikka) were selected to inform the evaluation.[66] Qualitative and quantitative data are provided in the form of electronic databases, interviews and focus groups.[67]

Among the many important findings uncovered in Barnett et al.'s evaluation, 'mobile phone improved accuracy of growth monitoring status classifications', 'the mobile phone increased the timeliness of the growth monitoring in all *posyandus*', and '[of] the mothers who received feedback, a significantly higher proportion (93 per cent) received it from [community health workers with] mobile phones'.[68] Some of the reasons for the mobile phone increasing feedback to mothers during the growth monitoring sessions were found to be context specific and related to issues such as 'trust in the feedback provided' and the 'objectivity of the calculations provided by the mobile phones'.[69] These are important findings and reflect just some of the contextual conditions under which the mobile application brought about desired improvements in growth monitoring and nutrition counselling.[70]

mHealth has the potential to help improve children's nutrition in other places where children can be weighed in community health clinics, but data is to be collected and curated and resources allocated at a higher (say, district) level. We have constructed a middle-level pToC that follows our format for such a programme, using, among other sources, information from a series of follow-up visits with health stakeholders relevant in mHealth Indonesia (e.g. World Vision implementation staff, caregivers attending the clinics, community health workers, government officials[71]). The Indonesian programme had a number of intended outcomes: children brought to the community health clinics being better nourished; more accurate childhood nutrition data being available at district and higher levels; more efficient allocation of district resources; pressure on the government to address local problems of childhood nutrition. For our illustration, we have chosen to focus

---

[62] A theory of change was constructed for Barnett et al.'s 2016 evaluation covering 14 *posyandus* in Indonesia, which has been extremely useful in reconstructing the six assumptions for a pToC detailed in this report.

[63] This programme of work and evaluation was led by Dr Inka Barnett at the Institute of Development Studies, Brighton (see Barnett et al. 2016, 18–19).

[64] Barnett et al. 2016, 13

[65] Barnett et al. 2016, 18–19

[66] Barnett et al. 2016, 32, 35

[67] Barnett et al. 2016, 23–24

[68] Barnett et al. 2016, 39

[69] Barnett et al. 2016, 41

[70] Barnett et al. 2016, 18–19

[71] This information can be found in the supplementary report in Munslow et al. 2016.

on the 'children brought to the clinics are better nourished' outcome. Similar middle-level pToCs can be constructed for the other outcomes as well. Our mHealth middle-level pToC appears in Figure 21.

**Figure 21: a representation of the mHealth middle-level pToC**
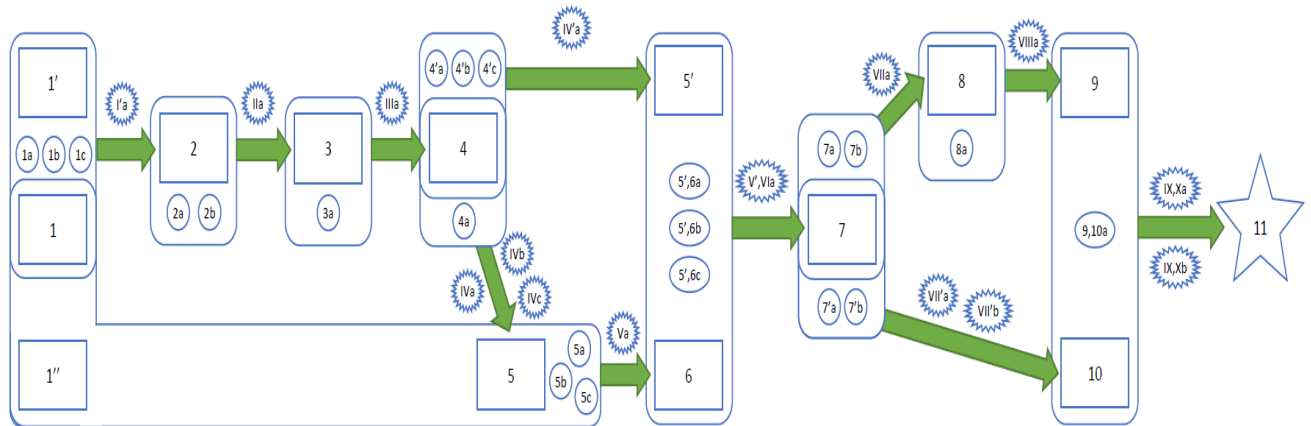


The key to Figure 21 is presented in Frame 1.

**Frame 1**

Overall programme theory (assumption type 1): mHealth is administered and used by community health workers for growth monitoring in community health clinics where weight data is stored in mobile phone technology. The technology improves the accurate classification of infant growth status and promotes a timelier response from district health centres. This improves feedback and response to mothers attending the clinics and provides them with the advice and help they need. This leads to the outcome of children attending the clinics growing better than before.

| Box 1: mHealth is administered. |
|---|

| Box 1': community health workers are mandated to use it. |
|---|

| Box 1'': district workers are mandated to monitor, curate and respond. |
|---|

Principle 1, 1'–2: health workers tend do what they can in their clients' best interest.

Support factors:

      1a      Community health workers have the capacity to use mHealth.

      1b.     Community health workers agree that using mHealth is good for their clients.

> 1c     Mothers and children attend the community health clinic on a regular basis

Derailer:

> I'a     external pressure to not perform the task or other priorities prevail

---

**Box 2: infants' weight data is recorded in phones.**

Principle 2–3: mHealth technology does accurate calculations of growth status.

Support factors:

> 2a     mHealth technology is well designed for the job.

> 2b     Community health workers input the correct data in the correct format.

Derailers:

> IIa     Technology fails to operate.

---

**Box 3: accurate classification of infants' growth status by mHealth technology.**

Principle 3–4: automated monitoring systems reduce error (e.g. there is a reduction in the misclassification of underweight infants with an automated monitoring system).

Support factor:

> 3a     community health workers understand how to interpret results recorded by mHealth technology.

Derailer:

> IIIa     external pressure to not perform the task or other priorities prevail.

---

**Box 4: accurate identification of underweight infants in the *posyandus*.**

Principle 4–5: automated methods for doing so promote data submission.

Support factor:

> 4a     community health workers are able to submit data.

Derailers:

    IVa       External pressure to not perform the task or other priorities prevail.

    IVb       Other tools dominate and/or are seen as more useful.

    IVc       Technology fails.

---

Box 5: more timely submission of data from local to district health clinics, especially in rural areas.

---

Principle 1, 1', 5–6: health workers tend do what they can in their clients' best interest.

Support factors:

    5a       District health staff know what to provide.

    5b       District health staff agree to the importance to clients of so doing.

    5c       District health staff are able to provide help.

Derailers:

    Va       External pressure to not perform the task or other priorities prevail.

---

Box 6: district level help (referrals etc) provided to mothers where indicated by data.

---

Principle 4–5': health workers tend do what they can in their clients' best interest.

Support factors:

    4'a      District health staff understand the results.

    4'b      District health staff are able to provide feedback and advice.

    4'c      District health staff agree on the importance of the feedback and advice for the clients.

    Derailers:

    IV'a     external pressure not to do so or other priorities/goals

Box 5': timely feedback and advice to mothers in clinics.

Principle 5', 6–7: people take in advice clearly given by people in positions of trust.

Support factors:

5', 6a        Advice and help are clearly presented.

5', 6b        Mothers can understand the advice.

5', 6c        Mothers trust community and district health institutions.

Derailers:

V', Via, ?        Here we put a question mark as a reminder that though none are identified, some may exist.


Box 7: mothers have knowledge and help they think can make their children better nourished.

Principle 7–8: mothers do what they think is in their children's best interest.

Support factors:

7a        Mothers think this is in their children's best interest overall.

7b        Mothers can follow advice.

Derailers:

VIIa        conflicting norms or priorities


Principle 7–10: mothers do what they think is in their children's best interest.

Support factors:

7'a        Mothers think this is in children's best interest overall.

7'b        Mothers and children can continue attending.

Derailers:

VII'a        conflicting norms or priorities

VII'b    physical or economic preventions

---

Box 8: mothers follow the advice.

---

Principle 8–9: children who are better nourished begin to grow more.

Support factor:

8a    The advice and help are a good way to better nourish those children.

Derailer:

VIIIa    Other factors (such as illness) impede growth.

---

Box 9: undernourished children from clinic begin to grow faster.

---

Box 10: mothers and children continue attending clinic and getting advice and help.

---

Principle 9, 10–11: children who are better nourished grow faster.

Support factor:

9, 10a    The advice and help are a good way to better nourish those children across time.

Derailers:

IX, Xa    Other factors (such as illness) impede growth.

IX, Xb    Failure at local or district level to supply advice or help.

---

Box 11: children attending clinic grow faster than before.

---

Let us now show how this could have been built  up one stage at a time to give you a sense of what is involved in building a pToC like this. It is important to realise that this will not really involve a linear progression of thought. Middle-level pToCs are often a good example

of grounded theory, where the theory emerges as it is developed. Each stage in the development informs the others. Work on one kind of assumption may provoke a rethinking of some assumptions made earlier, or as different assumptions are added, it may become clear that the initial sequence of intermediate stages needs to be modified. What matters is that at the finish, the pToC as a whole is coherent and credible. The simple input–output ToC looks like Figure 22.

**Figure 22: a simple input–output ToC representation**
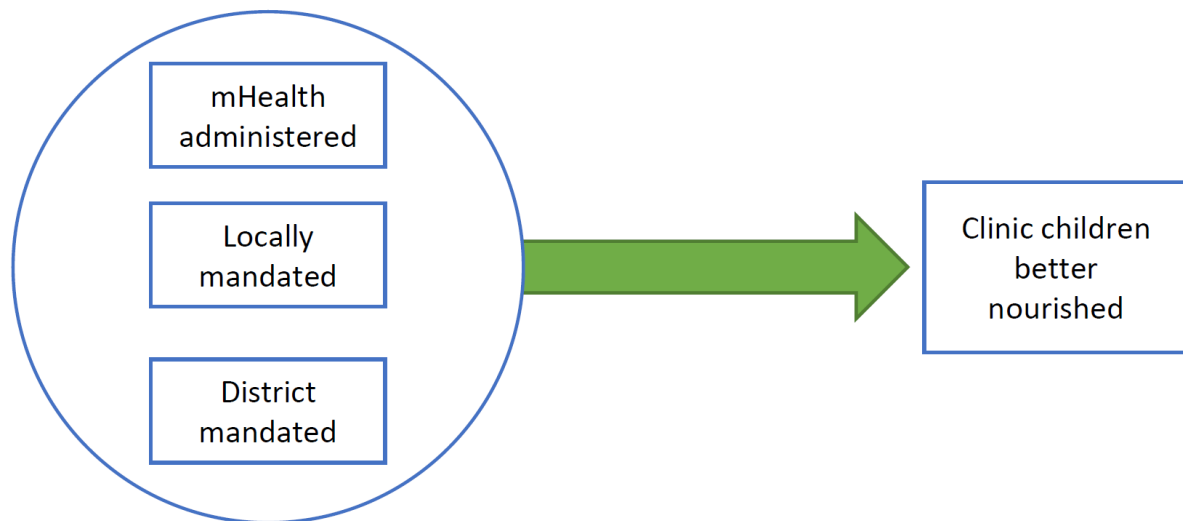


Next, we introduce the significant causal steps in between, as in Figure 23. Doing the necessary thinking and research to work out what the process is supposed to be and how the programme input initiates it is large part of the work of designing a programme that can be expected to carry through from start to finish.

**Figure 23: representation of significant causal steps in between**
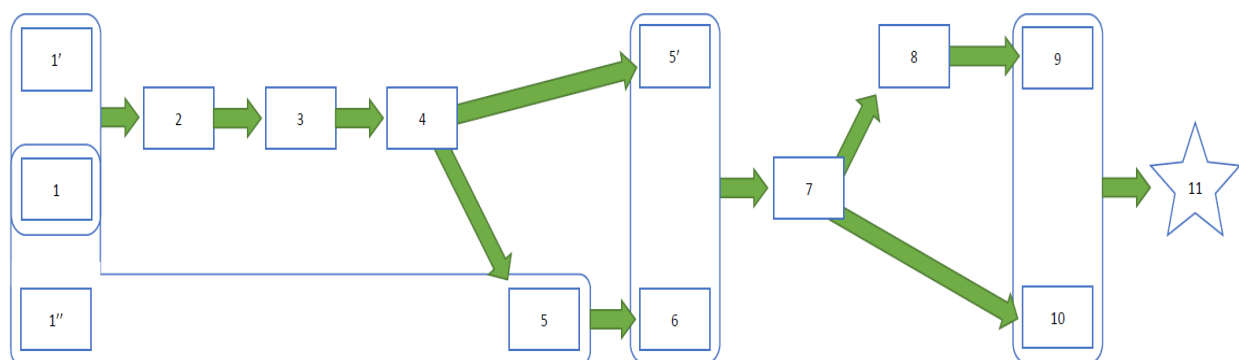


Next, all the support factors, derailers and safeguards must be added. Consider steps 2 to 3: 'Infants' weight data is recorded in phones' leads to 'Accurate classification of infants' growth status by mHealth technology'. This step will tend to work because of the principle that the mHealth technology does accurate calculations of growth status, as represented in Figure 24.
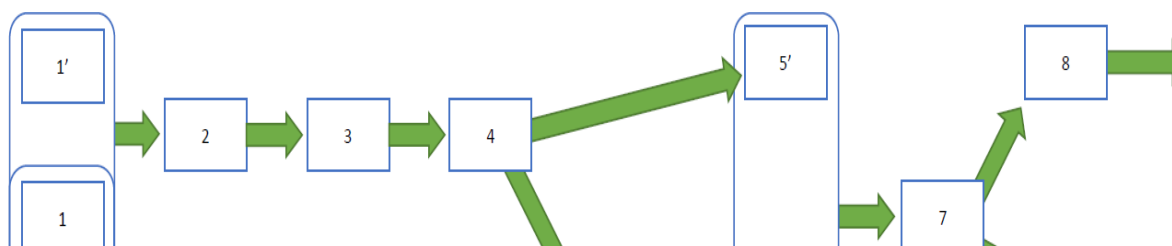
**Figure 24: adding in together the support factors, derailers and safeguards for this step**
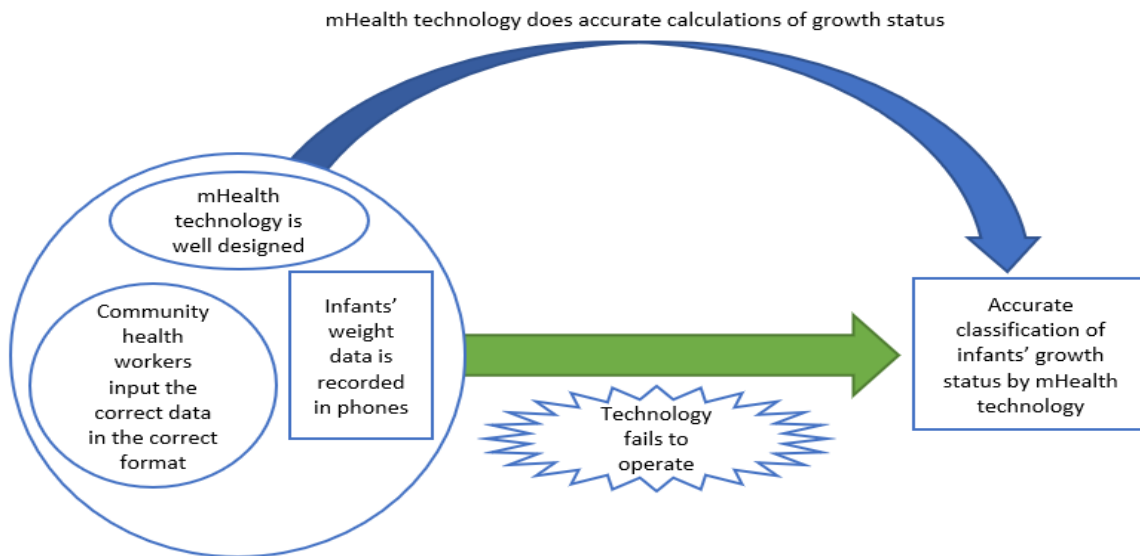


But we also see that in order to get from this step to the next, the mHealth technology must be well designed for the job, and community health workers must input the correct data in the correct format. The whole process will also be derailed if the technology fails to operate. For legibility, in the full middle-level pToC diagram for mHealth (Figure 21), all these descriptions have been abbreviated to their label as assigned in the key. Furthermore, principle arrows have been omitted, though these are detailed in the key. It is critical that principles are discovered and recorded, even if it may not be possible to include the arrows representing them in a legible diagram.

Building up the process in this way enables to see how many things can go wrong or prevent the programme from achieving the intended outcome. In the middle-level pToC for mHealth, there are more than 21 factors to consider if the programme is to work as intended. As always, decisions about how serious these problems might prove to be, whether they can be worked around, what the costs and benefits are and for whom of trying and failing (or trying something else or nothing at all) requires good sense and good judgement and can seldom be highly certain.

## 5.3 Case study 2: anti-corruption

This middle-level anti-corruption pToC is built from considerations, primarily by PERL (the programme that works with government and civil society to improve the delivery of public services), about how corruption can be or has been reduced in Nigeria. There are many high-level theories available for practitioners to interpret the Nigerian public sector. The more academic theory is captured by the concept of neopatrimonialism (where a distinction between public or state wealth and the private wealth of the ruler becomes blurred). The more practical theory is captured by the concept of good governance (where governance must become less corrupt before the state can intervene in the economy). However, high-level theories threaten to overlook the more fine-grained context-specific and time-specific facts on the ground that frame how particular governments can and should perform. So, the

ideal solutions that high-level theories produce tend to become suboptimal solutions and even harmful under real-world conditions.

In response, Merilee Grindle argues practitioners should avoid one-size-fits-all theories and idealised end states, and aim instead for more contextualised theories.[72] In a similar spirit, Future State and Sue Unsworth advise practitioners to set aside their own ideals or preconceptions.[73] Rather, practitioners should start from the country's reality on the ground and aim for more local and incremental institutional reform. In particular, researchers from the Overseas Development Institute's African Power and Politics Programme (APPP) recommends that the international community assesses institutional reform in the light of more rigorously researched causal linkages between political institutions and development outcomes.[74] So, practitioners need a more fine-grained conceptual framework that prioritises second-best or hybrid solutions that work with the grain of local communities.[75] However, it is unrealistic to expect practitioners to conduct a deep contextual analysis of every new context. Consequently, practitioners need access to middle-level theories, especially middle-level overall programme pToCs, drawing on practical principles that are supported by a body of theory and that are empirically robust across a range of contexts.

Within the Nigerian public sector, PERL has identified what Roll calls 'pockets of effectiveness'.[76] PERL uses a 'positive deviance'methodology, which identifies anomalous islands of success within a dysfunctional system.[77] The pockets of effectiveness were not from international donor interventions but emerged from within the system. The good governance reforms promoted by international actors often misunderstand the realities of how power is ordered in a given local context, and they also neglects the political barriers that derail policy implementations.[78] Competition for oil rents dominates the political economy of Nigeria, and the (mis)allocation of oil rents often worsens the ethnic and political divisions within Nigerian society.[79] So successful public sector reform is not purely a technical activity but demands political skills and organisational knowledge. In particular, these tend to emerge from the interplay between political, organisational and societal factors.

APPP argues that the two competing frameworks – the supply-side approach and the demand-side approach – both share a common principal-agent framework. The implicit but central assumption of the principal-agent framework is that individual actors – whether public servants or private citizens – possess a strong and simple motivation to make public services less corrupt and more effective. According to APPP, the supply-side approach assumes individual public servants have a commitment to provide effective public services. With this high-level conception of individual public servants in mind, the supply-side approach questions how to enable well-motivated public servants to overcome external

---

[72] Grindle 2011

[73] Future State 2010; Unsworth 2009

[74] Booth 2012; Andrews 2008; Khan 2007

[75] Crook and Booth 2011

[76] Roll 2014

[77] Abah 2012; 2017; Roll 2014; Bureau of Public Service Reforms 2015; de Gramont 2015

[78] Porter and Watts 2016, 260

[79] Porter and Watts 2016, 252

structural barriers within institutions. The preferred instruments include budget support, technical assistance, policy monitoring and dialogue.

However, APPP criticises the supply-side approach as managerialist and insufficiently sensitive to the political factors at play within poor public sector performance. In response, the demand-side approach aims to counterbalance the weaknesses in the supply-side approach by politically empowering citizens and politically mobilising civil society to hold governments to account. Nevertheless, the demand-side approach repeats the same mistake of the supply-side approach by assuming that individual private citizens have a simple desire and capability to hold governments to account. Either way, the principal-agent approach often produces generic analysis that risks the false hope of silver bullet solutions. Worse, its high-level conception of individual actors black boxes complex political decisions which are often framed by context-specific and time-specific facts on the ground.

In response, a 'collective action' approach provides a middle-level general programme theory, as in our assumption type 1, which allows practitioners to open up the black box of complex political decisions and provide more localised and incremental solutions. The collective action approach views effective public services as public goods that create free-rider problems. The basic public good or free-rider problem is that everyone would benefit from less corrupt and more effective public services, but everyone would prefer that someone else, somewhere else did the intensive work necessary to implement successful reform. The collective action approach stresses the significance of arrangements that assist with local problem-solving, instead of the single-stranded solution or silver bullets that have undue influence over development practice.

Good institutions solve problems arising in specific circumstances, meaning that generic remedies will often miss the point and may well do harm. The collective action approach allows for a more sophisticated interpretation that emphasises the overcoming of problems of coordination, credibility and collective action among sets of actors with complex interlocking interests. APPP suggests that good development institutions enable more efficient and less corrupt provision of public services by safeguarding against locally specific collective action problems. This promotes second-best or hybrid solutions that reduce the costs of institutional innovation by working with the grain of local practices and norms.

In the light of a collective action framework, one small strand we have decided to pick out is the significance of dedicated leadership. In particular, integrity is often powerful enough to overcome free-rider problems. Individual leaders are in a position of power to implement successful reform, and dedicated leaders are internally motivated enough to do the necessary work. They do not wait to free-ride on someone else, but take it upon themselves to attain effective reform. Once given the opportunity, dedicated leaders are able to change mindsets and behaviours within the institution, spilling over into changing the mindset and behaviour of the wider public, to attain successful reform. The most common indicator of effective pockets is active political pressure. Competition for oil rents often derails cooperation among political elites.[80] Moreover, when the political elites do respond to public demand, they tend to prefer providing discretionary handouts instead of public

---

[80] Lewis and Watts 2015a

goods.[81] As a result, public sector reform is often enabled through dedicated political leaders with long-standing records of integrity. Reform-minded leaders are hard to identify in advance, so it is better to identify indirect signals of integrity instead.

Two of the clearest cases of dedicated political leadership are President Obasanjo and Edo State Governor Oshiomhole.[82] President Obasanjo co-founded the anti-corruption agency Transparency International, and the first bill he introduced as president was on anti-corruption. Under his presidency, anti-corruption agencies became much more effective. In particular, the anti-corruption agency, the Economic and Financial Crimes Commission (EFCC) visibly arrested high-profile criminals thought untouchable, the Federal Inland Revenue Service (FIRS) reduced the under-reporting of revenue and the National Agency for Food and Drug Administration and Control (NAFDAC) reduced the number of counterfeit drugs.[83]

The EFCC was ranked the most effective anti-corruption agency in Africa. It was created as a response to the Financial Action Task Force adding Nigeria to the list of non-cooperative countries or territories for not combating money laundering. Nigeria was removed from the list in 2013. Second, Governor Oshiomhole was more extrinsically motivated to achieve effective reform. It took a long legal dispute for him to win his election victory. After this his local legitimacy rested in part on him getting better results than his predecessor.[84] To do so, Oshiomhole strategically targeted road construction for quick, visible and attributable results to increase public support before he addressed less tangible reforms such as creating job-friendly economic growth.

Abah and Roll identify that successful organisations commonly head hunt leaders in light of their proven integrity over technical expertise, and integrity tends to be prioritised throughout the pockets of effectiveness.[85] Akunyili, leading the NAFDAC, is Nigeria's public sector executive with most awards. With the support of President Obasanjo, Akunyili was determined to reduce counterfeit drugs after her sister died from counterfeit insulin.[86] Akunyili carefully recruited on the basis of integrity and won NAFDAC autonomy from the Ministry of Health. From 2001 till 2006 the sale of counterfeit drugs dropped from 41 per cent to 16.7 per cent.

Subsequently, Ndaguba, leading the NAPTIP (National Agency for the Prohibition of Traffic in Persons), handpicked her staff with integrity as a priority, and they remained intrinsically motivated to combat human trafficking by meeting the victims.[87] Ndaguba lobbied for legal amendments to give the NAPTIP powers to prosecute traffickers in-house, which enabled it to avoid delays, coordination problems with external agencies and political interference.[88] The NAPTIP secured 57 convictions for human trafficking in 2003–9, more than half of all convictions in Africa. Finally, Oshiomhole benefited from internal disputes within the

---

[81] Keefer and Khemani 2003
[82] Abah 2012
[83] Abah 2012, 254
[84] Porter and Watts 2016
[85] Abah 2012; Roll 2015
[86] Roll 2015
[87] Simbine et al. 2015
[88] Lewis and Watts 2015b; Simbine et al. 2015, 140

opposition party, the People's Democratic Party (PDP). With no coherent opposition, he could implement his reform agenda unimpeded.[89]

Increased autonomy was often necessary to enable dedicated leaders to act effectively. Rogger found that decentralised organisations had a 40 per cent higher rate of project completion than centralised ministries because of their higher degree of autonomy.[90] The organisations were free from political interference and from delays caused by a need to collaborate with outside organisations. This higher degree of institutional autonomy enabled dedicated leaders to implement targeted reform efforts, and effective political management safeguarded the more autonomous organisations from the risk of corruption. Moreover, the organisational leaders often gave more autonomy to lower-level bureaucrats. This often went with the grain of their intrinsic motivation to perform well, despite increasing the opportunities for corruption.

Rasul and Rogger, after comparing the completion rates of 4,700 projects across the Nigerian civil service, identified projects where managers were willing to delegate decisions to more junior civil servants and give them more autonomy that correlatively had higher rates of project completion.[91]

The centralised Ministry of Budget and National Planning often left out decentralised bureaucrats from reform discussions and as a result they resisted implementation. In contrast, The Public Financial Management (PFM) reforms in Nigeria included other bureaucrats as agents of change, instead of targets, to build trust. The PFM amended policy collaboratively with these bureaucrats. They then delivered the earliest executive budget to the legislature in ten years.[92] In Ogun State, tight deadlines motivated bureaucrats to respond rapidly, and secured their continued buy-in to reform efforts.[93] The introduction of impersonal and predictable technology, backed with high-level political support for effective sanctions, safeguarded reform efforts from corruption and enabled sustainability.[94]

The World Bank (2003) blames the lack of social accountability for poor government performance. However, this demand-side approach overlooks the complexity of the local collective action problems that citizens face, and there is little evidence that civil society enables effective reform. On the other hand, coalitions often form once citizens see early signals of reform to safeguard it from external corruption.[95] With systemic failure in mind, citizens tend to become motivated to value reform only once they see tangible signs that reform efforts are meaningful. In particular, the media can identify the tangible signs of reform that enable public support. Public support enables an organisation to gain local legitimacy, which safeguards that organisation from political interference and enables it to get sustained funding.

---

[89] Porter and Watts 2016, 258
[90] Rogger 2014
[91] Rasul and Rogger 2017
[92] Andrews 2018
[93] Bankole et al. 2018
[94] Abah 2017
[95] Lewis and Watts 2015a, 20

First, Abah details how the Bureau of Public Service Reforms (BPSR) used citizen voices through surveys, key informant interviews, pictures with geotagging, focus group discussions and mystery shopping (where bureau staff disguise themselves as members of the public trying to obtain public services) to increase demand for change within the Nigerian public sector.[96] Second, NAFDAC gained more political support than the Standards Organisation of Nigeria since there was more public outrage over counterfeit drugs than over counterfeit consumer goods.[97] Third, once the media communicated that the EFCC had made high-profile arrests of untouchable criminals, public support rose; this protected the EFCC from political attempts to derail or hinder its work, and secured its funding. These success stories allow us to start constructing a middle-level theory of how dedicated leadership can contribute towards successful public sector reform.

From PERL's recent research on Nigerian public sector performance, we extracted specific claims that we judged as significant for a middle-level programme pToC. With these building blocks to hand, we constructed an initially plausible pToC for such a programme (see Figure 25, the key for which is in Frame 2). However, the central purpose of our ToC below is merely to show how practitioners can start to develop middle-level pToCs of their own. Our pToC is not intended as a finished article. Our steps and their support factors and derailers are primarily based upon our own desk-based theorising from a distance in the light of PERL's current research.

It is important to note that the pToC in Figure 25 explores only one pathway to successful public sector reform out of many explored by PERL. It is thus not a general theory for overall public sector reform but rather a general theory for one strand of public sector reform, centred around the significance of dedicated political leadership over and above well credentialed leadership. As already explained, PERL explores the anomalous real-world cases of good governance on the ground. One commonality among the anomalies is dedicated leadership. Here we construct a sample pToC for how dedicated leadership can spearhead public sector reform.

**Frame 2**

Overall programme theory (assumption type 1): the dedication of leadership is to make reform efforts resilient to the complex derailers that it needs to overcome. This is to lead to institutional reform that is visible to the public, which in turn is to lead to the outcome of widespread public support and to the embedding and stabilising of the reform within the wider society.

Box 1: leaders and bureaucrats with credentials signalling integrity are hired (over those signalling expertise).

Principle 1–2: if leaders and bureaucrats with credentials signalling integrity are hired when the credentials actually signal integrity and there are no high payoffs for false integrity signals or a lapse in character, then leaders and bureaucrats with integrity enter into the positions of political power.

---

[96] Abah 2017
[97] Abah 2012

Support factor:

1a    The credentials signal integrity.

Derailers:

Ia    high payoffs for false integrity signals

Ib    lapses in character between signalling and being hired

---

Box 2: leaders and bureaucrats with integrity in positions of political power.

Principle 2–3: if leaders and bureaucrats with integrity are in positions of political power with firm characters and no overpowering or subtle corruption, then the leaders and bureaucrats are resistant to external corruption.

Support factor:

2a    Character remains firm in the positions of power.

Derailers:

IIa    distributive competition for rents

IIb    overpowering corruption

IIc    subtle corruption

---

Box 3: leaders and bureaucrats resistant to external corruption.

Principle 3–4: if the leaders and bureaucrats are resistant to external corruption, and decentralised bureaucrats recognise leaders' resistance and desire to change and are included in reform discussions by centralised organisations, then decentralised bureaucrats believe in change.

Support factors:

3a    Decentralised bureaucrats recognise the leaders' resistance.

3b    Decentralised bureaucrats desire change.

Derailer:

IIIa    decentralised bureaucrats excluded from reform discussions by centralised organisations

---

Box 4: Decentralised bureaucrats believe in change

---

Principle 4–5: if decentralised bureaucrats believe in change and the reform-minded leaders motivate them; if the apathetic bureaucrats go with what they believe is the flow of the majority and they are equipped to work towards reform and there are no excessive barriers; then decentralised bureaucrats work towards reform.

Support factors:

4a    Reform-minded leaders motivate bureaucrats with tight deadlines.

4b    Apathetic bureaucrats go with the flow.

4c    Apathetic bureaucrats believe the majority believe in change.

4d    Decentralised bureaucrats are in positions to work towards reform.

4e    Apathetic bureaucrats have knowledge and skills to work towards reform.

Derailers:

IVa    excessive penalties for working towards reform

IVb    other priorities dominate

---

Box 5: decentralised bureaucrats work towards reform.

---

Principle 5–6: if decentralised bureaucrats work towards reform, disempowered bureaucrats believe more are working towards reform and there are no big threats from outside, then disempowered bureaucrats work towards reform.

Support factor:

5a    Disempowered bureaucrats believe more are working towards reform.

Derailer:

Va    big threats from outside

| Box 6: disempowered bureaucrats work towards reform. |
|---|

Principle 6–7: if most bureaucrats work towards reform, that majority has sufficient power, they interact with each other and there are no excessive obstacles, then a pro-reform organisation forms.

Support factors:

6a      Majority has sufficient power.

6b      Majority interacts with each other.

Derailer:

VIa     strong structural, physical, financial obstacles to the existence of such an organisation

| Box 7: a pro-reform organisation |
|---|

Principle 7–8: if there is a pro-reform organisation and inter-organisational interaction, then a pro-reform culture is formed.

Support factor:

7a.     inter-organisational interaction

| Box 8: a pro-reform culture |
|---|

Principle 8–9: if a pro-reform culture forms, the organisation gains the legal power to act, political approval and approval from a silent public, then the organisation acts.

Support factors:

8a      legal power to act

8b      political approval

8c      silent public approval

| Box 9: organisation acts. |
|---|

Principle 9–10: if the organisation acts and gets early, big, visible and attributable successes that signal legal credibility to citizens, then citizen coalitions form.

Support factors:

9a      early, big, visible and attributable successes

9b      signals of legal credibility to citizens

| Box 10: citizen coalition |
|---|

Principle 10–11: if citizen coalitions form as embodiments of public opinion with a re-energised citizenry and organisations can get quick and easy access to public opinion, then organisations respond to citizen demand.

Support factors:

10a     Public opinion embodied.

10b     Citizens (re-)energised.

10c     Quick and easy access to public opinion.

| Box 11: organisation responds to citizen demand. |
|---|

Principle 11–12: if the organisation responds to citizen demand, communicates with the local community and serves community needs, then the organisation gains local legitimacy.

Support factors:

11a     communication with the local community

11b     serving community needs

| Box 12: locally legitimate organisation |
|---|

Principle 12–13: if the organisation is locally legitimate through a sustained reform mindset, established feedback loops and an established harmony with the local community, power remains decentralised and local community voices remain heard, then the reform is socially stable.
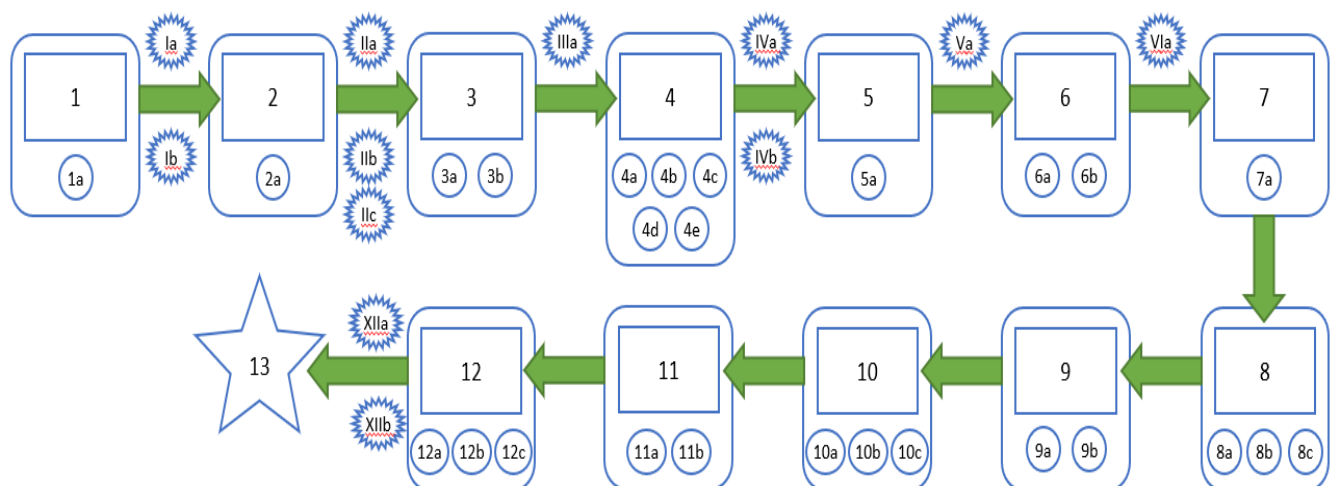
Support factors:

12a    Remaining reform-minded.

12b    Established feedback loops.

12c    Harmony with the local community.

Derailers:

XIIa    Power becomes centralised.

XIIb    Political elites crowd out local community voices.

<div style="border:1px solid black;padding:10px">
Box 13: socially stable reform
</div>

**Figure 25: one possible middle-level pToC for PERL**



## 5.4 Case study 3: conditional cash transfers aiming to increase school enrolment

The use of CCTs to support essential household consumption and to increase investment in health and/or education is a very well-studied development intervention. Since the implementation of PROGRESSA by the Mexican government in 1997, similar CCT programmes have been adopted by low- and middle-income country governments around the world. A 2015 systematic review by Snilstveit et al. identified 50 experimental or quasi-experimental impact evaluations of 38 distinct programmes from all over the global south. While CCTs originate in Latin America, Snilstveit et al. identify seven programmes in sub-Saharan Africa, six in the East Asia and Pacific region, and five in South Asia.

As well as being elaborated and tested in very numerous publications, the general assumption type 1 theory that justifies and underpins CCT programmes is fairly simple and is

homogenous between authors. This makes CCT programmes a good case study for the construction of a middle-level pToC. The theory synthesised and represented through the pToC below was drawn mainly from six narrative and systematic reviews of CCT programmes aiming in part or whole to increase levels of school enrolment of children.[98] As mentioned above, CCTs often target outcomes in health and education. In order to render the pToC developed here comprehensible, we focus only on the components of a CCT aiming to increase school enrolment through payments conditional on school attendance.

The central theoretical insight that motivates cash transfers for school enrolment is that households may be liquidity constrained in their investments in children's education by the absence or poor functioning of credit markets (Fiszbein and Schady 2009). They may also face financial barriers specifically to education in the form of direct costs such as school fees, uniforms, transport etc. and/or indirect costs in the form of returns from activities that are constrained by enrolment in school, such as child labour (Edmonds 2008; Snilstveit et al. 2015).[99]

In the presence of financial barriers to enrolment, settings will tend to be characterised by a level of enrolment that is below what the literature refers to as the household's 'privately optimal level' – the level of enrolment that the household would choose had that choice not been constrained. However, there is a further distinction to be made between the 'true' privately optimal level of investment in education (and therefore school enrolment), and the level that is perceived to be optimal by households.[100] The 'true' privately optimal level of investment in education for a given agent, i.e. an individual or a household, is the level of investment that maximises rationally expected returns – expected benefits from higher income compared to costs. There will be a difference between this level and the level that is perceived to be optimal by the agent if the agent suffers from failures of rationality or holds erroneously low beliefs about the returns from education. It will also diverge if the agent is optimising on dimensions other than future returns, e.g. in the presence of familial or community norms about what is an appropriate level of education for whom. It is this perceived privately optimal level of enrolment that is important for household decision-making, rather than the 'true' level.

As well as the privately optimal level of enrolment and the 'true' optimal level of enrolment, states often target a socially optimal level of enrolment, often 100 per cent (Das et al. 2005). It is the existence of a difference between the privately optimal level of enrolment and the 'true' or socially optimal levels of enrolment that motivates the use of CCTs. By making payments to households conditional on children's attendance at school above a certain threshold, normally 80 per cent, CCTs increase the cost of not sending children to school, providing an incentive over and above the household's perception of the returns from education. The strength of this incentive varies with the extent to which households understand programme conditions and expect them to be enforced.

---

[98] Baird et al. 2013; Das et al. 2005; Fiszbein and Schady 2009; Garcia and Saavedra 2013; Grosh et al. 2008; Snilstveit et al. 2015

[99] It is not correct to say that child labour is ruled out by school attendance, let alone enrolment. This is especially true in a context such as Brazil, where children attend school in four-hour shifts and often work around them. However, schooling at the very least constrains this activity.

[100] Fiszbein and Schady 2009 is the most complete treatment.

It has been compellingly argued by Gaarder (2012) and Baird et al. (2013) that cash transfer programmes exist on a continuum of conditionality from unconditional cash transfers, with no explicit conditions regarding health or education behaviours, to CCTs with conditions that are known to recipients, and well monitored and enforced. Between these two extremes are labelled transfers that are associated to some extent with use for a particular purpose, either through their name, the location of disbursement or even an explicit set of directions for use or a stated conditionality, but without any enforced conditionality. Such transfers may induce an increase in enrolment above a purely unconditional cash transfer in settings where an increase in the importance of education is sufficient to induce behaviour change. Benhassine et al. (2015) have tested the strength of the labelling effect in one setting, finding almost no difference on that occasion between a labelled transfer with no stated conditionality and an enforced conditional transfer in that setting.

Cash transfers conditional on school attendance, then, tend to cause increases in school enrolment through three distinct causal channels. First, transfers tend to have a direct income effect, relieving liquidity constraints and allowing households to invest more in education to the extent that this is the household's priority. Second, the conditionality associated with transfers tends to increase the cost of not enrolling children, creating a substitution or price effect that motivates households to educate children even to a level above the household's privately optimal level. Third, the labelling of transfers as intended for education will tend to induce a nudge effect, increasing the importance of education to households in some settings and making them more likely to enrol their children in school.

**Frame 3**

Overall programme theory (assumption type 1): Households tend to spend in their children's best interests, but financial barriers (insufficient resources to meet the direct and indirect costs of education) and non-financial barriers (such as incorrect beliefs about the returns from education, excessive future discounting and intra-household bargaining problems) can cause them to under-invest in schooling. Therefore, alleviating these barriers by providing resources, nudges and incentives to enrol children in school tends to lead to higher levels of enrolment.

| Box 1: resources transferred to poor households. |
| --- |

Principle 1–2: increased resources allow households to spend more on their priorities and enrolling unenrolled children in school tends to be a high priority for poor families, especially at the start of a new academic year.

Principle 1–2: households tend to make education spending a higher priority when they perceive education to be high value in general, when they perceive local provision to be of high quality, and when they perceive the child in question to have higher ability.

Principle 1–2: programmes tend to help most people with the available resources when errors of inclusion (unnecessary transfers) are balanced against errors of exclusion (transfers not made to recipients who would benefit from them).

Support factors:

1a    Transfers are low cost to recipients, e.g. they are easy to collect and to prove eligibility for.

1b    Transfers are large enough to overcome costs of schooling.

1c    Transfers are well timed in the academic year.

1d    Children in the household are perceived by their parents to have higher academic ability.

1e    The household considers education in general to be highly valuable.

1f    The household perceives the local school to be high quality.

1g    Transfers are well targeted to poor households.

1h    Households must have some non-enrolled children in order to decide to enrol more children in school.[101]

Derailers:

Ia    Excessive future discounting may undermine households' desire to invest now for future returns.

Ib    Other spending priorities may exceed children's schooling.

Ic    Temptation of spending may divert resources.

Id    The direct costs of education, such as fees, uniforms and transport, and the indirect costs of education, such as forgone child earnings, may prevent households from enrolling children in school, even in the presence of transfers.

Safeguard:

αa    Transfers may be made to mothers, who may be more likely to prioritise children's interests and less likely to indulge in temptation spending.

Box 1': transfers labelled as education related.

---

[101] This may seem too obvious to mention, but for many households recipient of CCTs in many locations, this is not the case. It is important to remember that a CCT is a mass treatment and effects are concentrated in what may be a small minority of households, depending on targeting strategy. This is particularly important to remember in the context of the statistical detection of effects in large populations.

Principle 1 + 1'–2: labelling a transfer tends to increase the importance of the indicated good and increase spending on that good, even in the absence of a requirement to spend the transfer on the good.

Support factors:

1'a       the location in which the transfer takes place being related to education

1'b       the name of the transfer being associated with investing in children's education

Box 1'': conditionality of transfers on school attendance communicated.

Principle 1 + 1''–2: making continued transfers conditional on some requirement (in this case, attending school above a certain threshold) increases the expected cost of not meeting that requirement, inducing a substitution or price effect, which tends to increase the desirability of meeting the requirement.

Support factors:

1''a      Conditions are announced.

1''b      There is a credible threat of enforcement of the conditions.

Derailer:

I''a      Conditionality not understood.

Box 2: households decide to enrol more children and deliver them for enrolment.

Principle 2–3: when a school is presented with a child for enrolment, that child tends to be enrolled.

Support factors:

2a       School able to enrol a new student.

2b       Place available or the possibility of creating one for a new student to be enrolled.

Box 3: more children are enrolled in school.

Box 3': background level of enrolment.

Principle 3, 3'–4: some enrolled children tend to drop out of school.

Box 4: some children cease to attend school.

Principle 4–5: programme administrators tend to enforce programme conditions when they are informed of non-compliance.

Support factors:

4a    Non-attendees identified by the school in order to inform the programme administrators.

4b    Programme administrators informed of the identities of non-attending students to enforce programme conditions.

Derailers:

IVa    Programme administrators feel compassion for or solidarity with households with non-attending children and are unwilling to enforce program conditions.

IVb    Long time lags in low- and middle-income country administrative systems (which may mean transfers to identified non-compliers continue for as long as a year after dropout, diminishing the incentives to comply with conditions).

Safeguard:

δa    A highly formalised system with oversight and accountability mechanisms will tend to limit programme administrators' ability to exercise compassion or enact solidarity.

Box 5: transfers to households with dropouts stopped.

Principle 5–1''b: awareness of the enforcement of programme conditions on others increases the credibility of the threat to enforce program conditions on oneself.

Support factors:

5a    Visibility to other recipients of the cessation of transfers.

5b    The reason for the cessation of transfers is understood to be the enforcement of programme conditions.

The middle-level pToC for CCTs for child school enrolment described above is represented in Figure 16.

This diagram helps the reader to understand several features of the model of programme causation laid out here. First, box 1, along with its support factors, causes box 2 independently of boxes 1' and 1''. This reflects the fact that unconditional, unlabelled transfers can be effective. The conditionality and labelling attached to CCTs are additive features that may help the transfer of resources to be more effective in the right sorts of settings. Therefore, the conjunction of box 1, box 1' and their respective support factors

also causes box 2. The same is true of the conjunction of box 1, box 1'' and their respective support factors.

Second, there is a causal loop in the model. More children being enrolled will lead to some of those children dropping out, in addition to the background level of enrolment also leading to some children dropping out. This in turn tends to cause transfers to households with dropouts being stopped, which tends to reinforce the credibility of the threat of enforcement of programme conditionality.

We can also see the various derailers present in the middle-level pToC, which causal relations they threaten, and what safeguards might be implemented to inhibit their action. Both the diagram and the written model are necessary to clearly capture the patterns of causation implied by the pToC and the causal tendency principles that underpin the model.

## *6 In sum*

Predictions that a development programme will make the expected contribution in a given local setting are always highly uncertain and it is difficult to provide good justifications for them. Here we have offered a systematic account of the type of information that can make these predictions more reliable, and a framework that can be used to think about this information and express it. This framework is also useful for *post hoc* evaluation of whether a programme contributed to the outcomes as expected.

There are two categories of middle-level theory that can help in predicting a programme's effectiveness in a target setting. First: pToCs, both general and locally thickened. Second: the middle-level principles that justify each step in such stepwise theories of change. In this paper we have specified the features that general programme theories must have in order to be useful for improving the reliability of and rationale for predictions of effectiveness. We have shown how they can be built and how they can be thickened to become locally specific. We have also provided three case studies that illustrate the explanatory and predictive power of this approach.

The account presented here of how to create an adequate pToC might seem very demanding. However, designing effective programmes in complicated social settings is challenging. In order for theories of these programmes to have predictive power, those theories must be very well thought through. We hope that with practice, thinking in the way that we suggest can become second nature. If such thinking were widespread, the quality of design and evaluation of social programmes would be enormously improved.

On many occasions, for a variety of reasons you may not wish to invest the time, energy and resources in constructing these pToCs, or you may not be able too. Nevertheless, the things represented in the pToC are just what must transpire *in loco* if the programme is to succeed. This is why having a well-developed pToC for the programme provides solid grounds for predictions about its success in local settings, be they positive or negative predictions. For this reason, it is certainly a good idea to know about them. More generally it is important for the representation to be as good as possible in order to better understand how to implement the predictions.

***Appendix: a primer on middle-level principles***

In this appendix we attempt to provide a better background understanding of the middle-level principles we have in mind, what can be done with them and what their limitations are. Many of the middle-level principles that can be useful in designing development interventions and in evaluating their chances of success in targeted settings are research based.

In general, middle-level principles of use in development planning do not tell us what must happen or even what does happen. Rather, they are what John Stuart Mill called 'tendency' principles,[102] which indicate what effect a feature *tends* to cause. What a feature tends to cause is what it will cause if all the right factors to help it do so are in order, and no other causes of the same effect are in play. The word 'tendency' is meant to imply a kind of regularity; Mill used the word to stress that what the cause produces 'on its own' is generally what it 'tries' to produce when other causes act as well. It 'pushes' towards the same effect even if the other causes dilute or enhance this effect. Therefore there is no regularity between the operation of the tendency and what outcomes actually occur, but only between the operation of the tendency and a push in the direction of the effect we associate with it. We all know cases where an intervention has been highly effective but still outcomes got worse because so many negative factors were building up at the same time.

To mark that these are tendency principles and that we cannot expect their effect to be what actually appears when they operate, the words 'may', 'can' or 'tends to' are often used. Here is an example from the 2019 programme level report on the PERL project in Nigeria in which 'will' is used, though 'tends to' or 'pushes towards' (or 'may', which the report uses elsewhere) would have been better:

> PERL: Working on the connections between different levels of government (federal, state and local) will enhance outcomes. (p. 3)

Often the tendency principles of use in social planning describe the operation of psychological or social dispositions that are widespread in individuals or institutions, or widespread in specific settings or cultures. Sometimes they are dispositions particular to local institutional structures. In either case, where present, they still may need triggering. As we have discussed, they will also need the requisite support factors to be in place and derailers to be guarded against if they are to act as we expect

There is generally some overall principle that suggests that the programme can produce the targeted outcome. This is the first assumption on our list of assumptions that are needed to lay out a good middle-level programme pToC. But beyond that, each step in the pToC is expected to produce the next, and that cannot be by accident. There must be some reason that an earlier factor can produce a later one if we are to rely on that happening – some principle under which it happens. But the causal principles involved are rarely those that get labelled 'laws';nrandomised controlled trials. Nevertheless, they are what make our policies work.

---

[102] The social/political theorist Jon Elster calls these 'mechanisms', but this is a word with dozens of different meanings in social science and in the evaluation literature, so we avoid it.

pToCs for the same kind of intervention can be at different levels of generality. One of our aims in this paper is to describe what researchers and programme developers should provide in a middle-level programme pToC that will help local individual policy decision makers and implementers thicken it to produce a local pToC they can use for predicting if and how the interventions will work. But clearly middle-level pToCs can vary dramatically in level of generality too.

Just as middle-level pToCs themselves can vary in levels of generality, the middle-level tendency principles involved in them can vary too, from fairly general to fairly local. Principles that are too high buy generality at the cost of utility. Some ideas of Jonathan Fox (2015) from a paper on social accountability strategies (SAcc) to improve institutional performance help to illustrate this point. Fox discusses four conceptual frameworks imported from other intellectual agenda that have not served well. One, at least, is a paradigm of higher as opposed to middle-level theory. It is the  gent theory, which the World Bank's 2004 *World Development Report* (WDR) advocates. Principal-agent theory assumes the relatively high-level tendency principle, which we shall label 'PAP':

> PAP: Public sector performance can be improved if incentives-based contracts (or as WDR suggests, 'compacts') are implemented between citizens (principals) and government institutions (agents).

Fox suggests that in many cases where improvement is needed, citizens are not principals in the relevant sense of being in charge, nor do they have homogeneous interests. This points to one of the standard problems with using high theory: the concepts involved tend to be so abstract that it is hard to tell when they apply. PAP may be a good general tendency principle, one that applies widely to agents and principals. But whether someone can be considered an agent/principal depends on the local setting, to be decided case by case on some concrete criteria that are no part of the theory. Middle-level theory uses far more concrete concepts whose applicability is easier to adjudicate. This is one reason why theory at the middle-level can be so useful.

Many of the tendency principles we can call on for development planning and prediction are research based. The research that supports these will be highly varied in form and method. The principal-agency principle cited above comes out of a tradition of rational choice modelling. It is a more detailed version of an even more general principle that economists are always reminding us of: 'pPeople respond to incentives'.

Priming theory provides an example of a tendency principle that is based primarily on psychological laboratory studies, one that the UK nudge strategy relies on heavily: 'sSubtle cues in the environment may have significant, reliable effects on behaviour'. The PERL principle cited above is said to have moderate evidence in its favour in Nigerian settings (PERL 2009, p. 3).

Fox offers a number of propositions he thinks important to consider for social accountability projects, some of which are tendency principles, such as this one:

> Voice can be constrained by 'the fear factor'.

These, he explains, were 'initially developed inductively from both top-down and bottom-up SAcc efforts in Mexico over more than two decades' (p. 352).

Other tendencies that we regularly rely on are familiar and well known, for instance that mothers seek to nurture and protect their children. These do not need research to establish them. Our problem in using them is not that we have insufficient reason to think they obtain widely, but rather that we are often not sure what will reliably trigger them, what might get in their way or what mothers themselves may conceive as nurturing or protecting.

We apologise if we are making heavy weather of this. But the tendency principles needed for pToCs are causal principles, and it has become a mantra in evidence-based policy that causal principles need randomised controlled trials to back them up. Tendency principles do not have the right character to be tested by randomised controlled trials. They are instead backed up, as most of proper science is, byrequire a tangle of conceptualising work, concept validation studies, theorising and a great variety of case studies.[103] Because of this, some policy developers and researchers may be reluctant to be explicit about them, for fear of appearing to offer low-quality work.

Whether they are acknowledged or not, tendency principles are what will allow the programme to work, if it does. These principles need to be clearly and explicitly stated in the middle-level programme pToC so that local decision makers and planners can consider whether they can and will operate in the local setting. Adding these principles can, in general, be a powerful new tool for effectiveness prediction in the toolbox of evidence-based policy.

---

[103] See for example Cartwright et al. Forthcoming.

## References

Abah, J. 2012. *Strong Organisations in Weak States*. PhD Dissertation. Maastricht University.

Abah, J. 2017. *Making public sector reform a reality in Nigeria*. ARI Anniversary Lecture.

Andrews, M. 2008. 'The good governance agenda: beyond indicators without theory', *Oxford Development Studies*, 36(4), pp. 379–407.

Andrews, M. 2018. 'New connections and better performance in Nigeria's budget process'. Building State Capability blog, 24 January. Available at: https://buildingstatecapability.com/2018/01/24/new-connections-and-better-performance-in-nigerias-budget-process/ (Accessed: 21 July 2020).

Anscombe, G. E. M. 1971. 'Causality and determination', in Sosa, E. (ed.) *Causation*. …, pp. 88–105.

Armand, A. and Carneiro, P. 2018. *Impact evaluation of the conditional cash transfer program for secondary school attendance in Macedonia*. 3ie Impact Evaluation Report 69. New Delhi: International Initiative for Impact Evaluation (3ie).

Baird, S., Ferreira, F. H. G., Özler, B., and Woolcock, M. 2013. 'Relative Effectiveness of Conditional and Unconditional Cash Transfers for Schooling Outcomes in Developing Countries: A Systematic Review', *Campbell Systematic Reviews*, … .

Bankole, A., Fasanya, I., Wahab, O. and Adeleke-Adedoyin, P. 2018. *'Study of Ogun State's experience of improving ease of doing business*. Ibadan School of Government and Public Policy (ISGPP).

Barnett, I., Sulistyo, S. S., Befani, B., KariSari, K., Sharmin, S. and Dewi, D. 2016. *Mixed-Method Impact Evaluation of a Mobile Phone Application for Nutrition Monitoring in Indonesia*. …: Institute of Development Studies.

Benhassine, N., Devoto, F., Duflo, E., Dupas, P., and Pouliquen, V. 2015. 'Turning a shove into a nudge? A "labeled cash transfer" for education', *American Economic Journal: Economic Policy*, 7(3), pp. 86–125. https://doi.org/10.1257/pol.20130225.

Bex, F. J. 2011. *Arguments, Stories and Criminal Evidence – A Formal Hybrid Theory*. Law and Philosophy Library 92.

Booth, D. 2012. *Development as a Collective Action Problem*. London: Overseas Development Institute; published on behalf of the Africa Power and Politics Programme.

Boudon, R. 1991. 'What Middle-Range Theories Are', *Contemporary Sociology*, 20(4), pp. 519–522. doi:10.2307/2071781.

Bureau of Public Service Reforms. 2015. *How Nigeria Contained Ebola: Lessons for Institutional Reform*. Available at: https://www.bpsr.gov.ng/index.php/publications/bpsr-resources/evaluations?download=41:how-nigeria-contained-ebola-lessons-for-institutional-reform (Accessed: 21 July 2020).

Cartwright, N. 1989. *Nature's Capacities and Their Measurement*. Oxford University Press. https://ideas.repec.org/b/oxp/obooks/9780198235071.html.

Cartwright, N. and Hardie, J. 2012. Evidence-Based Policy: A Practical Guide to Doing It Better. Oxford University Press.

Cartwright, N., Hardie, J., Montuschi, E., Soleiman, M. and Thresher, A. Forthcoming. *The Tangle of Science: Reliability Without Rigour*. Oxford University Press.

Collier, D. and Mahon, J. 1993. 'Conceptual "stretching" revisited: adapting categories in comparative analysis', *American Political Science Review*, 8, 7(8), pp. 45–55.

Crook, R. C. and Booth, D. 2011. 'Working with the grain? Rethinking African governance', *IDS Bulletin*, 42(2), ….

Das, J., Do, Q.-T. and Özler, B. 2005. 'Reassessing conditional cash transfer programs'. *The World Bank Research Observer*, 20(1), pp. 57–80.

Davies, R. 2018. 'Representing theories of change: a technical challenge with evaluation consequences', Inception Paper 15, CEDIL: London.

de Gramont, D. 2015. *Governing Lagos: Unlocking the Politics of Reform*. Carnegie Endowment for International Peace.

Deaton, A. 2019. 'Randomization in the tropics revisited: a theme and eleven variations'. Available at: https://scholar.princeton.edu/sites/default/files/deaton/files/deaton_randomization_revisited_v2_2019_01.pdf (Accessed: 21 July 2020).

Deaton, A and Cartwright, N. 2018. 'Understanding and misunderstanding randomized controlled trials', *Social Science & Medicine*, 210, pp. 2–21. https://doi.org/10.1016/j.socscimed.2017.12.005.

DFID. 2016. *Jordan CSSF: Community Level Conflict Resolution Programme Document*. Available at: http://iati.dfid.gov.uk/iati_documents/5654077.odt (Accessed: 21 July 2020).

DFID. 2017. *Local Government Development Programme, Afghanistan (ACCESS: Accountable, Citizen-Centred Sub-National Governance Strengthening)*. Available at: https://devtracker.dfid.gov.uk/projects/GB-GOV-1-300437/documents (Accessed: 21 July 2020).

Dhillon, L. and Vaca, S. 2018. 'Refining theories of change', *Journal of MultiDisciplinary Evaluation*, 14(30), pp. 64–87.

EBM+ *Evaluating Evidence in Medicine*. Available at: http://ebmplus.org (Accessed: 21 July 2020).

Edmonds, E. 2008. 'Child Labor', *Handbook of Development Economics*. Elsevier.

Elster, J. 2007. Explaining Social Behavior. Cambridge University Press.

Fiszbein, A. and Schady, N. R. 2009. *Conditional Cash Transfers: Reducing Present and Future Poverty*. World Bank Publications.

Fox, J. 2015. 'Social Accountability: What Does the Evidence Really Say?' *World Development*, 72, pp. 346 -61.

Future State, Centre for the. 2010. *An Upside Down View of Governance*. Brighton: Institute of Development Studies.

Gaarder, M. 2012. 'Conditional versus Unconditional Cash : A Commentary'. *Journal of Development Effectiveness* 4 (1), pp. 130–33. doi:10.1080/19439342.2012.658635

Garcia, S. and Saavedra, J. 2013. 'Educational impacts and cost-effectiveness of conditional cash transfer programs in developing countries: a meta-analysis', CESR Working Paper No. 2013-007. Available at: http://papers.ssrn.com/abstract=2333946 (Accessed: 21 July 2020).

Grindle, M. S. 2011. 'Governance reform: the new analytics of next steps', *Governance*, 24(3), pp. 415–418.

Grosh, M., del Ninno, C., Tesliuc, E. and Ouerghi, A. 2008. *The Design and Implementation of Effective Safety Nets for Protection and Promotion*. Washington, DC: World Bank.

Haiku Analytics. 2018. *Logic Models and Complex System Mapping for Evaluation, Performance Measurement and Planning*. Available at: http://haikuanalytics.com/dynamic-system-mapping/ (Accessed: 21 July 2020).

Hitchcock, C. (2018) 'Probabilistic causation', *The Stanford Encyclopedia of Philosophy* (fall 2018 edn), Zalta, E. N. (ed.). https://plato.stanford.edu/archives/fall2018/entries/causation-probabilistic/

Hussein, K. and Sharp, S. 2018. *Public Sector Reform in Nigeria – What Works?* London: DFID – Partnerships to Engage, Reform and Learn.

International Rescue Committee. No date. *Outcome and Evidence Framework*. Available at: http://oef.rescue.org/#/outcome/24?_k=iamnjm (Accessed: 22 July 2020).

Keefer, P. and Khemani, S. 2003. *Democracy, Public Expenditures and the Poor*. World Bank Policy Research working paper 3164. Available at: https://elibrary.worldbank.org/doi/abs/10.1596/1813-9450-3164… (Accessed: 21 July 2020).

Khan, M. H. 2007. 'Governance, economic growth and development since the 1960s', in Ocampo, J. A., Jomo, K. S. and Vos, R. (eds.) *Growth Divergences: Explaining Differences in Economic Performance*. London: Zed Books/United Nations, pp. 285–323.

Layne, C., Steinberg, J. and Steinberg, A. 2014. 'Causal reasoning skills training for mental health practitioners: promoting sound clinical judgment in evidence-based practice'. *Training and Education in Professional Psychology*, 8(4), pp. 292–302.

Lewis, P. and Watts, M. 2015a. 'The Politics of Policy Reform in Nigeria'. Doing Development Differently Discussion Paper. World Bank.

Lewis, P. and Watts, M. 2015b. 'Nigeria: The Political Economy of Governance'. Doing Development Differently Discussion Paper. World Bank.

Mackie, J. L. 1980. *The Cement of the Universe: A Study of Causation*. Oxford University Press.

Mahoney, J. and Goertz, G. 2006. 'A tale of two cultures: contrasting quantitative and qualitative research', *Political Analysis*, 14, pp. 227–249.

Merton, R K. 1968. *Social Theory and Social Structure.* 1968 enlarged edn. New York: Free Press.

Mill, J. S. 1848. *Principles of Political Economy*. Oxford University Press.

Monitoring and Evaluation NEWS. 2011. *M&E Software: A List*. Available at: https://mande.co.uk/2011/lists/software-lists/me-software-a-list/ (Accessed: 22 July 2020).

Munro, E. 2011. *The Munro Review of Child Protection: Final Report. A Child-Centred System*. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/175391/Munro-Review.pdf. (Accessed: 21 July 2020).

Munro, E., Cartwright, N., Hardie, J. and Montuschi, E. 2016. *Improving Child Safety: Deliberation, Judgement and Empirical Research*. Available at:

https://www.dur.ac.uk/resources/chess/ONLINE_Improvingchildsafety-15_2_17-FINAL.pdf. (Accessed: 21 July 2020).

Munslow, T., Barnett, I. and Dewi, D. 2016. The Potential of M-health for Improved Data Use: Workshop Report, IDS Evidence Report 201. Brighton: Institute of Development Studies.

NHMRC *Building a Healthy Australia*. 2018. Available at: https://www.nhmrc.gov.au/guidelinesforguidelines/plan/adopt-adapt-or-start-scratch. (Accessed: 21 July 2020).

Ostrom, E. 1965. *Public entrepreneurship: a case study in Ground Water Basin Management*. PhD dissertation. University of California, Los Angeles.

Ostrom, E. 1990. *Governing the Commons: The Evolution of Institutions for Collective Action*. New York: Cambridge University Press.

PERL. 2019. *Theory of Change and Theory of Action Update Note*. Programme Level Report. …

Pawson, R. and Tilley, N. 1997. Realistic Evaluation. London: SAGE Publications.

Pawson, R. 2009. 'Middle-range realism'. *European Journal of Sociology / Archives Européennes de Sociologie*. https://doi.org/10.1017/S000397560000705

Pawson, R., Greenhalgh, T., Harvey, G. and Walshe, K. 2004. Realist synthesis: an introduction. ESRC Research Methods Programme.

Porter, D. and Watts, M. 2016. 'Righting the resource curse: institutional politics and state capabilities in Edo State, Nigeria', *The Journal of Development Studies*, 53, pp. 249–263.

Rasul, I. and Rogger, D. 2017. 'Management of bureaucrats and public service delivery: evidence from the Nigerian Civil Service', *Economic Journal,* 128(608), pp.413-446.

Reed, J. E., McNicholas, C., Woodcock, T., Issen, L. and Bell, D. 2014. 'Designing quality improvement initiatives', *BMJ Quality & Safety*, 23, pp. 1040–1048.

Rogger, D. 2014. 'The causes and consequences of political interference in bureaucratic decision making: evidence from Nigeria'. Job Market Paper. University College London.

Roll, M. 2014. 'The state that works: a "pockets of effectiveness" perspective on Nigeria and beyond', *States at Work*, pp. 365–397.

Roll, M. 2015. *The Politics of Public Sector Performance: Pockets of Effectiveness in Developing Countries*. Routledge.

Rudra, S. 2000. 'The division of labor in social science research: unified methodology or "organic solidarity"?', *Polity*, 32(4), pp. 499–531.

Salmon, W. 1998. *Causality and Explanation*. Oxford University Press.

Sartori, G. 1970. 'Concept misinformation in comparative politics', *American Political Science Review*, 6, 4(1), pp. 33–53.

Simbine, A., Attoh, F. and Oladeji, A. 2015. 'Taming the menace of human trafficking', in Roll, M. (ed.). *The Politics of Public Sector Performance: Pockets of Effectiveness in Developing Countries*. Routledge, pp. 128–146.

Snilstveit, B., Stevenson, J., Phillips, D., Vojtkova, M., Gallagher, E., Schmidt, T., Jobse, H., Geelen, M., Pastorello, M. G., and Eyers, J. 2015. 'Interventions for improving learning outcomes and access to education in low-and middle-income countries: A systematic review, *3ie Systematic*

*Review*. London: International Initiative for Impact Evaluation (3ie).
https://www.3ieimpact.org/sites/default/files/2019-05/SR24-education-review_2.pdf

Unsworth, S. 2009. 'What's politics got to do with It? Why donors find It so hard to come to terms with politics, and why this matters', *Journal of International Development*, 21(6), pp. 883–894.

White, H. 2009. 'Theory-Based Impact Evaluation: Principles and Practice'. 3ie Working Papers. …: International Initiative for Impact Evaluation (3ie). https://www.3ieimpact.org/evidence-hub/publications/working-papers/theory-based-impact-evaluation-principles-and-practice

White, H. 2018. 'Theory-based systematic reviews', *Journal of Development Effectiveness*, 10(1), pp. 17–38. 10.1080/19439342.2018.1439078.

White, H. and Booth, R. 2003. 'Using development goals to design country strategies', Black, R. and White, H. (eds.) *Targeting Development: Critical Perspectives on the Millennium Development Goals*. London and New York: Routledge.

World Bank. 2004. *World Development Report*. *Making Services Work for Poor People.* Washington, DC: Oxford University Press/World Bank…. https://openknowledge.worldbank.org/handle/10986/5986

World Vision International. 2015. *mHealth: Indonesia*. Available at: https://www.wvi.org/sites/default/files/Indonesia%20mHealth%202014%20v5%20pdf.pdf (Accessed: 20 October 2019).

World Vision International. 2016. *mHealth: Indonesia*. Available at: https://www.wvi.org/sites/default/files/Indonesia%20mHealth%20Factsheet%202016.pdf (Accessed: 20 October 2019).

World Vision International. 2018. *Global mHealth Report*. Available at: https://www.wvi.org/sites/default/files/2018%20mHealth%20Report-FINAL-electronic.pdf (Accessed: 20 October 2019).