



SNL Working Paper

# **SCALE-UP AS IF IMPACT MATTERED**

**Learning and Adaptation as the Essential  
(often missing)  
Ingredient**

Stephen Hodgins

Katy Quissell

DECEMBER 2016

Save the Children

899 North Capital St NE, suite 900

Washington DC 20002

## Contents

Executive Summary.....	1
Introduction .....	9
Methods.....	10
Interview Findings.....	14
What is the Initiative Intended to Achieve? .....	14
Measurement .....	24
Ownership.....	30
Barriers to Acknowledgement and Response.....	33
Learning and Adaptation .....	39
Conclusions .....	48
References .....	51
Other Resources.....	52
Annex I: Literature Review.....	53
Annex II: Interview Guide .....	66
Annex III: Initial Coding Used in Text Analysis .....	67



## Executive Summary

Over the past generation, globally, there have been enormous gains in health and well-being. In large part this has been a consequence of improvements in economic conditions. And there have been synergistic effects across social sectors; for example, improvements in women's level of education and literacy have had further benefits with regard to child health, nutritional status and well-being. But specific program efforts in health have also made an important contribution. Primary healthcare programs and services have been developed and in many important instances are reaching a large proportion of those who can benefit from them and have made an important contribution to improved population health status. Immunization is a case in point.

Nevertheless there are also instances *of initiatives or program efforts that develop momentum and continue, drawing considerable resources, despite not making the impact they were intended to*. There is a failure to acknowledge and remediate.

This phenomenon of *dysfunctional persistence of efforts that in fact are not delivering* is seen across human endeavors. Most people in the United States are aware of the DARE anti-drugs program which has been running for over 30 years and, at its peak, was active in 3/4s of all school districts in the country. Billions have been spent on the program and it continues to chug along despite rigorous evaluations that have systematically demonstrated that it's not been working (Lynam 1999, Birkeland 2005).

There is now a well-developed literature on the phenomenon of *escalation*, the process of increasing commitment to a failing course of action (Ross and Staw 1987, 1997; Brockner 1992; Arkes and Blumer 1985; and others). But beyond factors identified in this literature from other sectors, there are also dynamics more specific to global health and development assistance. This discussion paper presents results of an initial effort to investigate this set of issues. It is intended as part of a broader ongoing effort to better characterize the problem, identify solutions and galvanize efforts for improved effectiveness.

This paper presents findings of a series of interviews with global health leaders with substantial on-the-ground experience. With input from an advisory group, a list was drawn up consisting of 30 potential interviewees. The general criteria for selection included:

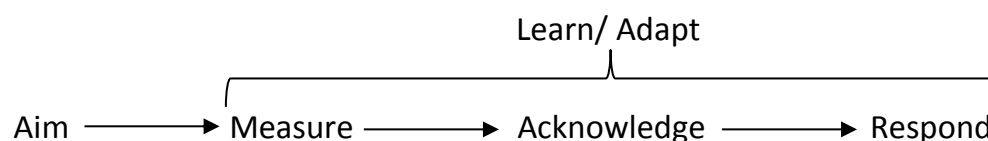
- Anticipated depth of insight into the questions of interest,
- Breadth and depth of experience, and
- Diversity across the sample with regard to: geography of program experience (focusing mainly on low-income countries), global health content areas, and institutional role (ministry of health, donor agency, technical assistance agency, evaluator).

Of the 30, in the end 22 were interviewed; several, on more than one occasion. The interviews focused on *detection of and response to indications of significant performance problems in global health program efforts*. At one end of a continuum, interviewees were asked to report on experiences they were familiar with in which program efforts continued unchanged, with failure either to recognize or respond to indications that the program was not achieving its intended objective. At the other end of the continuum, they were asked to recall cases in which there was a serious effort to track the extent to which the program effort was achieving its overall objectives, where there were indications of significant performance problems followed by timely and appropriate course corrections. And they were asked to recall and comment on cases lying between these two extremes. They were asked for their view on important factors contributing to these varied patterns of response and to suggest remedies to create conditions more favorable for a more functional pattern of vigilance, early recognition and suitable response.

Although the methodology used was not in-depth studies of particular cases, certain program experiences (positive and negative) were discussed by multiple interviewees. Those most commonly addressed included: the Integrated Child Development Scheme (ICDS) from India, Integrated Management of Childhood Illness (IMCI), Helping Babies Breathe (HBB), BRAC, and immunization program efforts. To protect interviewee confidentiality, in some instances there is masking of program identity in presentation of interview findings in this paper.

Below, we summarize the main findings from these interviews. Results in the paper are organized using the following model (which is explained in the Methods section):

Figure: Conceptual Framework for Learning and Adaptation



An important theme emerging from our interviews is the need to clearly think through and specify *what* a program effort is intended to accomplish, particularly in terms of a population health benefit. Related to this is a non-perfunctory effort at elaborating a realistic program theory that spells out *how* the effort is expected to give rise to the intended population health effect. This will necessarily oversimplify; the real world is invariably too messy and complex to be adequately captured in our plans. But, even if put forward on a tentative basis, the program theory needs to be critically elaborated as a serious early step. Failure at this stage is a common contributor to dysfunctional escalating program efforts.

In clearly specifying the overall population health intention for an initiative, we state an *explicit* goal. In the real world, it will always be the case that key stakeholders will have other *implicit* goals or side-bets. In some instances key players will have implicit goals that are, in fact, more important to them than the explicit goals of an initiative. This cannot be entirely avoided. However, to the extent that high-priority goals of the main players are not fully consonant with the explicit goal, this can undermine effectiveness. Important implicit goals highlighted by our interviewees included: money, relationships and maintaining the appearance of success.

Clearly, at all levels *money* can be a critically important driver which can push an initiative in a direction that undermines its effectiveness in achieving its population health goal. Our interviewees described such circumstances affecting individuals, organizations, and between-organization relationships. Keeping the money flowing—in many instances—can trump other considerations.

Perhaps a little less obvious, as our interviewees report, *maintaining and cultivating relationships* can also take precedence over the effectiveness of a program effort. There may be important political reasons for a donor to “play together nicely” with a host government, or vice versa, continuing to support an initiative that they are well aware is not performing.

A very important factor discussed by almost all our interviewees is maintaining the *appearance of success*. This is at work at a variety of levels. For an individual program manager, attention to such appearances can enhance one’s career prospects. Implementing agencies are incentivized to “show results” and tell “success stories.” A donor agency will sell its work to its country’s legislature, as justification for continued funding. Maintaining the appearance of success is often closely linked with ensuring the continued flow of funds. So, across our interviews we heard repeatedly about spin, hype, and appearance of success driving programs and inhibiting critical, reflective practice.

Another pervasive problem identified both by our interviewees and in papers included in our literature review is the recourse to *fixed universal strategies*. As we've seen, this approach tends to lay the ground for escalation and empty scale-up. Instead, what's needed is a context-responsive, learning/ adaptive stance.

A requirement for effectively managing a program effort and helping ensure that it genuinely achieves its intended population health goal is serious *measurement along the presumed causal path*. As our interviewees pointed out, providing that the efficacy of the technical interventions is well established one needn't necessarily be tracking actual health impact if, instead, it is feasible to track the proportion of the population needing the intervention that actually gets it (effective coverage). To the degree that the effectiveness of the intervention depends on some aspect of quality, that too needs to be considered as an integral aspect of effective coverage. But to ensure that the overall effort is effective generally also requires tracking what's happening further upstream. For example, if an intervention is commodity-dependent, it may be necessary to track supply-chain performance. These various aspects of systems performance which may need to be tracked can be considered aspects of "strength of implementation." But the only reason for measurement at these various points along the causal path is to inform efforts for detecting and correcting error, including "double-loop" learning, i.e. reflection on and possible correction to the overall strategy.

The first requirement for effective measurement—to serve this needed learning/ adaptation function—is that such measurement and data use be taken seriously for informing decision-making. And, as we heard from our interviewees, this condition is frequently not met. Often, the bar is set low, using easy-to-collect input indicators. But even when taken seriously, measurement can be problematic. There can be important inherent difficulties including weak capacity and systems, difficulty with observability of key elements of the intervention, and lack of timeliness using current data sources. Likewise, even with seriousness about using data for improving program performance, it can be inherently challenging to work out the key underlying remediable drivers of poor performance (i.e. answering the 'why' question). An important problem identified by our interviewees in some program efforts has been the use of misleading proxies of overall program performance, which may be readily measurable but poor indicators of how well the program effort is doing in achieving its population health status goals and can ultimately drive misplaced program effort (for example, tracking "skilled birth attendance").

It is evident that it is hard to objectively evaluate one's own work. As much as we need to cultivate openness to negative findings and a genuinely learning culture, we also need independence—not in the sense of evaluators that stand completely apart from implementers—but we need a partnership between those charged primarily with implementation and those assessing performance, in a joint, ongoing, problem-solving/ improvement effort.

Structural impediments and misaligned incentives are important contributors to the dysfunctional patterns our interviewees described and although they may not be immediately amenable to change, in principle they can be, and need to be, changed.

Pressure to rapidly scale up program effort (paying less serious attention to how a program is actually performing) is often a prescription for empty scale-up. Governments, donors and others need an appreciation of what pace is realistic—for a given program effort under the conditions obtaining in a particular setting—if they also have a serious interest in program effectiveness.

As we have found, ownership cuts two ways—*no sense of ownership*, for example on the part of government, will tend to result in relatively little concern whether or not an initiative is actually achieving its public health goal. A *strong sense of ownership*, on the other hand, will create commitment and this can include seriousness about the fundamental objective. However, *over-*

*investment* [my baby] can be a barrier to a critical reflective stance. Genuine successes are normally characterized by good *shared ownership*—across host government, donor, technical agency and evaluator.

High-profile, politicized initiatives can open very little space for the acknowledgement of failure or poor performance. It is inevitable that certain initiatives will be politicized. However, we heard of such cases where courageous program staff were able to negotiate space for critical review and revision.

Current patterns and practice, for example within organizations or between donor and implementer, often include incentives that are aligned such that they discourage acknowledgement of problematic performance (giving rise to faked data, etc.); one is punished, not rewarded, for calling attention to problems. This is, in principle, remediable.

As our interviewees report, not uncommonly our program initiatives proceed to the point where there can be very big costs to calling an effort into question. Initiatives can, in effect, be forbidden to fail. It can be far easier to simply carry on. Typically, the standard of evidence required to support continuing can be very low but to call for a halt, very high.

Interaction between donors and implementers often contributes to dysfunctional escalation and resistance to take action for endeavors that are, in fact, not delivering. How serious a problem this is can depend on the relationship between the donor agency and its own source of funds and—in particular—the degree to which learning and adaptation is acceptable to or supported by the holder of the purse-strings. To the extent that institutional norms are inimical to learning/ adaptation, program effectiveness will be compromised. But the difficult task of changing these norms must be engaged.

Available literature provides useful insights into barriers and facilitators influencing dysfunctional escalation vs. learning/ adaptation—particularly at the levels of individual psychology and within organizations. We face challenges of hard-wired psychology (notably self-justification and our inability to look objectively at our own work) and with the cultures that tend to develop within organizations, but it's clear that these challenges can be at least to some extent overcome, achieving less dysfunctional patterns of behavior, to avoid getting locked into failing endeavors.

In both the interview material and the literature considered, it is clear that although more functional patterns are possible, this requires culture change and committed leadership. There are good examples of genuine learning organizations and instances where donors support such a stance. Change and improvement is possible. The best strategy may be incremental; we eat an elephant one mouthful at a time.

There is a well-developed literature on creating conditions favoring the development of a learning culture in organizations that can be drawn on to remedy the dysfunctional patterns discussed by our interviewees. It behooves those in governments, donor agencies, technical assistance agencies and evaluator groups who care about population impact in global health to tackle dysfunctional patterns in our collective work that undermine our effectiveness.



## **Key Points**

### **The Challenge:**

- Acknowledging that a program effort is not yielding its intended benefit and appropriately changing course.

### **The Causes:**

- Simplistic, context-blind, rigidly implemented plans (lack of a learning/ adaptive stance)
- Measurement:
  - Not taken seriously as a means for directing adjustment/ improvement
  - Not looking at the right things
  - Inherent challenges in discerning how we're doing with regard to ultimate goals
- Self-justification and lack of objectivity with regard to one's own work
- Momentum, having a heavy investment in the current effort, maintaining appearance of success

### **Possible Remedies:**

- In the first instance, acknowledging this is an issue worth attention
- New culture, expectations, norms, procedures

The table below summarizes important findings from the interview material, including identified problems and corresponding solutions:

Problem	Solutions
<b>Poorly specified causal models/ program theory</b>	More serious effort to characterize what all would need to happen to achieve sustained impact at scale; need the right goals (i.e. closely linked to impact)
Use of <b>poor proxies</b> for overall performance, resulting in misdirected program effort	Need <i>appropriate</i> means of tracking overall performance (impact, effective coverage)
<b>Blind implementation</b> , not seriously tracking/ managing for performance, adapting	Measurement along the causal path: “effective coverage,” key implementation factors
<b>Implicit objectives</b> can take precedence over explicit	Need greater honesty; efforts to reconcile side objectives to ensure effort on explicit objective(s) is not compromised
Inherent <b>measurement problems</b> : observability, capacity, timeliness	Don’t give up; grapple seriously with this issue, otherwise we set ourselves up for ineffectiveness
<b>Compartmentalization</b> (with specific partners responsible only for particular pieces of the overall program effort), with lack of attention to our accountability for overall performance; corresponding fragmented focus on inputs; diffusion of responsibility.	Mechanisms to ensure a coordinated effort, keeping the big picture in view. Avoiding letting available toolkit drive strategy.
<b>Structural disincentives to openness</b> – heavy emphasis on mere <i>appearance of success</i>	Need to recognize this as an important systemic problem, requiring changes in organizational culture & procedure.
<b>Donor–host government relationship</b> can create a dynamic in which program effectiveness is not prioritized, with broader political considerations taking precedence (and intended program beneficiaries losing out)	This may be a reality that can’t easily be changed, but we should certainly be striving for greater honesty, as well as accountability to end-beneficiaries for whom these programs are supposedly created.
Recourse to <b>supposedly universal solutions</b> and associated misplaced confidence that “of course, it must work” and therefore doesn’t require careful ongoing validation	While large-scale implementation requires some standardization, the typical recourse to prescribed universal strategy must be challenged. There needs to be a fundamental shift towards program solutions that are responsive to the actual situation on the ground.  Need to differentiate: complicated/ complex, decomposable/ non-decomposable, well-structured/ ill-structured problem—we tend to assume things are simpler than they really are, not adequately taking into account context and systems. We need a more sophisticated approach in developing solutions. We need more modesty in our predictions, checking to see what’s happening and then adjusting based on what we observe.

<p><b>Fixed program trajectory</b> (implementing the Plan) – without adequate provision for checking/ adjusting.</p> <p>Typical dynamic:</p> <ol style="list-style-type: none"> <li>1) a relatively fixed (imported) strategy;</li> <li>2) perfunctory “results reporting”, generally focused on inputs;</li> <li>3) incentivizing the appearance of success;</li> <li>4) comparatively little scope for learning &amp; adaptation.</li> </ol>	<p>Need more realism about how change happens at scale in the real world (take context and systems fully into account). Changed expectations on the part of donors, host governments. Changed culture. Changed norms for project funding. From the beginning, expect there will be problems, build in provision for review/ adjustment.</p> <p>Need to move away from plan rigidity: No plan survives contact with the real world fully intact therefore, holding to the overall goal, but with scope to respond to the real, evolving situation on the ground.</p> <p>Recognition that we will never get it perfect from the start. Instead need to continuously design as we learn. Creating a culture of learning/ adaptation.</p> <p>Need change in donor expectations, better acknowledging real-world complexity and need for a flexible learning approach.</p> <p>Need <i>continuous</i> iterative learning and adaptation. Need a middle piece between small-scale, intensively supported pilots and programs operating at national scale, learning how to effectively implement at scale.</p>
Legislatures appropriating funds for health/ development work have simplistic views on what’s required for impact, resulting in overly simplistic, context-blind program strategy and inflexibility (“just implement the Plan”).	A big problem, but shouldn’t be dismissed as an immovable given. This is a structural, systemic problem that nevertheless needs to be effectively engaged.
Lack of <b>objectivity</b> concerning our own work; confirmation bias, post-hoc self-justification, etc.	Need to recognize this and explicitly build in mitigation mechanisms, notably by providing for <u>independence</u> ; fresh eyes, no skin in the game
<b>Misappropriated ownership</b> – undermining investment in what’s actually happening	Need to acknowledge/ recognize the issue, and appropriately deal with it when the right balance isn’t struck
Seemingly unstoppable <b>momentum</b> – too big to change course (escalation/ entrapment)	Certainly, try to avoid getting into this situation in the first place; recognize that if this <i>is</i> the situation, it will be a tough slog to change course but the sooner the situation is fully acknowledged and an appropriate effort begins the better. Otherwise it will tend to only get harder.

As has already been stated, this document captures work conducted as the first stage in what is expected to be a larger effort. Primarily, it serves to sketch out the contours of an important problem compromising effectiveness of global health programs. Although the interviews and literature review also sought out positive cases and remedies, this dimension is less well developed. A key question for this work, as we move forward, is:

### What concrete steps can move us towards sounder, more effective program efforts?

It is intended, as a next step, that one or more small consultative working meetings will be convened, drawing in a broader community of global health practitioners and thinkers, to better characterize key remediable drivers and to begin to formulate a program of action. One such meeting was held in Washington DC in May 2016.

## Introduction

In normal daily life, things seldom go strictly to plan. But this unpredictability needn't be a big problem, providing we're poised to recognize and appropriately respond when events begin to deviate from our expectations. What's common sense in daily life unfortunately is somewhat less common in our collective life. This applies in global health as it does in many other enterprises. As one of the interviewees for this study stated:

"I have more examples where things aren't going well and nobody's worried about it (chuckles):  
I have hundreds of those."

Global health work over the past several decades can be seen as consisting, in part, of many initiatives or program efforts, each of which has charted a certain course over time. They may begin with results of an effectiveness study or a pilot project, or simply from someone's bright idea. Whatever their origin, a judgment is made that the intervention or approach is promising, momentum begins to build, resources are committed, and a process of scale-up begun and carried forward. There are certainly important global health success stories that have followed this trajectory and had genuinely important impact at large scale.\* But there are also cases where all the other ingredients are there, including lots of positive language about "achieving results," but where—if one looks closely—one finds that the effort is in fact not at all achieving its explicitly stated objectives. Nevertheless, the efforts continue to chug along without course corrections or indeed any real acknowledgement that they're not delivering.

What's missing in these cases is an effective learning and adaptation function, where there's vigilance to detect important performance problems, willingness to acknowledge them, and an appropriate response. This paper investigates this phenomenon in global public health, drawing also on lessons documented in the literature from other fields beset by similar challenges, and offers some preliminary suggestions on how to remedy this problem.

In an American domestic setting, almost everyone is familiar with the DARE (Drug Abuse Resistance Education) anti-drug program which, at its peak, was being implemented in three quarters of all school districts in the country. It has now been running for more than 30 years. Billions have been invested. And there is a rich evaluation literature documenting that it has been systematically ineffective (Lynam 1999). It doesn't work. But it keeps on going. Of course, this is a fairly extreme case. There are such cases in global health, as we shall see. But there are far more in the "muddled middle." This paper will explore experiences across a continuum, with such paradigmatic negative cases at one end, the common cases of the muddled middle, and the "positive deviants" at the other end of the continuum in which there is a genuine learning culture, where there is serious effort from the beginning to get a sense of whether or not an effort is on track, there is vigilance in monitoring performance, problems are acknowledged early, and needed action is taken even if it requires painful decisions.

The effort of which this working paper is primarily addressed to ministries of health, donors, technical assistance agencies, and evaluators. Its purpose is to draw on experience in global health programs and in the relevant literature to help inform more effective program efforts. In particular, it is intended to investigate and better understand the conditions that help or hinder timely *acknowledgement* that an initiative is actually not on track to achieve its intended objectives and an appropriate *response* to the identified problem. And, more fundamentally, it is intended to contribute to more consistently effective program work in global health.

---

\* E.g. [Ten Great Public Health Achievements Worldwide, 2001—2010](#), [Millions of Lives Saved – Case Studies](#)

## Methods

With input from an advisory group, a list was drawn up consisting of 30 potential interviewees. The general criteria for selection included:

- Anticipated depth of insight into the questions of interest,
- Breadth and depth of experience, and
- Diversity across the sample with regard to:
  - Geography of program experience (focusing mainly on low-income countries),
  - Global health content areas, and
  - Institutional role (ministry of health, donor agency, technical assistance agency, evaluator).

Of the 30, in the end 22 were interviewed, 16 of whom were men and 13, American. Four had 20-29 years of experience in global health-related work; 8 had 30-39 years; and 10, 40 years or more. Median number of years of such experience was over 35, so the interviewees consisted largely of grey-beards and their feminine counterparts.

Several interviews were done face-to-face but most were conducted by phone or skype. Those interviewed were sent briefing materials in advance, informing them of the objectives and of the key issues to be discussed during the interview (in Annex, see interview guide). Many came prepared with notes. Interviewees were given a description of the characteristics of a paradigmatic negative case in which there is a failure to acknowledge and take action addressing fundamental performance problems. They were asked to recall such cases from their experience, reflecting on the factors contributing to this pattern. They were also asked to recall cases along a continuum, with the other pole being exemplary positive cases in which there was vigilance, early recognition of problems, and timely and appropriate action taken. And they were asked, drawing on their program experience, to suggest remedies to this problem.

The consent process included a request to audio-record the interview. In one case, consent was given only for written note-taking. Verbatim transcripts were done. Text analysis was done using NVIVO, beginning with a set of codes (see Annex) based in part on a categorization of influences on *escalation*, defined as: *the process of increasing commitment to a failing course of action*, adapted from Ross and Staw (1987, 1997). Through an iterative process of coding and reviewing, new codes were developed and systematic coding done.

Note that due to sensitivity of some of the interview material, in many instances quotes have been edited to mask individual or program identifiers.

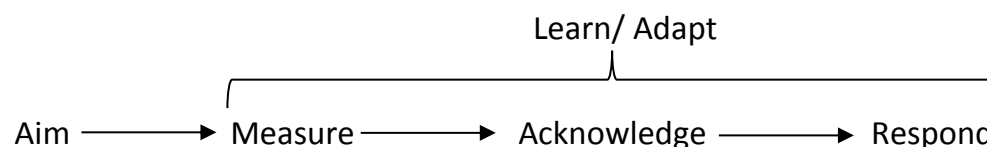
The final conceptual framework used to organize the presentation of findings is a further elaboration arising from the major codes initially developed for the interview material. This is informed by a framework of categories of influence on organizational escalation developed by Staw and Ross (1987), with some modification reflecting the study data. Further, it reflects the focus in the interview questions on factors influencing acknowledgment and taking appropriate action, in the face of evidence of important performance problems.

This is supplemented by concepts emerging from the data on steps logically preceding acknowledgement, notably: the aim or objectives of the initiative, and measurement of performance. We propose this sequence of steps as constituting a process of learning and adaptation. In analyzing the interview material, most of the data were found to map into a matrix consisting of these two dimensions (the Staw framework and the steps in a learning process). As presented in this paper,

findings are organized into major topics or themes emerging from the material. There has been a continuing process of elaborating conceptual models.

The figure below outlines the process that we were investigating in our interviews (see interview guide in the Annex).

Figure: Conceptual Framework for Learning and Adaptation



Although this is intended to reflect a logically sequential relationship between steps, we do not mean to imply that this is necessarily a temporal sequence. Instead, it should be understood that all of these are simultaneously in operation. In this figure, by “Aim” we are referring to what a program initiative is intended to accomplish. For “Measure,” we are concerned about how performance is tracked, particularly with regard to the main objective(s) of the effort. “Acknowledge” refers to recognition that a program effort is, in fact, not on track and that important action is needed. By “Respond,” we are referring to whatever action follows acknowledgement of the problem. The 3 final steps in this sequence, we see as constituting a learning or adaptation process, which may or may not proceed in such a way as to ensure that an effort achieves its intended objective.

As we will see in literature reviewed, a range of factors have been identified as giving rise to a dysfunctional pattern of escalation and entrapment, on the one hand, or appropriate learning and adaptation, on the other. As mentioned earlier, we use the term “escalation” in the sense used by Staw (1976): *the process of increasing commitment to a failing course of action*; and the term “entrapment,” as defined by Weick and Sutcliffe (2003): *a process by which people get locked into lines of action, subsequently justify those lines of action, and search for confirmation that they are doing what they should be doing*.

As part of their work on these phenomena, Ross and Staw (1993) have developed a helpful set of categories of drivers: *project, contextual, psychological, social, and organizational*. This served as a useful starting point for our work, which we have retained in large part. What Ross and Staw categorize as project-related factors correspond roughly to the *characteristics* of the initiative or program effort as well as its *trajectory* as it unfolds over time. A second important dimension that has helped structure our analysis concerns the *setting* in which an initiative unfolds. This can be seen as including *characteristics of the system that delivers the content* of the initiative, as well as *relevant factors impinging from the broader environment*. This corresponds to Ross and Staw’s “contextual” category.

Ross and Staw have 3 other categories, “psychological,” “social,” and “organizational.” These can be seen as referring to important *dynamics within and between the players* involved in an initiative. This we break down further as follows: **intra-organizational** (including social), **individual** (corresponding to Ross and Staw’s “psychological”), and **inter-organizational** (which does not really have a parallel in Ross and Staw’s framework but is an important dynamic for large global health programs). We also focus on several specific *categories of players*, corresponding to the target audiences for the current work, which include: *host governments, donors* (including their

sources of funds), *technical assistance agencies*, and *evaluators*. An important area of interest for us is the contribution of *interactions between these players* to the patterns of dysfunctional escalation vs. learning/ adaptation that are the central concern for this paper.

Although the study was not designed to do in-depth case-studies of particular efforts, certain initiatives or programs came up across multiple interviews, providing some opportunity to get multiple perspectives on a particular program experience. The initiatives most frequently discussed:

- The *Integrated Child Development Services* program (ICDS), in India
- *Integrated Management of Childhood Illness* (IMCI) and the related, more recent initiative, *Integrated Community Case Management* (iCCM)
- *Helping Babies Breathe* (HBB)
- Oral Rehydration Treatment (ORS), particularly the *BRAC* experience
- Immunization services
- Two linked HIV programs involving use of antiretroviral drugs: Prevention of Mother-to-Child Transmission (PMTCT) and antiretroviral treatment (ARV)
- Skilled Birth Attendance (SBA)

Following are details on several of the initiatives listed above:

**Box 1: The Integrated Child Development Scheme (India)**

- By various measures (e.g. number of beneficiaries, number of workers), it can be considered the world's largest child welfare program.
- Launched in 1975 under the federal government, Ministry of Women and Child Development; implemented by the states. Has had major support from World Bank, Unicef, World Food Program, USAID and other donors, though most funding is from the Government of India.
- Addresses nutrition (including food supplementation) and psychosocial development for children 0-6 years of age, with linkages to public health services for immunization and health checks.
- Main services provided by Anganwadi Workers. Over 1.2 million functioning Anganwadi centers.
- Annual funding approximately US\$3 billion (2012-2013).
- Numerous evaluations have shown little, if any, population level impact, with exceptions in some settings (notably Tamil Nadu, which has implemented an adapted, more targeted strategy).

**Box 2: Integrated Management of Childhood Illness (IMCI) & Integrated Community Case Management (iCCM)**

**IMCI**

- First developed by WHO and UNICEF in 1992.
- Constitutes a syndromic approach to outpatient care of the sick child with use of clinical classification algorithms (as well as assessment of status for certain preventive services, e.g. immunization).
- Has a particular focus on assessment and appropriate care for diarrhea and possible pneumonia, as well as for malaria in settings endemic for the condition.
- In principle, the approach includes three major elements: clinical, health systems and community. However, in practice, most program effort has been on off-site, group-based, clinical training, usually lasting 11 days. This training includes classroom work, videos and exposure in a clinical setting.
- Has been widely rolled out (adopted in >100 countries), mainly as in-service training, but also taken up into pre-service training.
- Rigorous evaluation has generally showed little evidence of impact (Bryce 2004, 2005) although there have been examples of more effective delivery in some countries.



#### iCCM

- There has been a decades-long history of programs involving CHWs treating diarrhea with oral rehydration solution and, in some settings, treating malaria with chloroquine. There were important early experiences documenting potential effectiveness of community-based use of antibiotics for treatment of possible pneumonia in Nepal (Pandey 1989) and India (Bang 1990).
- The government of Nepal has implemented a program involving CHW assessment and treatment of diarrhea and possible pneumonia since 1994.
- The current iCCM approach is based in part on IMCI, classifying cases syndromically using clinical algorithms. It can, in some sense, be seen as an extension or further development of IMCI.
- It has been receiving substantial funding support from several donors, including USAID and the Canadian DFATD, and has been supported by WHO and UNICEF since the late 2000s.
- There has been significant concurrent evaluation with early disappointing evidence on population effective coverage and impact.

#### Box 3: Helping Babies Breathe (HBB)

- In 2010, the American Academy of Pediatrics finalized development of a hands-on, two-day training curriculum on care of the non-breathing newborn, with an algorithm covering clinical care up to and including bag-and-mask ventilation (but not including intubation or the use of drugs) based on the concept of the “Golden Minute.” This was intended particularly for more peripheral level settings and cadres of health workers. The training makes use of an inexpensive training and practice manikin produced by Laerdal (Neonatalie).
- Since 2010, with support from USAID and other donors and technical agencies, there has been large-scale roll-out of HBB training (reported to have been introduced in 77 countries as of November 2015). This initiative has consisted mainly of delivering the two-day training to health professionals and, in many cases, making available Neonatalie practice manikins.
- Evaluations to date draw attention to problems with skill retention, post-training.

Another important case discussed by a number of our interviewees was BRAC...

#### Box 4: BRAC

- By a variety of metrics the world’s largest NGO, BRAC was founded by FH Abed in 1972, in the wake of Bangladesh’s war of independence. It has been involved in a broad range of areas notably economic development (including micro-credit), education, and health.
- In global health, it is known for (among other things) its effective pioneering role in household-level care for childhood diarrhea using oral rehydration treatment.
- Relevant to this paper, it is also known for the emphasis it places on applied research and program learning.

Note that for various reasons, notably to protect interviewee confidentiality, in many instances the interview material presented in this report does not specify which program experience is referred to. We recognize that this decontextualization has the effect of rendering the argument somewhat less coherent and compelling. A further phase of the work is to include a set of case-studies.

In addition to the interviews, a literature review was done on escalation, entrapment and related phenomena, drawing largely on psychology, sociology and management and organizational studies. Important insights from this literature bearing on the questions of interest for this paper are documented in a literature review section found in the Annexes.

## Interview Findings

From the interview material, factors bearing on or influencing the phenomena we are investigating largely map onto the frameworks above. However, in most of the cases discussed in our interviews we see multiple interacting factors contributing to the observed process over time, relating to different components of the framework. We have chosen to present the findings in a way that allows us to look at these interactions by considering in turn several major themes or dynamics that our interviewees report in the cases they discuss, largely mapping along our sequential aim > measure > acknowledge > respond model. The notes at the beginning of each of the sections below reflect the final set of codes used for the interview material.

## What is the Initiative Intended to Achieve?



### Outline:

#### I) Causal models/ program theory: what's the problem, how is the solution expected to work?

Addressing the whole thing.

- The right goal, clarity; operationalizing success
- Measurement as it relates to the causal model—measuring the right things at the right time
  - Problem of simple-to-measure but misleading proxies for overall program performance

#### II) Implicit objectives: side-bets, mixed motives, incentives

- Money (keeping money flowing), including keeping the keeper of the purse-strings happy
- Reputation (and power)
- Relationships (keeping host-government or donor happy), political behavior
- The importance of the *appearance of success* (what's incentivized)/ "success stories," "happy talk," spin/ hype/ touting; unacceptability of failure

#### III) Pressure to scale-up; pacing

#### IV) "Blueprint" strategies (therefore blind, uncritical scale-up; little perceived need for learning/ adaption); fixed/ rigid (vs. responsive) designs; dogmatism; blindness to context and systems

### I) Causal Models

One might assume that major initiatives in global public health necessarily have clearly worked out objectives, specified in terms of expected population-level health impact, and that they are carried forward mobilizing and optimizing efforts for the achievement of those overall goals. However, from their experience with many programmatic initiatives, our interviewees paint a rather more complex picture. One interviewee commented, categorically, that...

"Decisions about the design, implementation, and scale-up of projects and programs are seldom made based on the potential contribution to the health impact. They are made for other reasons."

If not impact, then what other considerations drive planning and decision-making? Although there may be a perfunctory nod in the direction of some population health status goal, in many instances the explicit central goal sits further back along the causal path, and the assumption is made (often uncritically) that achievement of this goal will unproblematically give rise to a desired improvement

in population-level health status. So an important first set of issues to be addressed here is the *causal model* or *program theory* underlying a program initiative. *What is the problem and how is the solution expected to work?* Insufficient work at the conceptualization and planning stage and overly simplistic causal models can put an initiative on a trajectory towards failure.

One interviewee commented on thinking through and making explicit a model for how an effort is expected to produce impact...

"There is almost no example I can think of where an outcome is identified, a causal chain is designed all the way back to the beginning of the causal chain, every link is evidence-based, and every link is combined to achieve a health impact, and the attempt is made to maximize the effectiveness of every link in that causal chain. This causal chain is never designed, developed, understood, or documented. What are the 3 or 4 most important things that together will cause an outcome and in what sequence will those things be most effective? What are the elements that truly produce an impact? Can we discard the rest of it? It forces everybody involved along that causal chain to ask: 'What outcome do we want? What are the various things we need to achieve that outcome?'"

However, as one interviewee pointed out, though we need as solid conceptual models as we can develop, at the same time we need to acknowledge their limited descriptive or predictive power with regard to real-world complexity.

"I think that it's not easy because these interventions are extremely complicated. They have many components and rarely can you control them all. So I would even say: how do we even anticipate that these things will work in the first place? I idealize much less about the things that we can do. I think we have to be much more modest."

As a number of interviewees commented, in many program efforts there is insufficient work done up front to think through and operationalize what success would look like and how a program can be expected to produce such a result. One commented...

"In the absence of any upfront consideration of what would constitute good performance, there was a failure to track the measures that would have allowed for ongoing, mid-term correction."

In designing and launching a program initiative—as several of our interviewees indicated—often a full causal picture is not worked through. Faced by the complex real world, we oversimplify and, in effect, take it on faith that if we do a good job on a particular piece the other parts can be assumed to take care of themselves and impact will be achieved.

"There's little data that can link many of these projects to health outcomes, so sometimes it's a matter of *belief*. For example, with those doing management and leadership work. They so strongly believe that good management and good leadership will lead to good policies; and good policies will lead to good programs; and in good programs they will choose the right interventions; and those interventions will be funded by the right amount of money; and the right amount of money and the right interventions will attract the right staff. You know, this whole chain; they believe in this causal chain. And they believe that the driving factor is proper management and good leadership. And so in the absence of data, that belief is enough to keep some of these programs chugging along. I think that there are a lot of *believers* out there."

One interviewee gave—as a specific example of an inadequately worked out causal model—the case of ARV treatment strategy focusing largely on enrolling new patients...

“Paying really good attention to retention and adherence is super important. It may have been just the emergency status of the funding effort in the beginning—where we really were trying to get X number of patients on treatment—that we didn't pay close enough attention to retention and adherence.”

Keeping the big picture in view is not just a matter of developing a good plan. One interviewee, commenting on how implementers lose sight of the overall objective of the efforts they're involved with, said:

“You need to *stay focused* on what you're trying to achieve ultimately because it is very easy to get distracted by what you're doing every day. You forget that day-to-day work is not what you are trying to achieve. It's supposed to get you to what you are really trying to achieve but it becomes very easy to forget that it's not *necessarily* getting you there.”

Part of the task of fleshing out the program theory for an initiative is determining **how performance will be measured**. How will we know we're on track? Operationalizing an effective effort to get to impact at scale requires sound and appropriate measurement—but this is often lacking. One interviewee, discussing iCCM, commented ...

“I think we must maintain focus on *population level coverage* because without it, we will never know whether we can expect to have population impact. Coming out of that meeting, some went too far, saying we can't measure impact and we shouldn't be even trying to measure coverage; we should focus on *implementation strength*. No, I don't think so. I think that we need to be attending to all the pieces. Now measuring implementation strength has been an enormous focus of our work with the implementers in all those countries—to have them agree on a core set of implementation strength measures that were not limited to inputs but also looked at intermediate outcomes such as available stocks and drugs, workers in place and trained and supervised, and the utilization of the services. I think there should by now be an understanding that you can't select one piece of the impact model, evaluate that in isolation and think you have the answer.”

An important problem identified by some of our interviewees is the use of simple-to-measure but **misleading proxies for overall program performance**, for example tracking performance of maternal health programs using the proportion of births happening in the presence of an appropriately qualified health worker (“skilled birth attendance”), discussed further in the section below on “blueprint strategies.” Another notable example given was the focus in ARV programs on number of patients enrolled. As one interviewee reported,

“One of the most frustrating issues I've faced has been losing people to follow-up once you've got them in care. When we started our big scale-up activities we focused on the metrics we were asked to by the donor and that was *getting as many people on treatment as possible*. I think we put the first patient on treatment in April of 2004. The project officer I was working with essentially said, ‘Give me a thousand patients on treatment by July.’ That's what he wanted. But very early we learned we were *losing patients*. By losing, I mean they don't come back and you don't know what their status is. The proportion of such patients in the beginning was worryingly high, but we weren't dinged for it by the early reporting metrics because really what the donor wanted was just to get as many people *enrolled* as possible. There wasn't

really much regard for the outcome. I think the expectation was: once they're into care, they're okay. So we sounded some alarms about that but no one seemed to care. It wasn't really that they didn't care, but it wasn't what people were focusing on."

As the interviewee reported, although the problem was evident early on, the emphasis continued to be on enrolling new patients...

"Once it had been demonstrated that it was possible to enroll a large volume of patients, there was eventually a shift to perhaps quality of care and retention and these other things. But it went on *far too long*, and I think we sacrificed quality for scale-up for a long time. There were these super ambitious targets of getting as many people *enrolled* as possible. *That didn't change until many years into the program.*"

## II) Implicit Objectives

It is not only specifying an explicit goal and how an initiative is expected to achieve this goal that determines how it proceeds. This being the real world, there are almost always **other objectives or incentives that are not explicitly acknowledged** which nonetheless can have a very important influence. These various **implicit objectives** can end up taking precedence over the achievement of the *explicit* goals of an initiative. As indicated at the beginning of this section, it is not uncommon that—even though not openly acknowledged—the most important considerations are political. But there are diverse drivers that may take precedence over achievement of the explicit objectives of a program effort. Such implicit objectives can also include: ensuring the continued flow of money, prestige or reputation (both for individuals and organizations), maintaining the appearance of a successful initiative, and cultivating and maintaining relationships (for example between a donor and a host government).

As one interviewee reported... "Ultimately a lot of what they do is not based on effectiveness; it's based on some larger political-economic determination."

Another, commenting on decisions by governments and donors, stated that,

"Some of these programs are driven to scale without regard to whether the program has an impact or not. They might be politically motivated, or they might be pushed by a donor. Like Abstinence Only. The donor comes down and says "OK, we will give you money, but one of the programs has to be abstinence-only. We want you to take that to scale." So attention to impact is not what is driving a program to go to scale."

In many instances, **money** can trump explicit objectives. One interviewee described a common dynamic (but this reflects views expressed in many of the interviews),

"There was a very clear—if unstated—understanding on the part of the donor and the recipient that continued funding was linked to high performance [as reflected in reporting indicators]. That created a huge disincentive to be very transparent. I don't know that you can really blame that on either partner; it's the reality of the way donor funding works. I'm not sure the donors would have said that they in any way are condoning that, but of course the donors need to be able to defend to their stakeholders the investment of the funds they make, so I think it becomes a bit of the Emperor's Clothes roundabout that isn't intentional but that is, in fact, a consequence of *how donor funding is constructed.*"

Money (including livelihoods) can be important at various levels, not only between donors and implementers, as we heard from various interviewees...

"It's not *necessarily* as crass as 'this will keep my job,' but a lot of times *it is*. Programs have to work. No space has been created for nuanced discussions."

"When the money is there to finance training, and when money can be made out of trainings, you're going to get training projects pushed forward as quickly as possible. So donors may create perverse incentives."

Almost all the program initiatives of interest for this study have involved multiple players. And, of course, the explicit objective of a particular program effort is only one consideration for decision-making related to that initiative. To varying degrees, maintaining and cultivating **relationships between involved partners** can also be an important driver. Many of our interviewees commented on the role this plays in the management and evolution of program efforts. As one interviewee reported (discussing a program effort that a donor continued to support despite evidence of poor effectiveness)...

"Donors are in a difficult situation when they're working with a host government that's invested in a certain model or approach to things and may be reluctant—even in the face of evidence—to change what they're doing. Donors face a tough situation because their role as partners is only partially dependent on doing the right thing (laughs). It's also dependent on the political optics of what they are doing and on the larger political and economic relationships. I think the fundamental problem comes from the overall architecture of these relationships. We need a much more intensive effort to *create public information*, more public accountability. In cases like this, change has to come from the government side; it can't come from the donor side."

Another interviewee, on a similar situation, commented that...

"The external partners are well aware of these problems with performance of the program but have faced a difficult political problem: they are very reluctant to be seen as *not* supporting the program. And they don't have either the right answers or enough influence over adopting those answers to have fundamentally changed the program."

Another interviewee, commenting on a different country context, reported that though donors may have leverage, they don't necessarily use it:

"So the donor has to say that if you don't fix this, we're not going to continue funding it. And I'm pretty sure they'll fix it in 6 months. The problem is that donors want to stay goodie-goodie with the government."

Prioritizing maintenance of relationships over program effectiveness is also seen *within* organizations...

"There was a delicacy about appearing to contradict what was being reported by supervisors; there was this snowball effect where the further it went along the less anybody seemed to be able to speak up. To have questioned what was being reported would have been very threatening and it would have undermined certain hierarchies that were in place."

An important problem highlighted by most of our interviewees is that in many instances **maintaining the appearance of success** is, in reality, a fundamentally important goal, and can draw serious attention away from important performance problems.

“Because USAID must report to Congress, in my experience, the culture there is we want and must have positive results and it is not helpful, it is threatening to us, to have results that are less than positive.”

“With the government wanting to showcase a success, the international NGO wanting to showcase a success, all the donors wanting to showcase results, and an international technical agency writing this up as one of their 10 or 15 global major, successful, innovative approaches to newborn care, there was a lot of hype about it.”

“I talked to a key person involved in the initial work, and when one of the top-level people in this effort who was pushing more of the global shiny roll-out, came and said, ‘This is what we are going to do in [country X],’ and described a training/equipment drive approach. This guy who had been part of the initial work said, ‘I told the person behind the push that the way you’re talking about doing it is not the way we did it and it’s not going to work that way.’ And lo and behold a donor partner spent, I think, \$500,000 on this huge, fancy evaluation of this training and equipment roll-out and there was *no difference*. The warning signs were there. So I think people who ignore those kinds of warning signs are probably not going to change direction too much. I think they are in love with the spin.”

“But if we go back to the problem of people lying about programs, one of the reasons is because scale-up has become a buzzword. Therefore everything has to be described as scale-up.”

Describing a large multi-country effort in which maintaining the appearance of success seems to have taken precedence over other considerations, including actual performance, one interviewee reported...

“This seems to me a clear example of at least public expression on websites and things like that, of apparently enormously successful programs that just don’t bear any kind of analysis. In this initiative, to their credit they were funding the evaluation and their package of interventions was shown not resulting in change. Yet their last annual report still contains glowing language on what a great intervention this is and that’s mendacious. So that desire to put a fantastically positive spin on something is very important.”

### III) “Scaling Up”: Speed and the Expectation that Things Need to Keep Cranking

A common theme in our interviews was the **pace of scale-up**. Not infrequently, there is an expectation with major program initiatives that they be rolled out very quickly. This can be driven by concern about the appearance of success, by political or funding timetables, and by other factors. Pressure to rapidly scale can undermine interest in checking to see how things are actually performing and willingness to adapt to improve performance.

These problems were flagged in several of the interviews...

“The pressures to scale up sometimes come from the donors and sometimes from the government. And they can be very, very strong. The calendar of a political agenda may not coincide with the calendar of implementation. Some things take time. And so if you go too fast because of this pressure you jump some steps. And it is very courageous of public health



professionals to slow down.... People think scale-up happens within a time horizon of 3-5 years which is a funding cycle, isn't it? To me that's just a joke. By the time you get it a bit right you have to go out and people change strategy already. So health systems evolve if you study them at larger scales. They evolve over time so they need 10 years, or terms of 5-10 years, which is little. All of our dreams of rapid scale-up ignore that."

"There needs to be a much longer time frame. And acknowledgement and recognition that in order to meaningfully measure "performance impact" or coverage, you're really looking at a 5-7 year time frame to build that capacity in a country with a really weak information system."

"I asked our project, 'why don't you take just one local government area, of 200-300,000 people, and prove that you can reach reasonable rates of immunization in that area, and see what it takes? OK, so see what level of effort is necessary.' But the answer came back—and it was a quite reasonable answer—and it was, 'why are we doing things on a small scale? Nobody cares about that. If you can't do it on a big scale it doesn't count.'"

"I was recently looking at a community health worker program which is expanding across a whole country. And the advice I gave was: '*Slow down*. Stop where you are and look critically at what you've got. See what modifications you want to make.' —They weren't ignorant about the problems—'Solve those problems first. OK, you're committed to go to the whole country but make a commitment to *doing it right*.' They have 1200 community health workers now and they want to go to 3000+. And I said, 'There are enough problems; they are totally evident now. You need to get it right; don't spend your resources reproducing the wrong thing. I know you've committed yourselves to reaching the whole country in 5 years, but is it better to reach the whole country in 5 years with the wrong program or 7-8 years with the right one?'"

"But donors and agencies are very impatient. They don't want to do that sort of formative research. They don't want to spend time, I think. They have to 'produce results.'"

"Both from the donor and from government, the message was to go as fast as they could. So there was very little attention paid to trying to keep people in care once they were identified as ART-eligible and started on treatment."

#### IV) Fixed Universal Strategies

A common pattern referred to by many of our interviewees can be described in various terms, some examples from the interviews include: "blueprint," "certified," "enshrined," "cookie-cutter," "standardized," "silver bullet" and "fixed" strategies. This entails the elaboration of an approach or strategy which is to be uncritically accepted as a standard for universal implementation. This status can be accorded, for example, through recommendations from global normative bodies like WHO or support by a major donor. Such initiatives tend to give short shrift to contextualization and the need for serious monitoring and iterative adaptation.

"Our insistence that every delivery be attended by a skilled birth attendant, we've pushed that, and pushed that, and pushed that, and the scale is still really low but also there isn't any evidence that when it does go to scale it has any impact but we continue to push it. One of the consequences is that what got dropped, that could potentially have made a difference, was the training of traditional birth attendants."



In a similar vein, another interviewee said:

“There’s been a big effort to promote skilled birth attendants to address maternal mortality. The global indicators for maternal health programs measure this thing called “skilled birth attendance.” But from what we know about the major causes of maternal mortality, a large percentage of skilled birth attendants can’t do much about those problems. So I’m concerned that we’re scaling up an intervention that’s not likely to be effective in reducing maternal mortality. The global policy on skilled birth attendance is—in effect—recommending that women deliver in health facilities where, if they have an obstetric emergency, there is a high probability the facility can’t help them. And when they make a decision to go to a facility that can’t help them relative to one that could, they probably make a fatal decision. If my analysis is correct, I think the problem is that we have an overly simplistic analysis of the technical solution and we—the global community—are busy scaling this up and may discover that you can have very high levels of skilled birth attendance and not much impact on maternal mortality.”

Commenting on two different standardized nutrition program approaches pushed hard by their respective sponsoring organizations, one interviewee said...

“Behind the two promoted approaches lie also two of the major institutions: UNICEF and World Bank. And so this reflects profoundly different views, at a high level (I don’t want to say ideological war). Each of which is strongly able to substantiate their position with evidence. But when you leave it to an independent arbiter to make a decision on what does or doesn’t work, then if the evidence comes, none of them are ready to give up.”

Another interviewee described a dynamic including “universal” solutions and various associated factors inhibiting recognition and acknowledgement of performance problems:

“There is a tremendous appetite to roll out *technologies* (by which I mean not only devices or drugs but also *practices*) which may not always be well adapted to local conditions and circumstances. And to roll them out in very standardized ways. And when that’s done too soon, or with a technology that’s poorly developed, then it’s very difficult to roll back or to change that. And I think that’s a danger with the pressure we have to achieve goals and targets. So that’s one thing. I think there is a certain kind of fixed-mindedness or path dependence on the part of donors and governments when they’re invested in something. There is a high political cost to admitting that it might not be working that well. And especially when it has nice, measurable indicators which everybody is following. So we *achieve our indicators*, but then only find out later what our *outcomes* are. So I think that’s very hard to change. And I think more focus on outcomes, more accountability related to outcomes, and an improved mutual commitment to doing this, not to achieve *outputs* but to achieve *outcomes*. That might be helpful.”

Several interviewees commented on **dogmatism** with regard to strategy or approach as a factor driving program efforts and discouraging questioning about actual impact. An example given by several interviewees was what they viewed as, in effect, a proscription on HIV prevention efforts focusing on risk behaviors, in favor of “treatment-as-prevention.” One interviewee commented that,

“If they were to question the prescribed strategy they would be out on their behinds. Right? They would be out of a job. And there is nowhere to go in global health right now. There are very few places you can go if you don’t tow the top-down party line. Unless you want to be an

academic or unless you have some sort of august position somewhere you can take chances and be courageous.”

Another interviewee reported on a rigorously tested approach found *not* to work. But the donor was wedded to the strategy and pursued it anyway.

“This is a case about mobile phone technology. You know, to tell pregnant women what to eat, when to go to the clinic for their antenatal visits. A colleague designed and got a donor to fund an RCT in his area that he studies which is very poor but they’ve got that great set-up there. They are all set to do RCTs, they can RCT anything. And so they gave out these cell phones and some of the women liked getting messages and so on. But because of: network problems, the fact that the men got jealous when they saw the women getting all these calls, the fact that in order to recharge your phone you had to go into the neighboring town which was like a 4 hour walk and then have to stay the night there—I mean it was kind of hell to charge your phone—it *just didn’t work*. Only about 20% of women on average were getting their messages at the end of the pilot phase. And so my colleague went back to the donor and said, look, let’s redesign this in a different way so that we have a fixed phone line all the women can go to and it will ring for particular women with their messages and whatever. Anyway, it sounded like a practical solution, but basically, the donor wanted this to work *as originally designed* and so they took the grant away from them and gave it to another group which is of course less scientifically inclined and is now promoting it as a huge success.”

One interviewee, gave the following example of imposition of global strategies without regard for local context,

“We all know about the negative cases of community health worker programs in the 80s, when national programs were started in a number of countries and collapsed for all sorts of mainly contextual reasons. But the kind of thinking in many places then was, ‘Well, you know the barefoot doctors succeeded in China, why don’t we do this in all the other countries of the world?’ Again, there were *very specific contextual factors* in China, which sadly no longer exist there. So that’s another example of trying to implement—in that case—a rather complex intervention, *without taking into account rather important contextual factors*. I don’t know yet, but I do worry that iCCM is going to follow the same route. That it is seen as a *magic bullet*.”

One interviewee, discussing the ICDS reported that a different version was “implemented in the state of Tamil Nadu and was very successful. It had a somewhat different implementation model. But the decision was made deliberately—at national level—*not* to adopt that model to scale up into the larger ICDS program for largely political reasons. I think that it was perceived that the Tamil Nadu model was somehow undermining the *prevailing ideology* of the program. The successful Tamil Nadu model was focused specifically on child under-nutrition and targeted those who were malnourished. The ICDS program combines under-nutrition with pre-school education, and aimed to reach all children, so it has a conceptual basis which doesn’t focus only on the nutrition problem but also on child development. Which is all very delightful and nice, but in my view it is *very difficult* to implement successfully.”

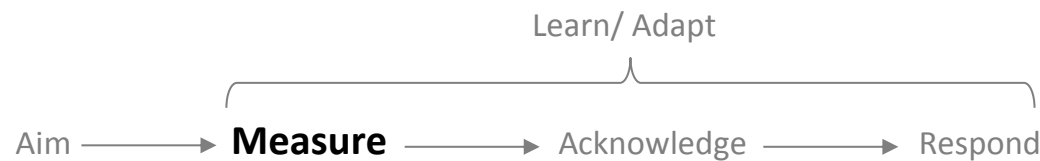
As another interviewee pointed out, more generally, part of the problem with universal blueprint “solutions” is that they generally ignore the complexities of real-world implementation, with all its context dependencies. The interviewee makes the very important point that,

“One of the big problems is that programs are often not designed based on evidence on **how to implement them**. We need to separate *technical interventions* from *delivery strategies*. For the technical interventions such as bed nets and that kind of stuff, of course there is evidence. But *how* you implement them *in a particular context* is more problematic.”

Calling for more local problem-solving and less recourse to overly simplistic, context-free, universal “solutions,” another interviewee commented that:

“We have a wrong perception of what health systems are. There have many moving parts. When we anticipate outcomes or impacts, and we design interventions for health systems at any level, what we envision, create, design, and implement is generally based on linear worldviews. So no wonder we run into trouble; it doesn’t work like this in the real world. We need much more humility with regard to assumptions about how health systems work. We need to admit that we don’t really know how things are going to turn out. It would be very healthy in my view to say, ‘I don’t know what I’m going to get.’ But we’re not allowed to say that in today’s world because we need commit to what we’re going to achieve.”

## Measurement



### Outline:

- I) Serious vs. perfunctory data-use, mere reporting
- II) Measurement difficulty: observability, capacity, problems with bad data, timeliness, data systems
- III) Compartmentalization – therefore too narrow a focus on particular pieces—inputs focus
- IV) Objectivity/ independence

As our interviewees described, in many instances there is relatively little interest in monitoring and data, beyond what is required for reporting. There can be *little incentive to use data to identify problems and find solutions*. Without an adequately specified program theory and associated approach to monitoring along the causal pathway, the focus for monitoring often ends up being on easily measurable *inputs*. This in turn drives program efforts and perceptions of what constitutes success.

An additional challenge is that important dimensions of program performance may be *inherently difficult to measure*. Even where there is more serious desire to understand what's really happening with a program effort, weak systems and capacity limitations can make it more difficult to see and understand what's actually happening.

### I) Serious Data Use

A first requirement for bringing evidence on program performance to bear on decision-making is that there is some **seriousness about use of data** as an input to decision-making. However, as many of our interviewees commented, motivation and capacity to seriously track performance for purposes of improvement (let alone to re-examine assumptions and strategy) is commonly lacking.

“There’s the donor’s need to showcase results. Implementers end up with these conflicting interests; they want impact too but they’re not expected to *report* impact. And so the level of evaluation is not as serious as it needs to be. They can just report districts covered, people trained, dollars spent, and projected numbers of lives saved. That’s a lot sexier than doing a complex evaluation where you actually look at mortality indicators.”

“If you look at these registers you’ll see there are a lot of problems with the data. The data’s not reliable. But part of the problem is that nobody even looks at the data; it is never used, so nobody questions it. If nobody questions it there’s no reason to improve it.”

“People don’t use it or trust it. It doesn’t get analyzed or it doesn’t get analyzed in such a way that allows for action to be taken. So there’s a failure to regularly use program data, to think evaluatively.”

Describing data use in a large CHW program:

“One of their mandates is case management for pneumonia and diarrhea. It’s now been 20 years but nobody’s been looking at it. If you ask the program people how they’re doing, they’ll tell you they’re doing great. So part of the problem is that they have a range of services like prevention, promotion, and curative care. Some of it *is* working. But when they talk about success, it’s a blanket success. They are not looking at details. I can’t pinpoint whose fault it is. But maybe the people who run the program—because there are frequent changes in the leadership—they have no idea about the impact of this intervention. But again, the donors are also at fault because they also don’t pay attention to this.”

Another interviewee, commenting on data use...

“A lot of large programs either are not funded to do good monitoring and evaluation or, for whatever reason, M&E is often merely for reporting purposes and data are not used for identifying problems or finding solutions.”

## II) Measurement Difficulties

Even if those responsible for a program are strongly motivated to look into actual performance, frequently this is not straightforward. As one interviewee explained...

“Measuring how you’re doing is hard. And it doesn’t get any easier over time. Meanwhile, your desire and your mandate to do something is always pushing you. I think often the reason why these situations develop is because you hope you’ve got a solution and you want to implement and scale it as quickly as you can. And you *can* measure certain inputs; you know how hard you are working. *But it’s not easy to measure whether you’re having an impact.* It would take a lot of management effort and a lot of money. So not consciously, but by default, by the push of the events, you end up not measuring effective coverage or impact. And then you’re led down the primrose path.”

Capacity to use data can vary considerably across settings:

“The level of system readiness affects the degree to which you can monitor program performance at scale. For example in the Latin America/ Caribbean region there is a standardized perinatal record, so in some of the work I’ve been involved with in maternal and newborn we were able to monitor at scale in a way that we just never could in Niger or Mali.”

Depending on the nature of the intervention, it can be more difficult or less difficult to observe to what extent it is achieving its intended objective. One interviewee points to neonatal resuscitation (the focus of HBB) as being inherently more observable. As the interviewee explained, this has been helpful in facilitating acknowledgement of a performance problem...

“Because newborn resuscitation is such a discrete activity, part of what has helped this growing recognition of a performance problem is that it has been quite simple to do follow-up assessment after training that demonstrates that provider skills are not maintained. I think that helped a bit—the nature of the clinical intervention; that helped to accelerate that recognition.”

Similarly, another interviewee commented on the comparative observability of contraceptive use, as helpful for tracking program performance.

“Behaviors can be extremely difficult to gauge, to measure. But in this case it’s about contraceptive use and the number of people who use them, who use them correctly, who make the right choice, and ultimately who has contraceptive protection. So it’s a bit more neat and simple.”

Another measurement difficulty that can impede appropriate recognition and response is **lack of timely data** on program performance. One interviewee explained...

“One thing I noticed with the DHS and the MICS surveys and the long delay in getting their data out, was that the state immunization programs were not getting the kind of timely feedback they needed. They knew they were going to get slammed on the household surveys. But then every single quarter all this data is rolling in, the routine data. And the routine data is telling them, ‘Oh, things are getting better.’ We did this, and we did this, and we did this. And some people are really working hard and they can’t see the households from where they are sitting; they can only see that they’re working hard. And the routine data keeps telling them, ‘Yes, yes, yes, you are succeeding.’ And that goes on for years and then you get a survey out that says, ‘We checked a year ago about services that took place two years ago and you weren’t doing good.’ And they say, ‘Well, that survey took place one year ago talking about services that took place two years ago, and we’re doing much better *now*.’ So even the household surveys—which are measuring actual population coverage—are not serving their purpose. Because then you go another 2-3 years down the line and you get another household survey and the pattern repeats itself. The survey isn’t serving its purpose which is to let them know, ‘No, what you are doing *isn’t* working!’ ... With the SMART survey, now they have—about every 6 months—an update on where their state is at and the hope is that this will help people to better keep their eye on the prize.

### III) Compartmentalized Effort

With the important programmatic initiatives in global health of interest for this paper, in all instances there are multiple actors playing important roles; no single actor is responsible for the whole thing. This can result in **compartmentalized efforts** in which individual actors give attention to and are held accountable for only their own small piece (tracked generally in terms of just their particular inputs). Without sound collective oversight and direction (and serious ongoing review of *overall* performance), the effectiveness of the whole effort can be much less than the excellence of the parts. This issue was raised by a number of our interviewees...

“So there are complexities within the government structure, and then the donors themselves are working at different points of intersection with the system. They are working with the government, but coordination is resource intensive. So everybody wants to coordinate but this is difficult. Everyone has to live within their administrative and bureaucratic realities.”

“Let’s take that *essential medicines* case. The job of the essential medicines people is to make that piece work perfectly, regardless of cost, regardless of everything. So you’ll throw all kinds of stuff into that ‘kit system,’ even if the things you are throwing in don’t maximize its real attributable *impact*. Because your job isn’t attributable impact, your job is only to make that link perfect.... Most people who implement these things don’t know they’re not having any

impact because they never have to measure whether they're really contributing to impact or not. Their job is just to push that little piece of the program that's supposed to be combined with all these other pieces that would altogether supposedly have that health impact."

"Often the people we ask to implement these programs aren't the people who can make a change. For example, you would never ask the person running the kit system to change the program if he or she realized it's not having an impact because that's not their job; their job is to implement just their piece of the program. So we don't empower people to make changes in these programs."

#### IV) Independence

An important barrier to learning and adaptation is that we have a natural tendency to believe in what we do; we have difficulty acknowledging that what we're doing may not be effective. Although implementers will generally be more intimately acquainted with the details of the effort they're involved with than are outsiders, their degree of investment can undermine **objectivity** when faced with evidence that things aren't going well. **Confirmation bias** can be powerful both at individual and organizational levels. Sound, balanced measurement and documentation of performance requires objective distance and that, in turn, generally requires some independence—whereby evaluative efforts are undertaken by actors without a significant prior investment in the effort. One interviewee, commenting on accepting supporting evidence but rejecting disconfirming evidence of poor performance, reported...

"You can't just stop it; people will just look for more evidence. They'll say that this evidence is really not conclusive; you need more. If unwelcome evidence comes, you're going to look for more evidence because you have no alternative in place."

A similar point is made by another interviewee... "So the biggest bottleneck was coming back with negative, disconfirming results. When we came with positive, confirmatory results in some areas, those were taken up immediately. Negative results we presented over and over and over, in different ways, and they were, frankly, never taken up until we published a sort of Emperor's New Clothes piece."

Clearly, a critical need for timely recognition and appropriate response when an effort is in fact off track is the *capacity for some critical distance*. But individual psychology and organizational culture can constitute important barriers. As one interviewee commented...

"People believe that what they've spent their careers doing must be effective. And they don't want to ask the question: is it really effective?"

Another (working within a donor agency) said...

"Projects, when they are finishing up—it's all human nature—they all want to put a good spin on things. They'll find somewhere and someway (laughs) to say that either: they succeeded—which is usually what they'll try to say—or, if they failed, it was due to circumstances beyond their control and what they were planning would have worked beautifully. But of course, that's human nature. And it's not to say that people aren't working hard. And it's not to say they're not doing the right things. But they may not be doing *what's necessary to get you where you want to go*.... Another thing I've noticed about people working in programs is that good people, with good ideas, doing good things, doesn't mean you're necessarily getting where you want to go. But if you tell them this is not working, they say, 'Look, I'm a good

person. These are good interventions. I'm working hard. What do you want?' (laughs) And they're right. From their perspective they're succeeding... People don't want to tell you bad news. And projects really, really, really do not want to report bad news. You've got to give these folks a break; they're out there working their tails off trying to implement something. And are they going to tell you, 'Well, we're probably getting only about 10% of what we promised here'? (laughs) No, they're *not* going to say that!"

Other interviewees made similar comments on the challenges of objectivity of evaluations done by those implementing programs...

"Everyone's trying to protect themselves by doing their own evaluations which don't hold water."

"This is actually an issue globally in immunization. There's a chronic tendency to overestimate your coverage. Everywhere I've worked I've seen this and—frankly—when I did a review looking at multiple countries, it pops up over and over. When your immunization rates are low, quite often your routine data is telling you you're getting 20 percentage points more coverage than your household survey shows you. Some of it is falsification. Those at the lower level know they'll get in trouble if they don't use the vaccine that was sent them. So they count up how much vaccine was sent and how many doses that is and, based on that, they tell you roughly how many doses were given. And record keeping is very poor in [country X, a country with poor immunization performance], but frankly even in other countries this was an issue. [Country Y] has a superb immunization program now, but I remember talking to some guys from DfID and the World Bank and they were asking if we could depend on the routine immunization data to tell us the immunization rates for that country. And I looked at it; I looked at 3 DHS surveys and I looked at routine statistics over the same 10 year period. And at the beginning of the period, routine statistics said coverage was around 80%. And the DHS said some much lower number, I don't know, 50%. And in the second survey, the routine statistics said 80%. And at that point the DHS said something like, maybe, 60%. And then at the time of the third survey, routine data was saying 80%. And the DHS agreed, 80%! Full immunization coverage! And so, it was so funny because the guy I was talking to was saying, 'Well, now they match.' And I said, 'Well yes, but if the immunization program were to stop tomorrow and *actual* coverage were to drop back down to 50%, what do you think the routine data would say? It would say 80%! It's *always* saying 80%!' The fact that it's correct now is not because people have now decided not to falsify reports; it's because they don't need to anymore."

## Closing the Loop

Although we can't change hard-wired psychology, as several interviewees comment, there *are* actions that can be taken that increase the likelihood of critical, reflective practice...

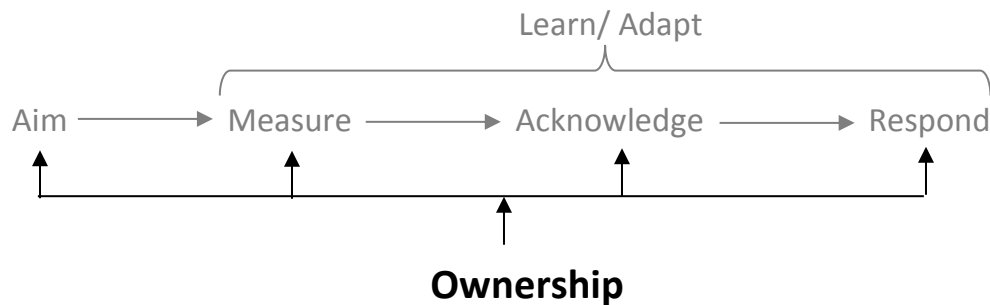
"We need independence of the evaluation group with feedback loops. We call it linked evaluation. There's objectivity because those conducting the evaluations are independent but they also have to be interacting constantly with those implementing the program to be sure they understand what the program is trying to do, to be sure that the intermediate data get back to the program."



It is not *either*: developing a learning mindset within an implementing organization *or* ensuring independent evaluation; it's *both*. As one interviewee involved in BRAC reported,

"We feel there's always learning to do and you have to nurture that culture of learning in the organization if you want to really move forward. So that's what has been guiding a lot of our activities. Right from the beginning we set up a monitoring and research division within the organization. It's an independent unit that does all kinds of research to help programs improve their performance or to do the evaluations that look at their impacts. So the whole idea of learning is there in the mindset of our staff.... Donors should suggest to NGOs or other technical partners, that there be an *internal, independent, research department*, which would be tasked to do evaluation and research and to help programs improve. Here I'm talking about the *improvement of implementation of programs because that's the important thing.*"

## Ownership



The kinds of initiatives considered in this paper tend to involve the following players: a host government (typically a ministry of health), one or more major external donors, technical assistance/ implementation partners (UN technical agencies, international NGOs, consultancy groups, universities) and, often, evaluators (consultants, universities). Behind these immediately involved players, other important actors can include: legislatures or philanthropists that make their funds available to donor agencies, and politicians whose support can drive initiatives forward but whose political needs can also set parameters on what is done and how.

Inevitably, power and control of resources play a very important role in how initiatives unfold. Where a player does not have much influence or control, that player will tend to feel less invested in an effort and less motivated to ensure genuinely good performance. Ownership plays an important role in data production and use; when these functions are undertaken by actors other than those directly involved in delivering and managing a service there may be less than optimal engagement in making use of data for improvement.

As both a potential benefit and a challenge, one interviewee commented that...

“The pressures to scale up sometimes come from the donors, and sometimes they come from the government. And they are very, very strong. Public health people need to recognize that things have a political dimension whether they like it or not.”

When government does *not* feel a sense of ownership, there may be less real interest in actual performance. One interviewee, commenting on a failure by government to take up and build on a successful program experience, explained...

“I think one of the weaknesses of the program—and it was a good program—was that it was seen to be run by foreigners because the World Bank had been involved in it, although it was run entirely by Indians; there was actually not a single expatriate involved in the program.”

In the best case scenario, all partners share a vision and are genuinely committed to realizing it.

“In the Indian case I’ve looked at, the parties that were involved at the national level had an extremely long history and understanding of context, and they were building on previous experiences. So scale-up was part of an effort where everyone knew where they were coming from where they wanted to get to. They were highly aware of potential problems and all committed to overcoming those problems. To be really crude, people gave a \$#@&. They were going to live with the consequences of it. And that’s partly having nationals running it, and

government that deeply cares, and is financing it by the way—almost entirely. But it's that sense of accountability."

With a sense of ownership and engagement within the host government, as one interviewee reported, even a big ship can be turned. The interviewee was discussing the case of an overly narrow focus, in ARV treatment programs, on patient enrollment and an associated relative neglect of retention. He noted that an important turning point occurred when the Ministry of Health technical officer responsible for the program drew the attention of partners to data suggesting there was a serious problem.

"It was brought up by a Ministry of Health technical officer, looking at the reports. If you looked at the data we had from our sites, we were reporting these phenomenally high lost-to-follow-up rates, and we had these electronic data systems. Then you had other sites reporting very low lost-to-follow-up rates, but they didn't have electronic data systems. I just remember going through those data at these technical working group meetings and the guy saying, 'Look, something's wrong here. We have to change the way we're thinking about this, to getting better information from *all* sites. Pay attention to patient care and not just scale-up.' So, again, it may have been a consensus thing, but it happened with the various partners pouring over the data and trying to understand what we were doing and moving out of emergency mode. It had to be owned by the Ministry and the local implementers... If it will be possible to introduce a culture—this has to be done at the leadership level and by somebody in government, really—of collecting data and looking at it in a careful way and everybody doing it together, in the same way, everyone having the same expectations and participating together, if you set your indicators correctly and you monitor correctly, then it's possible to get everybody on the same page. It was only after there was coordination by this technical officer who essentially insisted that we all sit down and look at this."

Several other interviewees also pointed to the importance of **shared ownership** of evaluation....

"So in the one country there was this tone of voice as they talked about "the *independent* evaluation," which meant us vs. them. Whereas in the other country where we were working, there was a lot of trust; we, the evaluation team, did everything *with* them. We know there's a better chance that results will be taken up, received, heard at all, if our implementation partners are engaged in the process of planning the evaluation."

Another interviewer commented...

"To be honest, I think I should also be critical about researchers doing evaluation, because they often don't want to spend that much time. They just want to get data and write papers. [laughs] Sorry to say that, but really, it takes a lot more effort to engage, but that's the only way to provide ownership to countries and implementation partners. That's why I don't like evaluation, where the evaluators stand apart from the implementers. Personally I think it's not that useful. I think what's more important is to work *with* large programs, to really be a partner with a program and basically *do the evaluation jointly so the data gets used.*"

"They said, 'Oh no, it can't be true that the coverage is like 6%; that's not possible.' We were able to overcome that as well but it took a year to convince them that this was not something we were just making up. This was work based on surveys and qualitative studies conducted by third parties so there was no reason to question it. It's best just to face it and to see what you can do about it. The program director said, 'It's easy for you to say that we *need* to change

course but it's very difficult for us actually to change course.' I think there were multiple layers to the problem. The first was they had their own M&E data, and the independent evaluation data and their M&E data were not really in sync. Their M&E data were showing that things *were* happening. One of the good things about findings emerging from the independent evaluation is that they then started looking at quality and validity of their own M&E data. Before that I think it was used mostly for reporting purposes."

Another interviewee described a case in which an agenda shared across government and partners made acknowledgement and appropriate action feasible...

"In Malawi supervision is a big issue for the Health Surveillance Assistants program. The assessment showed supervision was not working. So the Ministry, with the help of INGOs and donors, are trying their best to come up with a feasible supervision system. They've not succeeded yet, but they know it's a big issue and they have to do something about it."

On the same topic, but another case, this interviewee commented...

"So they're starting to create ownership, a dialog of 'Yes, this is part of what we do. This is how we care for our people, and we want to push it this way.' And only then do you start getting the kind of local ownership that will take it forward. So I think that helps create a sense of honesty, not spin."

In a small, foreigner-led project, the director: "would bring government experts and bureaucrats out to the site and they would look at the data together and say what it meant. So, it's important to involve people in the design to the point that key decision makers feel they have ownership even if they're not directly responsible for the program."

While discussing factors associated with real achievement in immunization programs in many settings, one interviewee explained that...

"One of the factors that helped them act on what they were learning was that it was a global priority that was pushed politically very, very hard. There was a lot of political commitment built at the local level so you would actually have people within the Ministry—often it was a combined ministry and donor or development agency group that was connected to the highest leadership. So they had a group, and if there was a problem in X district and the coverage was not increasing very rapidly or they had a couple bad months, the top leadership knew about it and acted on it. And that, in turn, enabled the program staff to act on it."

## Barriers to Acknowledgement and Response



### Outline:

- I) Willingness or ability to respond (entrapment); actionability, cost of reporting/acknowledging bad news
- II) Over-investment
- III) Momentum

### I) Willing and Able to Respond?

In our framework, we have a logical sequence running from *measurement* through *acknowledgement* (and understanding) and finally to *response* or *action*, which—together—we conceptualize as learning and adaptation. But this linear sequencing oversimplifies. As is evident in our interview material, acknowledgement of significant performance problems—even in the face of evidence pointing to such problems—is often impeded because the response that’s called for is unacceptable. So, in the face of inconvenient evidence, a standard response is denial or—as we’ve already heard from one of our interviewees—acceptance of confirmatory evidence (i.e. evidence suggesting things are going fine) but systematically calling into question any disconfirming evidence. Another impediment can be that having identified a performance problem doesn’t necessarily clearly point to the solution.

“I think the bigger problem is really in not looking at the data carefully. That data might be telling you what’s happening and what’s not happening but we often don’t know *why* something is not happening. This ‘why’ question is more problematic. And the documentation of implementation processes is really, mostly non-existent in large programs.”

In many of the instances reported by our interviewees, there was strong resistance to taking corrective action. Even in the face of potentially troubling evidence, action wasn’t taken. Describing conditions in a large national CHW program, one interviewee reported...

“So despite two evaluations showing that the CHWs are not good at diagnosing or treating pneumonia, nothing’s been done. (chuckles) I don’t know why they would have these evaluations done if they’re not prepared to act on their findings. I think it’s just a formality they have to go through—‘We’re funded for this program, we have to do evaluation.’ But what, after that?”

Even in program efforts that otherwise take data fairly seriously and are prepared to take action, there can be unhelpful dynamics at work as one interviewee, who had been involved in immunization programs, reported...

“When something is highly politically charged it can be difficult to acknowledge problems. People faked data like mad. They faked data at the global level and not just at the local level.

Because people's jobs were tied to that. So while there were teams and resources available to collect data, you couldn't really make problems public. And then acknowledging becomes a tricky thing."

Within major agencies, incentives can be so aligned to encourage misuse of data to sustain the image of a successful program. One interviewee described how data were handled on coverage in PMTCT programs...

"So we did a study in four countries. In this one country we found a 16% coverage of services and concurrently with that, there was a PMTCT survey done by a global agency which was praising the country for how well they were doing. They reported that 41% of pregnant women in the country received intervention to prevent MTCT. Now the reason they misestimated is that they were using program reports. You know how there's a chain of evidence: the data come from the facility; they go to the district, to the province, etc. By the time the global agency gets it, you don't know what it is; it's like mortgage-backed securities or something. So it was bad data to begin with and it got rolled up into this national estimate that was threefold higher than what it should have been."

Other interviewees reported similar issues...

"The evaluation found the intervention doesn't make a difference. Nevertheless, it wasn't stopped because people thought, 'You know, that's what we do; that's what we *have* to do.'"

Commenting on a large-scale, multi-country effort, one interviewee reported...

"What happened was, throughout this project, we would be coming with data; we would present it repeatedly. We provided direct feedback to those developing the training. We found there was strong resistance to listening to those data. Very strong resistance to hearing it. First of all we had to present the data multiple times. Multiple. Second, there was an absolute resistance to let these data go beyond the walls of the organization. They didn't want to hear it, they certainly didn't want us to share the information outside. And in fact when we wrote up a paper on the initial assessments, we were forbidden to publish it."

"The global organization pushing this effort did not want these data showing that the program wasn't working to go to the host governments because that would undermine their commitment. And they certainly didn't want it to go to potential donors. We were accused of killing the program. Explicitly killing the program."

"If they're going to sell the government something and get them all keyed up, they don't want to go back and have the government feel they'd been sold a bill of goods. But you can imagine how threatening that could be for the agency pushing this initiative."

Another interviewee pointed out that even with compelling evidence that an effort isn't working, unless there's a readily available alternative, there will be a tendency to keep pressing forward...

"Even in big programs and in large organizations that have a lot of experience implementing programs the underlying dynamic which prevents the M&E people from having influence is that despite language we use about being 'evidence-based,' to respond to such findings can have huge costs. It means that sometimes we have to say, 'Listen, we need to drop this.' But there may be nothing to offer in exchange. For us, there was no alternative and our

organization had put in a lot of time and energy. So in this case it was inertia and the cost of change. It's clear that we have no culture of changing when evidence comes."

Similarly, other interviewees commented...

"Even if they know it's not working, they have no solutions."

"Understanding the nature of the problem and what a good system looks like is not the same as knowing what it will take to get you from one place to another. People are not bad at problem identification, they are not bad at understanding what a good system looks like, but I haven't seen a good skill at identifying what effort is required to get from the problem to the solution and where that effort needs to be placed. So I think *that's* where some of the failures are happening."

Commenting on another multi-country effort, for which an evaluation was done, one interviewee noted diverse possible responses...

"What people did after the evaluation: ... [in country X, they] laughed. So it was like, we did all this hard work and then we get this killer evaluation. So they laughed, but those in [country Y], they cried—literally—in the dissemination, and those in [country Z] stood up and walked out of the room in anger. There, many of them were quite resistant to the negative findings. Immediately they seized on a couple of methodological issues in the surveys to explain away any negative findings because of course—as they saw it—it *was* working. Whereas those in [country Y], a few months after the report came out their view was, 'Well, what did you expect? It was a training and equipment intervention;' they were already buying into the findings that it wasn't working. But [country Z] never really accepted the results."

A related problem is that identifying and trying to fix problems are not incentivized.

"So let's say scaling up is going badly, well who sponsors that diagnosis? Who owns the process of fixing it? It's not simple. It's just not attractive to think about fixing something. It may have had other donor origins or some history that requires repair or recalibration. But you get ahead and get recognized and respected, not by fixing something but by starting something new and by doing something big. That's my sense."

Another interviewee, commenting on patient retention in ARV treatment programs, explained why it is difficult to get programs to give this critical issue the attention it warrants...

"Loss-to-follow-up is a tricky problem *because its solution is very unglamorous*. Essentially it's, 'You guys, go out and find these people.' There are all sorts of technical interventions to try to help with this, SMS's and incentive schemes and everything; these worked to some extent. They nibbled around the edges but, in fact, to really fix it is difficult. It's really very much a local organization kind of thing, which is hard to manage. You get CHWs and other kinds of agents to go out and find the people, track them down, figure out whether they're in care somewhere else. Have they died? If they're not in care, is there any chance of getting them to come back? It's very inefficient. We did a little study back in the beginning of all this where we worked out that in order to get one lost person to come back, you had to launch something like 18 home visits."

When the solution is unpalatable, acknowledging the problem is less likely.

“For those who were losing, there were incentives for them to resist. So understanding the winners and losers is important.”

One interviewee described a situation where it was widely known but downplayed that many CHWs—who were supposed to be resident in the areas they were responsible for—were in fact not resident. But nothing was done.

“Again, it’s the political commitment. I would say the will to make a decision: OK, we will fire all these people. Nobody wants to do that. Also maybe a lot of big people, like the minister of health, maybe somebody from his village had called him up and said, ‘I want my son to be appointed as a community health worker.’ And he gets appointed but he works from the town. So do you annoy the minister?... It’s a big issue. If you have to provide community case management for sick children, you have to have someone present in the community because a child can get sick any time. So the CHW gets appointed to a post for a particular village but ends up based in the hospital. So then that community gets underserved. The Ministry has known it for the last 3-4 years, donors know it, but nothing is happening.”

## II) Over-Investment

Normally, an important condition for successful initiatives is getting strong buy-in from key players. But buy-in cuts two ways, as we’ve seen in the earlier section on Ownership. Once committed, it can be hard to de-commit. This issue was raised by several interviewees.

“It was a government program and the government said, ‘Everyone is in.’ And they took—as they should have—the best advice of doctors and nutritionists in the country, who were good people. But once they’d developed a model, they were too invested to change it.”

In describing the consequences of over-investment of a key government official, one interview reported...

“One of those people was this very influential guy who was very senior in the Ministry. And he was getting toward the end of his career and people said he was looking for a legacy to leave. And this initiative seemed to have become something personal, the way he was going to be remembered. And people more cynically said it was also a way he was going to pay for a couple of houses and pay off some of his people.”

As one of the interviewees reported, a strongly politicized program can create incentives to deny evidence suggesting poor performance...

“There was a whole institute assessing it. There was also a very key person in the beginning who stayed with it for probably 30 years as its key intellectual advisor, a very good man who was totally invested in this. And he couldn’t take any criticism of the program; some people even think he cooked the data. Any time the data came out from this institute or anyone else doing evaluation of the program on the inside, they never showed the data to us on the outside. There was a real reluctance to accept some of the fundamental failures.”



*Not* being over-invested can make it possible to look critically at one's own performance. As one interviewee commented on an organization where he had worked...

"Maybe one of the reasons our group was able to acknowledge some of the especially problematic shortcomings we had with the poor coverage and all that, is because that wasn't all we were doing. We were also doing these rigorous evaluations. It was just a part of the culture of our place that we were trying to be—we were fascinated by the data and trying to figure out what it meant. We didn't feel we were being judged solely based on what the performance indicators were showing."

### III) Momentum

But individual investment (or over-investment) is just one of a number of factors raised by our interviewees that can contribute to **building momentum that can make it difficult to change direction**. As mentioned early in this paper, there were several major initiatives mentioned by numerous of our interviewees. Some of these efforts have received hundreds of millions or even billions of dollars of funding. Some have continued for decades. So they constitute very large enterprises. Interviewees used metaphors of stopping locomotives and turning ships...

"I think the reason why they continued it—against the evidence—is that you can't stop such a locomotive; it's too big. You can't just stop it because people will just look for more evidence. They'll say that this negative evidence is really not conclusive; you need more. IMCI, as a locomotive, is too big. You can't stop it because there are so many stakeholders and at that scale—which is a worldwide effort, still—if contrary evidence comes, then you're going to look for more evidence because you have no alternative in place. It's too costly to stop it; you know you can't stop it."

"Dismantling an entire system you've put in place, which is at scale, is so difficult that inertia just carries it on. I think IMCI is a good example. WHO promulgated manuals and training guidelines to every country in the world which were told to implement IMCI. They had trainers and 10-day trainings. The system behind that program was just a monstrous system trying to push IMCI through. And year, after year, after year, it showed no impact. But dismantling that system is really hard; it's very disruptive. I mean people's careers were built on IMCI; 5- and 10-year development strategies were built on IMCI. Governments had budget lines for IMCI and dismantling that was so difficult that inertia kept moving these things forward."

"Once you get some of these programs moving down the track, if they become enshrined in some policy document or some global strategy, or WHO norms and standards, it's really hard to turn the ship. It takes a lot of political effort, a lot of political will."

Similarly, at the country level...

"Once something gets into the health sector development plan, the 5-year health sector plan, it is really impossible to change that. So you're stuck with a strategy or a program for at least 5 years. And some of these things last longer than 5 years. As soon as something becomes policy, you're pretty well stuck trying to go to scale with something that doesn't work; you can't change it because it's policy."

Discussing the case of ICDS in India, several of our interviewees commented...

“When you go to scale in India you’re talking about services to a billion people. It’s so big that you can’t change anything.”

“Once its size grew, the government committed itself to grow so large, so fast, that it became a political necessity to just keep expanding and keep running it rather than trying to improve it. It costs, probably, billions of dollars a year. I mean, it’s a huge investment. I think it’s the biggest investment in the world in terms of child wellbeing. And it falls consistently short of its potential.”

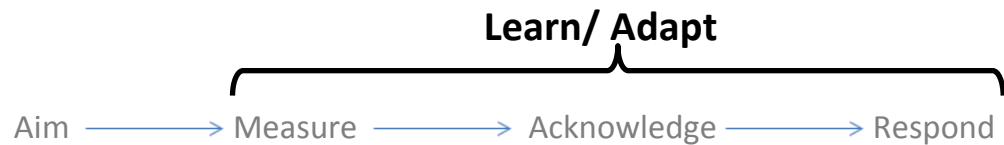
“There’s too much money, too much commitment. When you want to change the weight card, for example, you say, ‘Wait a minute. We have 26 million kids born in India every year. Are you telling me that the 26 million weight cards we print are not the right ones?’ What do we do? I always tried in these programs to say, ‘Look, take a longer view. Start with the kids born from today onward, or from next January onward. It doesn’t have to be done today. And give *them* a new weight card.’”

“I have some thoughts about why it’s so hard to change the program. It has a lot to do with institutional structures set up to implement the program at the national level, at the state level, at the local level. And the fact that this thing is already so scaled up, and has so many well-established operating procedures, that it is very difficult to change it.”

“There’s plenty of data. There are plenty of good analyses of what the problems are. There are plenty of ideas of how to solve those problems and there are examples of efforts to try to solve them. But they’ve not fundamentally changed the way this juggernaut operates.”

“I think one of the problems is that the bureaucratic structures created to deal with the program are insufficient in terms of their strength and their capabilities. And there’s been a reluctance on the part of the government to change that fundamentally. I think because it has a long history of operating in a certain way, it’s very difficult to change those processes and procedures. It would take a really strong momentum to do that. Now you have hundreds of thousands of front-line workers who’ve been doing things a certain way. So the change that’s needed is very dramatic. There are ideas and there are things happening to try to change, but one wonders if they’re sufficient to really solve the problem.”

## Learning and Adaptation



### Outline:

- I) Donor stance/ interaction between donor and implementer; incentives, responsiveness
  - Rigid vs. flexible (open to acknowledge performance problems)
  - Implementers not empowered to respond and make changes enforced single-loop learning, only; disincentivizing transparency; ostrich—head-in-the-sand, don’t-want-to-know
- II) Revisiting strategy based on learning—revalidating goals, assumptions, models; rigid vs. responsive stance
- III) Leadership and culture; organizational values

Across the numerous cases discussed by the interviewees, there was a relatively small group of major funders, consisting of bilateral donors, philanthropic foundations and multilateral donors, with some diversity in how they approach learning/ adaptation and how they relate to their implementing grantees or contractors. A dynamic seen in many of the cases entailed the following:

- a relatively fixed strategy,
- perfunctory “results reporting,” generally focused on inputs,
- incentivizing the appearance of success, and
- comparatively little scope for fundamental learning and adaptation.

### I) Donor-Implementer Relationship

Most of the interviewees pointed to the importance of **donor-implementer interaction**, as a determinant of the extent of learning and adaptation vs. persistence in an initiative when it is failing to deliver.

“One of the things that’s always bothered me—especially about AIDS but about the whole global health phenomenon in general—is that it is so donor-driven. I mean, in Africa donors pay for 80-95% of all the things that get done in AIDS. And for a massive proportion of health budgets as well, if you exclude salaries. That tends to take power away from local actors but also it takes away the intellectual leadership. Often these problems are really complex and only people who understand the culture can begin to think about how to solve them. The best programs have a deep collaboration between outsiders and locals, where the locals can recognize and analyze the problems.”

“In fact, in many projects, if you look at the mandated, required indicators of performance that donors ask for of their projects, they have nothing to do with health impact. They are about process and inputs, so we are not universally or systematically paying attention to impact.”

“We need the donors—including Congress, Parliament—to stop demanding short-term, bar-graph change with everybody having to report back and to look good or their money stops.

No. We need longer funding periods. We need *a culture of evaluation where learning from experience is a good thing.*"

"In some of these environments, especially in project environments, success is the only outcome allowed."

"Often programs are not funded in a way that really allows for making midterm adjustments."

As one interviewee working in a donor agency reported, structural constraints of the way a funder works can be an impediment to flexibility and effectiveness...

"Quite often, even if you know what needs to be done and have the funds to do it, you're not able to because it doesn't fit within the administrative and bureaucratic requirements of your organization. In most cases you're talking about a one year process, if everything goes well. So you can't respond quickly to new information with new projects. You can try to work with existing projects but that depends on your relationship with the project. From the contractual perspective, the project is supposed to know what they're doing when you give them the contract. (laughs) So the legalisms are as if you're buying a commodity. That doesn't necessarily produce flexibility in decision-making. It's difficult for any bureaucracy to change course. The implementer has a certain kind of staff, a certain investment, and a certain way of doing things. And if you say it's not working, what that implies is that they've got to completely change their staff, completely change their approach, and do things completely differently. They're going to say, 'Well, wait a second. If I have to make an extraordinary change then you better have an extraordinary proof that what you're saying is going to be better.' And often you don't; often you've just got a judgment. So in any case, these bureaucratic realities are important impediments to your ability to take action."

Though implementers tend not to, some are more open in discussing failure with their donors. As one interviewee involved in BRAC reported...

"Even in dealing with our donors we've found that they like it if we're open about our experiences, even if they're failures. They've never shied away when the things we're doing are not successful. What they look for, if we're having some failure, is: *what are we doing to try to overcome that?* In my opinion we're not afraid of telling the truth. 'What do you do about that failure? Are you changing your strategy to achieve a better result?' So long as that mindset is there, I don't think you have to worry."

"Donors would object: 'Oh, you said that you would do *this*.' The implementer replies: 'But we've learned that that's not the way to go!' Then the donor: 'But that's in your project proposal. You said this, this, this, and this.' BRAC is strong enough to say, 'I don't care, we're going to do it this way.' But most other people say, 'Well, you've funded us for 5 years and this is what was in the project paper, so we will do it. We have to do it.'"

## II) Revisiting Assumptions and Strategy, Based on Ongoing Learning

With regard to **remedies** for the widespread dysfunctional patterns described in our interviews, there was virtual unanimity that if we are to have programmatic initiatives in global health that actually achieve important sustained impact, we need to take **learning and adaptation** more seriously. Mere single-loop learning that is only open to minor operational adjustments is not sufficient. Instead, there needs to be recognition that the real world is almost always more messy and unpredictable than our plans anticipate. "No battle plan survives contact with the enemy." Consequently, though

we certainly need plans, we need to sit on them more tentatively, explicitly acknowledging from the beginning that we won't have gotten it all right; therefore we need an ongoing, iterative process of serious review of performance, including of our initial assumptions and goals. Several interviewees discussed this issue. One had this to say:

"I think things could have gone differently, particularly if *from the beginning* there was an explicit agreement among the various stakeholders that they would *expect* there would be ongoing problems and challenges, and that they would actually define *from the beginning* an explicit process, using regular measures of program performance, to together talk very openly and frankly about what was happening. And an understanding that everybody would be accountable and the donor also would have said, *from the beginning*, 'We recognize we're going to set some objectives but we're also going to be committed to *readjusting* workplans and expectations based on learning in real time, as the program is implemented.'"

### III) Leadership and Organizational Culture

The required attitudes and institutional values and norms can only be realized in our efforts through changes in organizational culture, which in turn can only happen if such an effort is championed by the leaders of those organizations.

"In that setting, there was no culture of adjusting to data. There was an evidence base, i.e. data on performance, but no culture of adjusting to it."

Where a learning culture is lacking, getting acknowledgement of problems will be like pulling teeth.

"There is, in my view, a real reluctance to admit results that are less than positive, to report those results, to accept those results, and there is a perception of us versus them, with those bringing evidence of performance problems."

One interviewee pointedly casts the problem in the first person...

"So it would be helpful for us to recognize that *we are in the system; we are part of the problem*. So then, with regard to evidence that we're failing to achieve impact, we should take full responsibility for being the engineers of those failures also. But many of these things that don't seem to work are not killed when the evidence is there because there's no accountability to the people. And so we live in our ivory towers, the programs, despite what we might say about being embedded in, and supportive of, and working with, and participating with communities. We are extractors without any penalty. And so when things don't work, there's no tribunal where we're obliged to say 'Listen, we apologize, we failed, we need to redo this, etc.' And since there's no real sanction, things are allowed to go on because nobody wants to take us on."

And yet it's possible for key actors to develop a learning culture.

"I think a huge factor is the donor's approach and understanding of program implementation and an acknowledgement that negative results are just as useful as positive results. I see different cultures in organizations that I've worked with over the years. Broadly speaking if we talk about donors, Gates has an extremely open and receptive approach to learning by doing. So they welcome negative results, they see this as part of the process, they don't overreact. Bill and Melinda are learners themselves and as long as you not only present areas where implementation is weak but in the same presentation you always say, 'Here's what can be done about it,' I've found that they're very receptive."

Our interviewees had numerous suggestions on how to **create the conditions needed for a more adaptive, learning stance**. They drew on case examples from a number of countries. One interviewee, discussing the development of a nutrition effort in Nigeria, reported on information he'd received from a colleague,

"He said he was glad they didn't have all the money at the beginning because they had time to work out a method that worked. Based on initial promising work at small scale using an NGO approach, they asked themselves, 'How do we get the *government system* to do this?' And he said that they ended up developing an approach that works at the local government level in Nigeria and allows this program to go forward, and it wasn't necessarily what we would have designed if someone had given us a big chunk of money and said 'Go, do it.' But it works. And then when the money came in, they were able to replicate it."

## Iterate Up

Discussing how to improve immunization service in Nigeria: "But could you go down to a few Local Government Areas and try different combinations of interventions, and then you ask, 'What does it actually take to succeed?' And then, 'What's the minimum it actually takes to succeed in terms of outside level of effort?' Sending in a 100 vaccinators who work for you might actually work pretty well, but what's a *reasonable thing that works within the system that exists*? What needs to be fixed in this system? How can you fix it? How to make it happen? And then once you have a model that doesn't break the bank, once you know what it takes to fix the system, then you have some idea what it would take to do it on a bigger scale."

Another interviewer, describing a program experience in India, made a similar point on iterating up:

"How can you change a program that's being implemented in close to 100 districts in India, with some 100 million people? And my response to that was, really, maybe you don't change everywhere all at once, maybe you change in one state, or only in a few districts, and see, 'Can we do a good job?' and try to take that lesson gradually to other districts. That was also difficult."

A positive example is given on the scale-up of contraceptive implants provided by peripheral level health workers.

"The Ministry of Health made a decision that to achieve acceleration in contraception use, the only way was to think dramatically out of the box and to delegate to these health workers—who are not medical people—the responsibility and authorization to insert implants. And that was based on no evidence; it was just a major, daring, policy decision. And it took some time before the various organizations would be able to support them—only a few consistently did—but there was one that committed itself to doing so and had an implementation strategy in place which they'd used to train nurses. The Ministry's expectation was that this should go *fast, fast, fast, fast*; that's how the leadership was. But the program people were able to slow it down a bit to give themselves time to think. For example, they wouldn't say 'yes' unconditionally; so you just buy time by following and helping the Ministry get where it wants but you try to get it on your terms. After some rounds it became clear that things were not really working. The number of women who needed to have the implants removed soon after insertion was not negligible; so removal was a key problem. Infection prevention also. And the evidence was there; it had gone too fast. The evidence was there because people kept track

of the monitoring in the field; the program was professionally monitored. So, very interestingly, there was this whole process of checking the initial assumptions and seeing that the workers weren't able to do the removals and that the program could not continue as it was—it needed to be revised. The trainings had to also include removal. It was extremely interesting because the pressure from the beginning was high to move at a fast pace. But then there was evidence it was not working. If any accidents happened, the whole program could be jeopardized. Indeed there *were* such cases. Then people took time to redesign, to reconfigure, to adjust the training curriculum, the monitoring, and the supervision of implementation. And they adjusted. You see, this is different from the negative cases I mentioned earlier because, here, people were building their very own program. It had no history, it was just recent and we were able to look at the first signs of failure and success. Normally we would do a pilot, but there was no time for that. I see this as an exemplary case where impact was initially threatened, and there were people who were awake and acted very responsibly and positively in response to evidence that the program wasn't working. In this case, in my view, the key factor is that there was a management stance that was inclined to act upon evidence."

A positive example from Tamil Nadu, on a state-level variant of the Indian ICDS program, the Tamil Nadu Integrated Program (TNIP):

"You need to go to scale with a built-in capacity to modify, even though it's difficult. TNIP did the monthly monitoring. And when things weren't working they went out to see what was going on. They found the feeding wasn't working, they changed the food; they changed the way the food was made. It changed constantly. TNIP was run by IAS officers. The IAS officers are the best and brightest of India. The IAS runs the tax system, the mines; they run everything in India. No male IAS officer wanted to be saddled with a child feeding project. But women in the IAS, they saw this as a responsibility of Indian bureaucrats and they had the power to modify it. And they could call on anyone they wanted; there was no question but that they were the boss. And they didn't have a vested interest in the design; they had a *vested interest in the outcome*—the measured outcome. Measurement was key."

Another example, on how BRAC works, for example, in the ongoing development of its child diarrhea work.

"With BRAC, there's the idea that *nothing we ever design is right from the start*. The only way we're going to get it right is to have strong monitoring and evaluation. I don't mean bean counting. I mean monitoring what's happening, people's attitudes, what's changing, what isn't changing, what's not working well. And to have the scope to try other things and to share the results. Always asking, 'What implications does this have for the larger program?' BRAC doesn't scale up until they feel they have a scalable model. And even when they scale up they're willing to change."

Double-loop learning (which entails the possibility of questioning underlying goals and strategies):

"So BRAC was characterized by David Korten years ago as a *learning organization*. He wrote a book on learning organizations. BRAC was one of his best examples. BRAC, from the very beginning, builds monitoring and evaluation into every program. When they go out with a proposal or they get help from a donor, or using their own money, they put 5% of it into

monitoring and evaluation. And with those results that then retools their program. What characterizes their programs is *constant learning and modification*. And that's what makes BRAC work."

Another interviewee, commenting on a BRAC program effort, reported: "The program developed over time; it didn't stay with one single intervention. I feel that this sort of thinking and ongoing program design is not very widespread anymore. Maybe it never was, but I think that this drive to scale up singular interventions through vertical programs where these interventions have their own funding, their own administration, and so on, detracts from thinking ahead, *designing as one learns*."

"They started off with *lobon-gur*, a home-made sugar-salt solution, and then later, they provided ORS in their own packets but locally manufactured. But there's a long process in between. And then they learn that they had to try to prevent diarrhea, so then there was that well construction initiative, and so on. So they kind of accreted to this centerpiece of ORT other interventions important to nutrition and diarrhea. *They adapted and learned and implemented and redesigned*."

Another interviewee who has been involved with BRAC explained that, "we saw that the issues one faces *in scale up* are much different from what you have to address *during the pilot phase*. It's not that what you do in the pilot is what you continue implementing at the scaling stage. You have to constantly look for problems and find solutions."

Our interviewees came to broader, more general conclusions about learning culture.

"So maybe it's having an environment both within government structures and within donor-recipient structures that is really transparent and committed to the *truth*."

As one of our interviewees eloquently stated, there is a need for a shift in mindset about negative results...

"What we need desperately is a *culture of evaluation* where negative results are understood as just as important as positive results. Where organizations are rewarded for showing what's not working and then changing. This has to be a culture change about being in this together and learning. ... What we don't need is having supporting partners—in the face of inconvenient evidence of poor performance—saying, 'It's impossible to evaluate these kinds of programs and we shouldn't try to do it anymore.' That's wrong."

Iterative learning and adapting as one moves progressively towards scale was proposed as part of the solution, but as one interviewee acknowledged, this can stand in tension with urgency to achieve genuine impact quickly.

"There's this tension between using a small scale to work out issues and then large scale. Because if you succeed only on a small scale you have still essentially failed. It's only large scale, in the end, that moves the numbers for a population. Furthermore, success on a small scale is often not replicable. Some of the conditions you were able to achieve when piloting cannot be scaled-up. So it's less informative about what you can do on a big scale than you would hope. So in this case then, the question became, 'Should they step back and find out what it actually takes to succeed in a small area or, since they know what's wrong, should they



keep trying to fix it at scale?’ And that’s something we battled back and forth but none of us were wise enough to know the answer.”

Another interviewee on iterative learning and adaption:

“What we want to see is a process where the scaling up is a phased, evidence-driven, process of replication and *learning about what works*, where the learning is drawn from primary indicators of survival and demographics, and morbidity indicators but also from implementation indicators. Shifting from experiment to scale requires clarifying what the scaling-up process should be. The impatience of donors to go to scale and to invest in countries in big, bold, blunderbuss programs, results in poorly-placed investment. *There should be investment in something that stands between a small-scale pilot or experimental study and a nation-wide program, which demonstrates the appropriate strategies and machinery for scale up.* What you hear from this particular donor, for example, is, ‘Oh, we don’t want to do any more experiments. We care about Tanzania.’ Well yeah, but getting from a small pilot to implementation across 120 districts in Tanzania requires strategies, appropriate milestones, management capabilities, leadership machinery, the tools of organization change. And these are not always well thought out or articulated by the people who design the original pilots. Scientists who conduct trials or small pilots aren’t generally thinking about what it takes to change a bureaucracy or what it takes to function at scale. That’s left to the people in the bureaucracy who aren’t usually thinking about how to change things in an evidence-based way. So there’s this great divide between program leaders and the innovators. And bridging that divide, I think, requires a replication and phased learning approach.”

Our interviewees discussed different stances that **leaders** can take with regard to negative results. One described a situation in which a senior government official was tenaciously defending a program, although there were evaluation findings suggesting important performance problems. He described what happened in a workshop where the evaluation was being discussed, the official and various partners were present, and some of the negative findings were presented...

“In the language being used there are various ways to say ‘you,’ depending on how small you want to make the person feel. With the government official’s use of language to a colleague, he went to a very low level, the way you would talk to a young child or an animal. It was awful to watch. It was addressed to this guy from a partner international NGO; he’d presented something suggesting problems with the program. It was quite a brave thing to do in front of this official. The staff person later said that he had to do it because there was another member of his team who was going to do it, but when he saw that this government official was going to be there he took it over because if the other person had done it the official would have killed him.”

One interviewee discussed the critical role that an agency director can play in ensuring appropriate acknowledgement and response in the face of indications that the program wasn't working. But these comments also illustrate that constructive action taken at one point in time doesn't necessarily lead to lasting change in institutional culture.

"The global agency director insisted that those data be brought to the table. She convened a huge meeting of all the program staff and had us present all the data over a whole day. She made a commitment that they were going to act on these data. There were some things we could see; one component was never implemented anywhere and they made a commitment that they were going to change what they were doing. She had me go to the regional meetings with country people. Now the country people of course knew that it wasn't being effectively implemented so they welcomed this. They felt the way these problems had failed to be addressed was emperor's new clothes and they used such language at the time. And they were glad that it came out. However, within other parties at the agency headquarters, this was perceived as killing the program and they were so upset that this would come out and make them look bad that they later created within the health program their own section responsible for evaluation. So never again would they go outside to have an independent evaluation. They evaluated their own."

Nevertheless, our interviewees pointed to the critically important, positive role leaders *can* and *do* play.

"One thing that I think is often overlooked is the culture of the country itself and the leadership in the Ministry of Health, with respect to how evaluations are understood. In Ethiopia for example, there was a learning culture and I think it was strongly promoted by the staff in the country office that evaluation was an important component of a good program, and that all results couldn't be expected to be positive and would instead be used to improve performance."

"The director was not averse to hearing negative news in his program. That's pretty important because you can go and tell him that this is not working and we need to do something about it. It's really important that leaders are willing to hear negative things about their own program. Not many people are bold enough to go tell their superiors."

"There has to be leadership that creates an environment for it to be safe to do this, especially in the beginning. You can cope with some of the usual leader sabotage once you have enough of a critical mass in either a facility or within a district/regional system. But there have to be some sort of champions there to go with this for long enough, to create the environment that allows one to acknowledge and act on serious performance problems."

One interviewee, involved with BRAC reported... "When we see that something's not working, we let our seniors know it."

He also commented that it is understood at BRAC that for all efforts, "Clearly when we do things there will be problems. It can't be all rosy."

Describing effective leadership of CHW program, one interviewee reported...

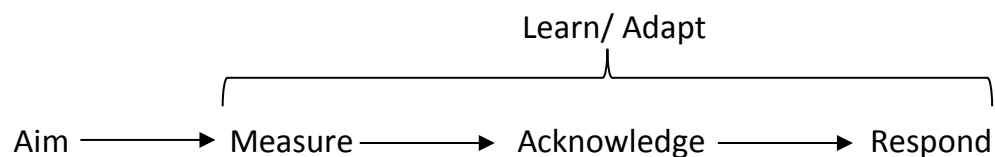
“The program manager is a very committed guy. And he had political support from the minister. He said that, ‘There were a lot of appointments that were not appropriate; I want this to be corrected.’ And what he did was send out people in his program to physically check each and every CHW to see whether they lived in the community or not. They found out that the people who went to do the checking actually signed off on the district health officer’s report that all of them were residents. And the program manager told me, ‘I think this is not right.’ And because I was working in one of the UN agencies he asked, ‘Can you go and check because you’re independent and you don’t have to keep the ministry people happy.’ I went out and found out that 700 of them were not residents; the program monitoring specialists had signed off on them because they were afraid of the district health officer. The manager then acted on this. And he got all that support from the minister at that time. Three important things about this guy: one, he was technically competent, so he understood the program. He was not only a program manager, he understood the interventions. Secondly, he was really a hard worker; he would sit in the office until 10pm, 11pm. And third, he was not averse to hearing negative news about his program. That’s pretty important because you can go and tell him this is not working and we need to do something about it. It’s really important that people are willing to hear negative things about their own programs.”

## Conclusions

Global health efforts have made important contributions to improved population health status. However, in global health it is not uncommon that program efforts proceed, sometimes at very large scale, without serious attention to whether or not they're actually yielding their intended benefits; there is a failure to acknowledge real and significant performance problems and to take appropriate action. Time and resources are squandered and we fail to prevent deaths that otherwise could have been prevented. This paper is intended as part of a broader effort to better understand the remediable drivers of such dysfunctional patterns and to begin to take action to address them.

This paper reports on findings of an inquiry which has included interviews with experienced global health practitioners, supplemented by a review of relevant literature. It has also benefited from informal discussion and feedback from numerous interested colleagues.

From our reading of the interview material and available literature, the key issues map onto a conceptual framework as follows:



By “aim” we mean that effective program efforts require, in the first instance, clarity about what they are intended to accomplish. In global health, there is usually a hope or expectation that a program effort will result in population-level improvement in some relevant health outcome. But clear, careful specification of the goal is a non-trivial challenge. With the moral imperative to do something about the hundreds thousands of people dying of AIDS who could have been returned to good health through ARVs, a program effort was initiated under PEPFAR that aimed to enroll large numbers of patients on treatment, under the banner of “3 by 5” (3 million by 2005). As we hear in the interview material in this report, not enough care went into formulating the goal for this effort. Framed as an *enrollment* challenge, attention was not directed at *adherence* and *retention*, with the result that for years the program effort was much less effective than intended in terms of bringing large numbers back to good health and sustaining this benefit.

Beyond clear, careful specification of the goal of an effort, similarly careful, well-informed thought needs to go into articulating a tentative causal model for how the program effort is expected to give rise to its intended benefit. This, in turn, will inform development of strategy. Necessarily this task amounts to speculation on causal processes in the real world and invariably we oversimplify. As we implement, aspects of the effort will certainly

turn out differently than expected; no plan survives contact with reality, fully intact. That needn't be a problem providing we recognize the divergence and appropriately adjust.

This step of clarifying goals and assumed causal processes of the program effort frequently trips us up and can set us up for failing to recognize when our efforts are, in fact, not delivering. Even when sound goals and causal models are developed, however, there are other ways that our intentions and plans put us on a track of implementation that's effectively blind with regard to whether or not it is achieving its intended benefit. First of all, it is almost always the case that major stakeholders have other objectives that may be more important to them than the explicit goal of the effort. For example, for all categories of players involved in such efforts, maintaining the *appearance of success* tends to be strongly incentivized. Efforts can, in effect, be forbidden to fail. Individual or institutional prestige or power can be very important drivers. Keeping money flowing is often an over-riding priority; this often entails overly simplistic pitches and white-washed results reporting.

Cultivating or maintaining good relations with partners ("higher level political optics") can take precedence over the explicit goals of a program initiative. Of course, implicit objectives are not necessarily a bad thing. The problem arises when they take precedence over explicit goals for improved population well-being. This can dis-incentivize looking very closely at actual program performance.

A common problem in global health (as in other development fields) is universal, blueprint solutions, often labeled as "best practices," which fail to adequately acknowledge context, systems or complexity. As certified, universal "solutions," their implementation is typically done with rapid "scale-up," with uncritical assumptions that, of course, they must be effective—after all, they're "evidence-based." And because they are already consecrated as best practices, monitoring at best addresses operational issues, with little openness to double-loop learning that revisits initial assumptions and design choices.

To be able to recognize and take suitable action once real-world developments start to deviate from expectation, we need ways of tracking or measuring what's happening along our posited causal pathway. That requires serious ongoing attention to try to understand what's actually happening. In many instances this is missing.

In global health, some of the program efforts we are involved with can be characterized as relatively simple, decomposable problems (like Coca Cola logistics) but many are, in fact, complex, interactive problems. Our planning and management need to appropriately respond to the kind of problem we're dealing with. Yes, we need well thought out plans with clear overall population-level goals and well worked-out, contextually-responsive, conceptual models about how we expect the intervention to achieve its intended result. But we need to sit lightly on the boxes and arrows of our conceptual models, revisiting assumptions and design choices in light of what actually

develops on the ground. That said, it's often unclear what's really happening. Serious effort to try to determine this is certainly needed but even then we may be faced by inherently difficult measurement challenges. A more deep-rooted problem is our natural tendency to attend to evidence that confirms what we've committed to and dismiss disconfirming evidence. We can't change hard-wired human nature but we can certainly explicitly recognize this limitation and structure our work in ways that we bring in others who don't have a dog in that fight and bring their perspective to bear in our decision-making.

At the institutional level, organizational norms and culture contribute to dysfunctional patterns of failure to acknowledge important performance problems and take appropriate action. Major players in global health, including multi-lateral, bilateral and philanthropic donors; technical agencies and academic groups; ministries of health; and major international NGOs; can take concrete steps to improve. Some of the needed actions are comparatively simple and straightforward; others may be very difficult.

In this paper, though we have reported on comments by our interviewees and recommendations from the available literature, we have not put forward our own specific recommendations. It is our intention to help facilitate a consultative process over the coming months to develop a concrete plan of action.

## References

- Ackhoff R (1981). "The Art and Science of Mess Management". *Interfaces* 11(1):20-26.
- Andrews M, Pritchett L, & Woolcock M (2012). ["Escaping Capability Traps through Problem-Driven Iterative Adaptation."](#) Harvard Kennedy School Faculty Research Working Paper 299.
- Argyris C (1977). "Double Loop Learning in Organizations." *Harvard Business Review* September
- Argyris C (1982). *Reasoning, Learning and Action*. Jossey-Bass. San Francisco.
- Argyris C (1991). "Teaching Smart People How to Learn." *Harvard Business Review* May.
- Arkes H & Blumer C (1985). "The Psychology of Sunk Cost." *Organizational Behavior and Human Decision Processes* 35: p. 124-140.
- Birkeland S, Murphy-Graham E, Weiss C (2005). "Good reasons for ignoring good evaluation: The case of the drug abuse resistance education (D.A.R.E.) program". *Evaluation & Program Planning*. 28: 247-256
- Brockner J (1992). "The Escalation of Commitment to a Failing Course of Action: Toward Theoretical Progress." *The Academy of Management Review* 17(1): p. 39-61.
- Brown A & Starkey K (2000). "Organizational Identity and Learning: A Psychodynamic Perspective." *The Academy of Management Review* 25(1): p. 102-120.
- Cannon MD & Edmondson AC (2005). Failing to learn and learning to fail. *Long Range Planning*. 38(2):299–319
- Dorner D, Nixon P, & Rosen SD (1990). "The Logic of Failure." *Philosophical Transactions in the Royal Society of London* 327(1241): p. 463-473.
- Drummond H (2012). "Organizational Entrapment," in *Routledge Companion to Organizational Change*. Eds. David M. Boje, Bernard Burnes, John Hassard. London, UK: p. 332-340.
- Easterly W (2006). *The White Man's Burden: Why the West's Efforts to Aid the Rest Have Done So Much Ill and So Little Good*. Oxford University Press. New York.
- Edmondson AC. Learning from failure in health care: frequent opportunities, pervasive barriers. *Qual Saf Health Care* 2004;13(Suppl II):ii3–ii9.
- Freedman L (2011). "Integrating HIV and Maternal Health Services: Will Organizational Culture Clash Sow the Seed of a New and Improved Implementation Practice?" *Journal of Acquired Immune Deficiency Syndrome* 57(S2): p. S80-S82.
- Glouberman S & Zimmerman B (2002). "Complicated and Complex Systems: What Would Successful Reform of Medicare Look Like?" Discussion Paper #8. Commission on the Future of Health Care in Canada. Ottawa.
- Haas P & Haas E (1995). "Learning to Learn: Improving International Governance." *Global Governance* 1(3): p. 255-284.
- Hammond JS, Keeney RL, Raiffa H. *The Hidden Traps in Decision Making*. Harvard Bus Rev. 2003
- Kahneman D & Tversky A (1979). "Prospect theory: An analysis of decision under risk." *Econometrica* 47: pp. 263-291.
- Korten D (1980). "Community Organization and Rural Development: A Learning Process Approach." *Public Administration Review* 40(5): p. 480-511.
- Lynam DR, Milich R, Zimmerman R, et al. Project DARE: no effects at 10-year follow-up. *J Consult Clin Psychol*. 1999 Aug;67(4):590-3.
- Main D & Rambo R (1998). "Avoiding Entrapment." *CPA Journal*.
- McNamara R. *In Retrospect: the Tragedy and Lessons of Vietnam* (1995) Vintage Books, New York
- Morrison E (2011). "Employee Voice Behavior: Integration and Directions for Future Research." *The Academy of Management Annals* 5(1): p. 373-412.
- Pritchett L (2002). "It pays to be ignorant: A simple political economy of rigorous program evaluation." *J Policy Reform*. 5:4, 251-269.

- Pritchett L, Woolcock M. Solutions when the solution is the problem: Arraying the disarray in development. *World Development* 2004 32 (2), 191-212
- Pritchett L, Woolcock M, & Andrews M (2013). "Looking Like a State: Techniques of Persistent Failure in State Capability for Implementation." *J Development Studies* 49(1): p. 1-18.
- Ross J & Staw B (1993). "Organizational Escalation and Exit: Lessons from the Shoreham Nuclear Power Plant." *Academy of Management Journal* 36(4): p. 701-732.
- Schein (2010) (*Organizational Culture and Leadership*, Jossey-Bass, San Francisco (2010) ch17: 299-313
- Scott J (1998). *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed*. Yale University Press: New Haven, CT.
- Senge P (1990). *The Fifth Discipline: The Art and Practice of the Learning Organization*. Doubleday. New York.
- Simon H. The Structure of Ill-Structured Problems. *Artificial Intelligence*. 4 (1973), 181-201
- Simon H. ch5, the Psychology of Administrative Decisions, in *Administrative Behavior*, 4th ed. 1997.
- Staw B (1976). "Knee-Deep in the Big Muddy: A Study of Escalating Commitment to a Chosen Course of Action." *Organizational Behavior and Human Performance* 16: p. 27-44.
- Staw B (1997). "The escalation of commitment: An update and appraisal." In *Organizational Decision Making* p. 192-215.
- Sydow J, Schreyogg G, & Koch J (2009). "Organizational Path Dependence: Opening the Black Box." *Academy of Management Review* 34(3): p. 689-709.
- Tavris C, Aronson E. *Mistakes were Made (but not by me)*. 2007. Harcourt, Orlando FL.
- Tetlock P (1985). "Accountability: the Neglected Social Context of Judgment and Choice." *Research in Organizational Behavior* 7: p. 297-332.
- Tucker A, Edmonson A, & Spear S (2002). "When problem solving prevents organizational learning." *Journal of Organizational Change Management* 15(2): p. 122-137.
- Tversky A & Kahneman D (1974). "Judgment under Uncertainty: Heuristics and Biases." *Science* 185(4157): pp. 1124-1131.
- Weick K & Sutcliffe K (2003). "Hospitals as Cultures of Entrapment: A Re-Analysis of the Bristol Royal Infirmary." *California Management Review* 45(2): p. 73-84.

## Resources on the Web

<http://www.admittingfailure.com/>,  
<https://failforward.org/>,  
<http://valuingvoices.com/face-our-fears-learn-from-failure/>  
<http://www.high-reliability.org/>  
<http://usaidlearninglab.org/>  
<http://doingdevelopmentdifferently.com/>  
<http://www.worldbank.org/reference/GDI/>  
<http://idinsight.org/>  
<http://www.resultsfordevelopment.org/>  
<http://feedbacklabs.org>  
<http://www.torontoevaluation.ca/evaluatingcomplexity/index.html>  
<http://www.kotterinternational.com/the-8-step-process-for-leading-change/>  
[http://www.debonogroup.com/six\\_thinking\\_hats.php](http://www.debonogroup.com/six_thinking_hats.php)  
<https://buildingstatecapability.com>



## Annex I: Literature Review

Although the issues addressed in this paper have not received much attention to date in the global health literature, the phenomenon of organizational efforts getting on a ballistic trajectory where there is great difficulty in redirecting them even in the face of evidence they are not achieving their intended goal has been well studied in other disciplines. In this section, we consider a range of relevant literature—mainly from the fields of psychology, sociology and organizational studies. Part of the intention of this section is to help direct interested readers to pertinent literature they may not otherwise have been aware of.

As has been explained, our particular interest is in factors that contribute to failures to acknowledge and appropriately respond in the face of evidence that an endeavor is not succeeding in its main objective, but instead to persist. And we are certainly also interested in the conditions that favor the opposite—a learning and adaptation stance.

In the interview findings section, we looked—in turn—at: 1) operationalizing success (including causal models), 2) measurement, 3) ownership, 4) barriers to acknowledgement and appropriate response and 5) learning/ adaptation. Our review of the literature largely concerns the final two of these themes.

In the Methods section, we referred to work by Staw and Ross (1987, 1997) which has informed a conceptual framework we have used to organize and make sense of our interview data. Staw and Ross categorize the various determinants of escalation and persistence in a failing course of action under the headings of: *project, psychological, social, organizational, and contextual*

Using Staw's taxonomy, *project* determinants can include costs associated with withdrawal. They note that extrication can be more difficult when project benefits are ambiguous. Important *psychological* factors include *sunk-cost* (or loss aversion), *optimism* and *illusion of control* cognitive biases. Optimism and control biases make people more likely to overestimate the likelihood that positive outcomes will occur and to underestimate the likelihood of negatives. Poor achievement is the fate of others, not oneself, leading to inflated project expectations. In an early paper, entitled “Knee-Deep in the Big Muddy: A Study of Escalating Commitment to a Chosen Course of Action,” Staw (1976) found that individuals continued to invest substantially greater resources when they were *personally responsible* for *negative consequences*. As a potential explanation for why this happens, he proposes *self-justification*—that individuals actively seek to maintain or restore the appearance of rationality to a previously chosen course of action—as the most likely explanation for escalation of commitment to a losing strategy.

Relevant *social* factors include: the desire to justify losses to hostile audiences (saving face); and leadership norms, where there may be social rewards for persistence and converting a losing project into a winner.

*Organizational* determinants include political support within an organization and the level of “side-bets.” Staw also includes among organizational determinants—institutional inertia and organizational values. Inertia can result from imperfect “sensory” systems, breakdowns in internal communications, and difficulties in mobilizing various constituencies, leading to slow responses to signals for change. Therefore, even if the need for change is recognized it may not occur. If it is necessary to alter long-standing policies, rules, or procedures, change can be especially difficult. Organizational values may hinder withdrawal from a losing course of action if a project becomes so

closely tied to the values and purposes of the organization that it is almost unthinkable to consider ending it.

*Contextual* factors include external political forces. Staw and Ross propose that the more external political forces become aligned with the project, the more difficult it will be for the initiating organization to withdraw.

Their theoretical work is based in part on a number of case studies they have conducted, the best known of which is that of the Shoreham Nuclear Power Plant (1993), a hugely money-losing effort characterized by escalating commitment. From these case studies, Staw and Ross have tentatively proposed a temporal sequencing in the importance of the different types of determinant, based on the phase of the effort, which Staw (1997) breaks into three: a *beginning phase* where project economics/data is more likely to drive decision-making; a *second phase* when initial indicators of poor performance are noted, triggering psychological and social determinants for persistence; and a *third phase* where results become so negative that outside parties become aware, triggering organizational and contextual determinants. It should be noted, however, that Staw's aggregate model emphasizes the accumulation and interaction of forces rather than the ordering of effects over time.

Ross and Staw (1993) propose possible solutions:

- 1) Change in top management can reduce psychological and social sources of commitment thus increasing the propensity for withdrawal from a losing course of action
- 2) Efforts to deinstitutionalize a project, or to separate it from the central goals and purposes of an enterprise (or to reframe as an experimental endeavor) can reduce organizational determinants of commitment thus increasing the propensity for withdrawal
- 3) Appeals to supporting organizational constituencies for new support that can change a project's economics so that withdrawal is not so costly

In an important review on theoretical and empirical work on escalation, Brockner (1992) makes the important point that, "complex phenomena such as escalation lend themselves to more than one explanation." As one important potential determinant, he discusses self-justification, which he summarizes as follows: "people do not like to admit that past decisions were incorrect, and what better way to (re)affirm the correctness of those earlier decisions than by becoming even more committed to them?" He proposes, further, that "decision makers may deny or distort the negative feedback they receive concerning their initial resource allocations, in an attempt to convince themselves that things do not look so bad." Brockner posits that the joint presence of two conditions leads to the greatest likelihood of escalation:

- 1) negative feedback concerning the outcomes of the original decision; and
- 2) a high need to justify the correctness of the initial decision.

He concludes that self-justification is one important explanation for escalation *but not the only explanation*. Another important set of drivers can be explained by expectancy theory. "According to such a viewpoint, decision makers assess the probability that additional resource allocations will lead to goal attainment, as well as the value of goal attainment (i.e., rewards minus costs), and thereby generate a subjective expected utility associated with the decision to allocate additional resources."

He offers the following *social* drivers for escalation:

- Whether decisions are made by individuals or *groups* (the concurrence-seeking of groupthink (Janis 1982), a process that tends to shut down independent critical thinking and to accentuate escalation; and see *group polarization theory*),
- Whether or not decision makers have been exposed to the escalation behaviors of relevant *models* (imitation more likely to lead to escalation),
- Whether an audience is observing the decision makers' escalating behavior (*self-presentation theory*), and
- Whether decision makers are competing against another entity in attempting to achieve their goals (competition more likely to lead to escalation).

Brockner also addresses *structural factors* that can perpetuate continuing involvement, drawing on work by Staw and Ross (1987). He points particularly to the role of “side bets” (factors other than the explicit objectives of the initiative that are nevertheless important for involved players) and their importance particularly later in the evolution of an escalating commitment, when they can contribute to a “locking-in effect ... when the initial course of action turns sour.” Such side-bets could include, for example, that offices have been opened up and staff hired.

As Brockner explains (drawing on Staw and Ross), important for self-presentation theory is the culture of the organization. If the culture of the organization is such that people are unwilling to admit failure or that values consistency in behavior, escalation is more likely to occur. If the culture allows people to admit they have erred, or if experimentation is highly valued, escalation is less likely.

## Individual Level

For individual-level or psychological determinants, Staw and Ross draw in part on the work of cognitive psychologists, including Tversky and Kahneman (1974, 1979). In their 1974 paper on Judgment under Uncertainty, Tversky and Kahneman identify heuristics or shortcuts that ease cognitive work but lead to decisions that are automatic and reactive, not optimizing (in the economist's sense). Biases or heuristics they have identified that are relevant to the problems of escalation and entrapment include:

- 1) The *representativeness* heuristic: where people overestimate the extent to which a small sample represents true values (oversimplifying causal models for an initiative);
- 2) The *availability* heuristic: where people preferentially attend to particularly striking, vivid, memorable information and overestimate the significance or relative frequency of such cognitively available information;
- 3) The heuristic of *anchoring*: initially provided information significantly ties down later judgments, limiting adjustments (eg. the information base used to start a program limits the scope of changes program managers will make as the program progresses).

The effect of these biases can be: undue attention to what is most salient, premature closure, and locking down.

As reflected in Staw's summary of determinants of escalation, Tversky and Kahneman have also drawn attention to the importance of *framing effects* in their seminal paper on Prospect Theory (1979). When a situation is framed in terms of holding on to what you've got, losses loom larger than gains. We are strongly loss-averse; it's psychologically hard to cut one's losses or to walk away from sunk costs.

Arkes and Blumer (1985) elaborate on the contribution of *sunk-cost bias* to maladaptive organizational decision-making. They have found that the *desire not to appear wasteful* explains the practice of continuing an endeavor that is not worth continuing. As noted by Staw (1976), Arkes and Blumer also found that *personal responsibility* for the situation tended to increase commitment to a floundering venture. And they found that those with no prior skin in the game—who were not involved in the original decision and investment—were more likely to avoid decisions unduly driven by sunk cost.

Dorner and colleagues (1990), based on experimental work with simulated cases, have investigated what can be seen as faulty problem-solving strategies contributing to failure in complex development scenarios, including:

- 1) Insufficient goal elaboration,
- 2) Insufficient formation of hypotheses about the structure of the system,
- 3) Insufficient ideas about the behavior of the system over time,
- 4) Insufficient coordination of different measures,
- 5) Ballistic action (i.e. with no adjustment in trajectory based on new information), and
- 6) No self-reflection.

With (1), people fail to take into account less conspicuous problems and focus, instead, on fixing only the obvious ones. For (2), people fail to treat a system as a whole and, instead, treat it as an accumulation of disconnected variables that can be manipulated in isolation from each other. For (3), people often have difficulty comprehending the temporal nature of events, particularly if they are nonlinear. With (5), people act “ballistically” when they take actions without checking the effects of these actions later, which leads to (6). Underlying these faulty moves, Dorner suggests that there are 4 main psychological processes:

- 1) Low capacity of humans to consciously process information,
- 2) Tendency to protect one's perception of being competent,
- 3) Preoccupation with *current* problems, and
- 4) Forgetting.

Dorner suggests, as a remedy, *strategic flexibility* in planning and approach (knowing when to do what; knowing what rules to follow under what conditions).

## Social/ Intra-Organizational

Tetlock (1985) addresses factors associated with accountability relationships of individuals *within organizations*, as contributing to irrevocable commitments. Demands for accountability can lead to vigilant information processing to identify the most defensible course of action or—if people have already irrevocably committed themselves to a course of action—it may motivate them instead to develop *rationalizations* for past courses of action. He hypothesizes that all else being equal, people will prefer the “least effort” solution: simply adopting positions likely to gain the favor of those to whom they feel accountable. He refers to this as the *acceptability heuristic*. This dynamic can have a negative influence on program efforts if it encourages people to adopt positions that will please those they are accountable to, as opposed to speaking up about problems (or making difficult decisions to address problems further down the chain). In the same way that political attention can be either a positive or a negative influence on creating effective programs, accountability is not universally good—it depends on how the accountability relationship is established, what the incentives are, and when explanations for actions are sought (prospectively vs. retrospectively).

Morrison (2011) addresses relationships within organizations and the effect this can have on acknowledging and addressing problems under the rubric of “employee voice,” which she defines as the discretionary communication of ideas, suggestions, concerns, or opinions about work-related issues with the intent to improve organizational or unit functioning. This definition implies the need to do something differently or to terminate a current practice, which typically involves risk to those voicing concern as they are challenging the status quo. *Deciding to speak up* is a deliberate decision process that considers both possible positive and negative consequences. As Morrison reports, the *voice* literature has conceptualized two categories of drivers of such decisions: (1) the individual's judgment about whether speaking up is likely to be effective (perceived efficacy); (2) the individual's judgment about the risks or potential negative outcomes associated with speaking up (perceived safety). These have been found to be affected by multiple factors:

- 1) Organizational structure (structures that are low in bureaucracy and that have formal mechanisms (such as grievance procedures) facilitate *voice*, hierarchy stifles voice)
- 2) Organizational culture (such as norms supporting *voice* and the extent to which the culture cares about employee well-being facilitates voice vs. contexts of intimidation or fear)
- 3) Supervisor behavior (trust, openness to *voice*, and approachability support *voice*)
- 4) Employee attitude/disposition (commitment to the job, conscientiousness, and extraversion support *voice*)
- 5) Employee status and performance (full time more likely to *voice* than part time, better-performers more likely to *voice*)

Weick and Sutcliffe (2003) conducted a case-study on the effects of organizational culture on entrapment in a healthcare setting, the Bristol Royal Infirmary. They describe a history of excess deaths which was in effect explained away by hospital staff as due to complexity of the cases. This reconstruction took place because adverse events prompted justifications for performance that preserved the positive identities of surgeons. This change-resistant rationale then persisted because the layers of hospital bureaucracy also benefited from this reading. The consequence was that the whole chain of decision-makers came to support an explanation that made it difficult for an underperforming unit to improve or change course.

The key lesson the authors drew was that autonomy of professional action (in this case, of surgeons) worked against appropriate learning and adaptation. They concluded that,

Interdependence and team accountability and collective responsibility should be emphasized. Hospitals need to create conditions for a learning orientation not a self-preservation orientation; and hospitals should try to weaken the committing context that surrounds adverse events so that people are not forced to justify inadequate performance.

Weick (2001) introduces a number of concepts useful for understanding organizational behavior, relevant to our concerns about dysfunctional escalation. In particular, he draws attention to not only the effect of cognition on action but on the reverse, the effects of action on cognition. *Behavioral commitment* arises from a choice made followed by an irreversible public commitment. Once commitment is in place, the rationale for the choice is strongly reinforced by *post hoc* self-justification. Such alignment of cognition with action taken resolves potential cognitive dissonance.

## Characteristics of the Problem (“project determinants”)

As a way to think about how organizations work, Weick talks about “loosely coupled systems.” In real-world organizations or programs, most of the circumstances we face can best be characterized as loosely rather than tightly coupled systems. Whether loosely or tightly coupled, as Weick points out, systems can be understood as problems in causal inference. In a loosely coupled system, inferences about causal relationships are made more difficult because “relations are intermittent, lagged, dampened, slow, abrupt and mediated.” When a system that is, in fact, loosely coupled is treated as if it were tightly coupled and readily predictable, and ambiguity is denied, we run into problems. As Weick points out, we live with real-world messiness by trying to impose order and meaning and this necessarily entails oversimplifications, which have the potential to trip us up.

Haas and Haas (1995) address implications for successful problem-solving arising from characteristics of the initiative and the specific performance problems. The article is helpful in that it points to the complexity of the issues that often need to be addressed, for example in global health program efforts. They classify problems as: (1) *Non-decomposable*—those for which effective solutions must take account of all linkages; (2) *Partially decomposable*—which have solutions that can ignore some of the links and concentrate on others; and (3) *Fully decomposable* problems—which have more discrete parts, such that they can be amenable to relatively discrete, compartmentalized effort. As with Weick’s loosely-coupled systems, the difficulty arises when one addresses non-decomposable problems as if they were decomposable. Similarly Simon (1973), in his work on artificial intelligence, differentiates between “ill-structured” and “well-structured problems,” and has also concluded that difficulties arise when problems that are in fact ill-structured are treated as if they were well-structured. Ackhoff (1981) makes similar distinctions between puzzles, problems and messes.

Along the same lines, Glouberman and Zimmerman (2002), in a discussion paper on problem-solving in health systems makes a distinction between simple, complicated and complex problems, arguing that problem-solving stances appropriate to merely complicated problems are generally inadequate for dealing with the complex problems that are characteristic of health services.

Simon (1997) introduces other concepts useful for better understanding the phenomena of interest for this paper. He observes that although much behavior within organizations “is intendedly rational, it is only boundedly so,” that is—it seeks good-enough solutions; it *satisfices*. So within organizations, decisions are made within the constraints of limited information, limited cognitive power, and limited time.

In their book *Mistakes were Made (but not by me)*, psychologists Tavis and Aronson pick up some of the points made by Weick, particularly emphasizing the potent effects of self-justification, impervious to disconfirming evidence.

Drummond (2012) identifies major factors contributing to organizational entrapment, which prevent changes in direction in the face of a failing effort, some of which were noted in papers already discussed:

- 1) Biased environmental scanning, where organizations miss cues because the methods used to filter information are biased towards what organizations see as important
- 2) Side-bets, eg. with contracts, over time it may be easier to keep contracting with the same organization rather than form a new relationship or educate a new organization that may offer better services

- 3) Costs of terminating a program: financial, psychological, social or political.
- 4) Passive or default decision-making
- 5) Reluctance to take a risk, change can be risky (or at least change presents unknowns, whereas staying the course is more known)

Drummond suggests five corresponding ways of avoiding entrapment:

- 1) Treat reality as fiction, ie. treat information (reports, data, etc.) as limited, look outside for additional sources of information
- 2) Count the costs of persistence
- 3) Create active decision points, e.g. funding renewed at each stage if the project passes assessment will require decision makers to confront options actively rather than passively drifting along
- 4) Create options, diversify products so if one doesn't work or becomes obsolete, other options are available
- 5) Avoid discounting the future, e.g. don't postpone a necessary investment because of short-term costs

Similarly, Main and Rambo (1998), drawing particularly on Staw, make additional suggestions for avoiding entrapment, as follows:

- 1) Seek disconfirming evidence;
- 2) Consider alternatives explicitly, e.g. specify several different scenarios for possible changes should short-term goals take longer than expected
- 3) Set limits up front on the amount of resources that will be committed
- 4) Reframe decisions to focus on future costs, not sunk costs (saving resources instead of losing resources)
- 5) Evaluate managers on *decision process, not decision outcome*

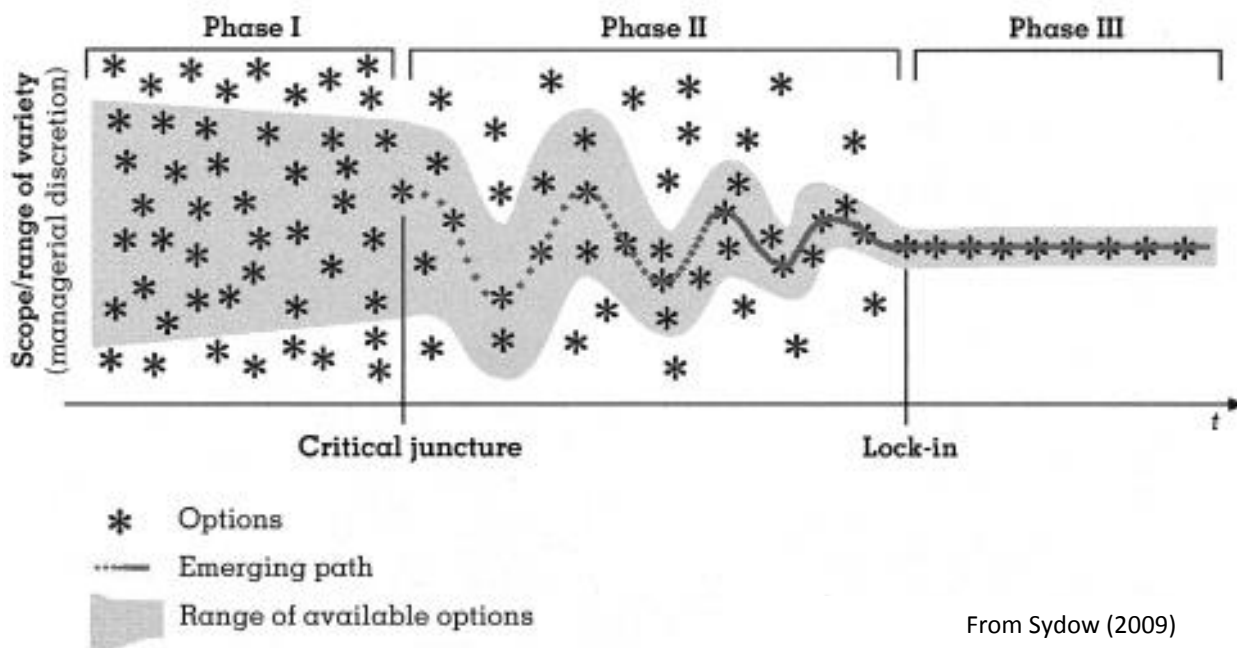
Sydow (2009) conceives of “**path dependence**” in terms similar but not identical to literature on escalation and entrapment. With path dependence, things may start off on a reasonable track but lock-in prevents adjustment when circumstances change and a corresponding change in approach is called for. In what has been described as an escalating commitment the course of action fails from the very beginning (though this may not be evident) and escalation implies an underlying pathological decision behavior based on the dynamics of self-justification and fears of losing face. But initiatives may have been well conceived and initially genuinely successful; with *path dependence* we see a persistent course of action that, at some point, shifts into ineffectiveness or inefficiency.

Sydow describes path dependence as a three-phase process. In the first phase (Pre-formation) there is broad scope for action, but once decisions begin to be made a self-reinforcing process may be established, which the authors call a “critical juncture.” These initial decisions will likely be shaped by organizational culture. In the second phase (Formation), a new regime takes over, a dominant action pattern is likely to emerge, and it becomes progressively more difficult to reverse the initial



choices or pattern of action. The primary driver of this new regime is positive feedback/ increasing returns (although the feedback may be limited to imperfect proxies due to non-observability), and it can become even more irreversible in cases of high investments and/or high fixed costs. In this phase choices are more constrained, but choices are still available. The transition from the second phase to the third (Lock-in) is characterized by a further constriction. In its extreme form, the dominant pattern gains a deterministic character, and alternative courses of action are no longer feasible for various reasons (high switching costs, sunk costs, monopoly, to name a few). By implication, further decisions are no longer actually decisions, but simply moving along a grooved track.

In summary, organizational path dependence can be defined as a rigidified, potentially inefficient or ineffective action pattern built up by the unintended consequences of former decisions and self-reinforcing processes. To break the pattern requires the effective restoration of a choice situation (at least one alternative course of action), although this may not be sufficient. As path dependency is a process built on self-reinforcing dynamics, to break the path will also likely require interrupting the logic/energy of the self-reinforcing patterns. This interruption can be problematic as these dynamics may be hidden (such as through subconscious blinders, perceptual defenses, and blind spots).



From Sydow (2009)

## Moving Beyond the Single Organization

In *Seeing Like a State* (1988), Scott makes the case that legibility, rationalizing and standardizing complex societies into a decipherable and administratively more convenient format is a central problem in statecraft, but that this frequently leads to poor outcomes.

Throughout this book, Scott points to the indispensable role of practical knowledge, informal processes, and improvisation in the face of unpredictability.



Korten (1980) levels a similar critique in his discussion of a “blueprint approach” common in development work that is inimical to learning and adaptation. He draws insights from five case studies (including BRAC) to offer characteristics of “a learning process approach,” which he contrasts with the blueprint approach. He defines the blueprint approach as emphasizing careful pre-planning, the use of pilot projects to develop the plan, and of subsequent *implementation fidelity*. Researchers, planners, and administrators are often separate people or organizations, and are often removed from the target population, and decision-making typically takes place far removed from community reality and day-to-day program functioning. This approach tends to be unresponsive to signals for change, lacking an effective capacity to detect and readiness to respond.

By contrast, in a learning process approach, organizations develop a capacity for embracing error, planning with target communities, and linking knowledge building with action.

Korten describes three characteristic responses to error: *denial*, *externalization*, and *embracing it*. He states that denial often happens in management environments that treat error as personal incompetence, which will encourage people to hide error in order to preserve a positive self-image. Externalization happens in organizations that speak openly of error, but emphasize how little control they have over outcomes. In this way, error becomes impotence—preventing change. Instead, Koren advises that...

To embrace error, organizations must treat it as a vital source of data for making adjustments. Korten further breaks the learning process down into three stages:

- 1) learning to be effective
- 2) learning to be efficient
- 3) learning to expand (relevant to scale-up efforts)

In the first stage, it should be considered normal for error rates and resource investments to be high. Once a program has found a way of effectively responding to a need, it can shift its concern to reducing the input requirements per unit of output. Once acceptable levels of effectiveness and efficiency have been achieved, an orderly phased expansion can occur. He argues that those supporting these different stages will require different approaches to development aid, with Stage 1 being closer to a venture capital commitment of resources, with the expectation that only 10-20% of programs meriting Stage 2 support. Large blocks of funding would not be necessary until Stage 3. Additionally, the demand for significant preplanning, adherence to budget line items, project plans, and schedules would be abandoned, and expected timelines for programs would increase.

In a somewhat similar vein, Pritchett and Woolcock (2003) discuss *the problem of centrally imposed solutions* in global development work. They argue that the primary problem in improving public service provision in developing countries has been the promotion and adoption of the strategy of “skipping straight to Weber” i.e., of seeking to quickly reach service delivery performance goals in developing countries by simply mimicking the organizational forms of developed countries (what they refer to as “Denmark”). When these types of development programs failed, the failure of the inherent design of projects (not creating and/or implementing the program with community input) was not initially recognized. After the first round of failures, failure was attributed to proximate causes and imperfect project design, and so a new round of “better of the same” solutions was

launched (project *intensification*). Only after a second round of failures were these failures recognized as systemic. But the logic of the solution was so seductive to governments and donors that it has taken decades of painful and expensive failures in sector after sector to see that the problem is not just a few mistakes here and there, but that the approach was “fundamentally wrong-headed from top to bottom.” The attachment to this solution was due to several reasons:

- It demonstrably works, it worked for Denmark (confirmation bias, representativeness heuristic).
- This approach made development rational, modern, scientific, technological, and controllable while rendering more legible (sf. Scott) a host of complex problems and solutions (availability heuristic, scientific paradigm).
- Governments adopted this approach because it aligned perfectly with the interests of the donor agencies.
- The governments promoting this strategy may have been sufficiently blinded by their own institutions’ creation myths to lack the historical knowledge and political savvy to successfully create alternatives.

Intensification solutions were proposed by intelligent, motivated, well-meaning professionals who were trying to address very real implementation problems. They considered this response as more appropriate than reform because: (1) even if those involved perceived the need for more fundamental change there was likely a narrow range of alternatives that would be politically feasible; and (2) if existing institutional and organizational forms of service delivery are *de jure* isomorphic to those in Denmark, and Denmark is the desired destination, then fundamental reform is not the obvious response to failure.

For programs to succeed, the authors argue that development professionals need to help *create the conditions under which genuine experiments to discern the most appropriate local solutions to local problems can be nurtured and sustained.*

In more recent work, Andrews, Pritchett, and Woolcock (2012) use the term “isomorphic mimicry” to describe a process of symbolic adoption of the *forms* of programs and structures without necessarily a corresponding concern about *function* (what the programs actually accomplish). This can lead to the continued flow of development resources even if performance has not improved, creating what they call “capability traps” where program/government capabilities stagnate or even decline over a period of time.

Further commenting on the phenomenon of isomorphic mimicry in global health initiatives, Freedman (2011) observes that the “fundamental assumption that function will follow form and that appropriate forms can be identified and promoted through so-called best practices... egged on by donors and experts, good policies are issued, annual workplans are designed, indicators are chosen, training packages are launched. From a global perspective, all looks well and the money continues to flow. But the actual functioning required for impact on health indicators remains elusive.”

As a solution to this problem, Andrews, Pritchett, and Woolcock propose what they call *Problem-Driven Iterative Adaptation*, which entails four main elements:

- 1) Solving particular problems in particular local contexts (instead of imposing universal "best-practice" models)
- 2) Creating an authorizing environment for decision-making that encourages experimentation and positive deviance, i.e. double-loop learning (see below)
- 3) Encouraging active, ongoing, and experimental learning that leads to the iterative feedback of lessons into new solutions (they distinguish this from more traditional M&E that is focused on disbursements, process compliance, and long feedback loops tracking outputs and sometimes outcomes. What they promote: shorter, more timely feedback loops focused on the problem, and experimentation in solving the problem)
- 4) Engaging broad sets of agents to ensure that reforms are viable, legitimate, and relevant

## Learning

Senge (1990) is a widely read source on what is required for learning organizations. He sees organizational learning as beginning with the individual. He uses the phrase "personal mastery" to describe the capacity of those who are able to continually grow in their ability to create the results they seek, and who do so through continual learning. A learning organization is not possible, according to Senge, without people at every level who practice personal mastery. He argues that compulsory training in personal mastery does *not* work. Instead, he advises that:

Leaders of organizations should focus on fostering a climate in which the principles of personal mastery are practiced in daily life. That means building an organization where it is safe for people to create visions, where inquiry and commitment to the truth are the norm, and where challenging the status quo is expected.

Argyris (1977) also addresses the kinds of processes needed within organizations to avoid persistence in unfruitful endeavors. He defines organizational learning as a process of detecting and correcting error. This requires what he describes as **double loop learning**, contrasted it with single loop learning, which is limited to operational adjustments. Double loop learning entails the possibility of questioning underlying goals and strategies. In public health programs, we could see single loop learning as consisting of detecting and responding to errors in program outputs. Double loop learning would be a more comprehensive inquiry into whether or not outputs were contributing to improvements in population health status or whether or not strategies should be changed. Two of the primary factors limiting double loop learning are: 1) institutional culture not supportive of challenging policies and strategies, and 2) norms emphasizing not upsetting colleagues.

In his book, *Reasoning, Learning and Action* (1982), Argyris discusses a number of issues relevant to the interest of this paper. As sources of ineffectiveness he lists: game-playing, distancing, protecting face, undiscussability, and inhibition of production of valid information for problem-solving and decision-making. He talks about the contribution of self-reinforcing covering up of error to dysfunctional escalation. In other work, Argyris (1993) has extended his reflection on organizational learning, using the concepts of *Model I* and *Model II theories-in-use*. Under the more typical but less functional Model I, he proposes that behavior is governed by the following *values*: achieving your intended purpose, maximizing winning and minimizing losing, suppressing negative

feelings, and behaving according to what you consider rational. The main *strategies* employed include: advocating for your position, evaluating the thoughts and actions of others (and your own), and attributing causes for whatever you are trying to understand. These actions require satisficing your governing values, i.e. ensuring attainment of a minimal level of control and winning. Argyris explains that this stance inhibits inquiry into one's positions, evaluations, attributions or tests of them using others' logic. The consequences include: defensiveness, misunderstanding, self-fulfilling and self-sealing processes.

In contrast, Argyris proposes an emphasis on the Model II theory-in-use, the governing values of which are: valid information, informed choice, and vigilant monitoring of implementation to detect and correct error. Embarrassment and threat are engaged not covered up. The result of adoption of this theory-in-use is that anti-learning, defensive routines are minimized and double-loop learning is facilitated.

In another article (1991), Argyris suggests that organizations can teach their staff to reason in a new way, what he refers to as "productive reasoning." *Productive reasoning emphasizes using data to support arguments, constantly testing the inferences that are drawn from data, and using the toughest tests for any conclusions drawn.*

Brown and Starkey (2000) consider individual-level processes needed for organizational learning. They point out that learning entails anxiety-provoking challenges to self-esteem; it requires critical reflection that can trigger ego defenses. They suggest that...

Strategic management helps organizations to prepare for an *essentially unknowable future*. To do so, managers must recognize the complexity of situations, challenge the assumption that the organization already knows all that it needs to know, and generate and *institutionalize constant self-questioning*.

Hammond et al. (2003) point to common pitfalls that can contribute to escalation and entrapment and corresponding possible remedies.

<b>Anchoring:</b> Giving disproportionate weight to the first information you receive.	<ul style="list-style-type: none"> <li>• Pursue other lines of thought in addition to your first one.</li> <li>• Seek information from a variety of people and sources after thinking through the problem on your own.</li> </ul>
<b>Status quo:</b> Favoring alternatives that perpetuate the existing situation	<ul style="list-style-type: none"> <li>• Ask if the status quo really serves your objectives.</li> <li>• Ask if you'd choose the status quo if it weren't the status quo.</li> <li>• Downplay the effort or cost of switching from the status quo.</li> </ul>
<b>Sunk costs</b> [plus self-justification]: Making choices in a way that justifies past, flawed choices	<ul style="list-style-type: none"> <li>• Get views of people who weren't involved in the original decisions.</li> <li>• Remind yourself that even the best managers make mistakes.</li> <li>• Don't encourage failure-fearing.</li> </ul>
<b>Confirming evidence:</b> Seeking information that supports your existing point of view	<ul style="list-style-type: none"> <li>• Check whether you're examining all evidence with equal rigor.</li> <li>• Ask a respected colleague to argue against your potential decision.</li> <li>• Avoid "yes-men."</li> </ul>
<b>Estimating and forecasting</b> [availability bias]: Being overly influenced by vivid memories when estimating	<ul style="list-style-type: none"> <li>• Be very disciplined in forecasting.</li> <li>• Start by considering extremes, and then challenge those extremes.</li> <li>• Get actual statistics, not just impressions.</li> </ul>

brief discussions to be added on:

- Bill Easterly's distinction between "planners" and "searchers."
- Greenhalgh T, Russell J, Ashcroft RE, Parsons W. Why national eHealth programs need dead philosophers: Wittgensteinian reflections on policymakers' reluctance to learn from history. *Milbank Q.* 2011 Dec;89(4):533-63.
- Miller RL, Shinn M. Learning from communities: overcoming difficulties in dissemination of prevention and promotion efforts. *Am J Community Psychol.* 2005 Jun;35(3-4):169-83.
- Ørnenmark C. GPSA Note: 'Learning journeys' for adaptive management – Where does it take us? 2015
- Gustafson, 2003, predictors of successful organizational change
- Heffernan, 2011, Willful Blindness
- Brinkerhoff D, Ingle M. Integrating blueprint and process: a structured flexibility approach to development management. *Pub Adm & Dev* 1989; 9: 487-503.

## Annex II: Interview Guide

### Background

We are investigating circumstances and factors contributing to the *failure to acknowledge and act on important performance problems as health programs continue to scale-up*. Note that we are not interested in planning or implementation failure, per se, but in the failure to acknowledge and address important performance problems. Furthermore, the important performance problems we are particularly interested in have to do with overall effectiveness or impact, rather than delivery of inputs.

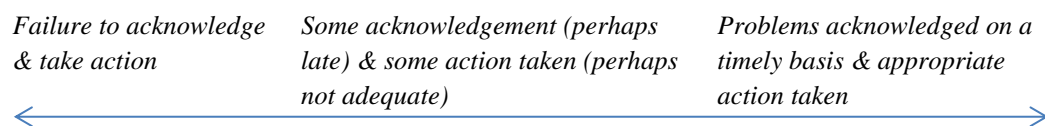
We have approached you because of your substantial experience with program scale-up in global health. In our interview, we will be focusing our discussion on cases lying along a continuum. At one end, we have cases in which scale-up has kept moving forward (this can even be for years), when in all likelihood the initiative was actually not achieving its stated population health objectives, but this was not acknowledged and acted on.

At the other end of the continuum, we have cases in which there was some serious effort made, as scale-up proceeded, to document coverage or impact (i.e. not just inputs) and—when significant performance problems were noted—appropriate action was taken, e.g. suspending the roll-out or significantly modifying the approach, to address the identified problems.

In between these more extreme cases, we will also consider the “muddled middle” where, perhaps at some point, those involved in the effort did start to ask questions, and (even if it was quite late) eventually took action (even if the action was not adequate).

In looking at a range of cases, we hope to identify key factors contributing to continued scale-up of unsuccessful programs, as well as insights into how to remedy this problem. We would also like to identify factors that have contributed to more timely identification of problems and appropriate action taken.

### Continuum of Cases



### Questions:

- 1) Reflecting on your experience in global health—especially program initiatives you’ve had direct contact with—we’d like you to think about particular cases that fall at various points along this continuum. We would like you first to think about particular cases in which, despite there being reason to suspect that the program was, in fact achieving far less impact than intended, there was *persistent failure to acknowledge problems, and scale-up kept chugging forward with no adjustment made to correct course*.
  - a. What specific cases, can you recall that fit this scenario? Please describe in detail.
  - b. What were the factors that encouraged or drove this pattern?  
(Potential prompts: What role did the host government play? The donor? The implementer? What were the key factors that contributed to their failure to recognize, acknowledge or act on important performance problems?)
  - c. What players could have made different choices in this case, to give a different result? What could have facilitated better choices?
  - d. What could be done differently to facilitate timely acknowledgement and action?
  - e. What key lessons should donors, ministries of health, and technical agencies draw from these examples?

- 2) Now for cases on the other end of the spectrum, i.e. more positive examples, we'd like you to think of particular cases in which *important dimensions of performance were somehow monitored or checked, revealing significant performance problems, and—based on these findings—either the effort was suspended or significant changes were made.*
  - a. What specific cases have you seen that fit this scenario? Please describe in detail.
  - b. What were the factors that enabled this pattern?
  - c. What key lessons should donors, ministries of health, and technical agencies draw from these examples?
- 3) Can you think of other cases that lie closer to the middle of this continuum, i.e. in which there may have been some acknowledgement and response to significant performance problems, yet the stance and action of key players was not optimal?
  - a. What specific cases have you seen that fit this scenario? Please describe in detail.
  - b. What were the factors that contributed?
  - c. What key lessons should donors, ministries of health, and technical agencies draw from these examples?
- 4) What general comments do you have on why these problems arise and what can be done to improve this situation? What advice to host governments? To donors? To technical assistance partners?

### Annex III: Initial Coding Used in Text Analysis

- Actor type
  - Ministry of Health
  - Donor
  - Technical agency
    - Implementing partners
    - Evaluators
- Timely recognition—valence
  - Positive
  - Negative
- Appropriate action—valence
  - Positive
  - Negative
- Drivers/ determinants/ inhibitors (from Staw's escalation framework)
  - Project (differences between intervention-based programs and health systems-based programs, path dependency, and any other differences that may shape ability to assess impact or make changes; phase of development—pilot, policy adoption, roll-out, ongoing delivery at scale)
  - Psychological/individual (optimism, illusion of control, self-justification, framing, sunk-cost, confirmation bias)
  - Social (external justification and binding, leadership norms)
  - Intra-organizational (institutional inertia, political consequences, issues of identity)
  - Inter-organizational (theories addressing the interactions between multiple organizations involved in implementing programs, such as dependency on funding, accountability, and trust)
  - Contextual (social, political, and economic factors in the broader environment, such as level of health system functioning, other competing political priorities)
- Possible remedies