

# Actually Navigating by Judgment: Towards a New Paradigm of Donor Accountability Where the Current System Doesn't Work

**Dan Honig**

## Abstract

This working paper explores how donors can move towards greater Navigation by Judgment, highlighting the actions people inside and outside aid agencies can work to make change—encouraging more Navigation by Judgment on the margin, starting today. It focuses principally on accepting the need for a very different way of measuring success and holding projects and personnel accountable, and when and why that might be a very good idea.

Center for Global Development  
2055 L Street NW  
Fifth Floor  
Washington DC 20036  
202-416-4000  
[www.cgdev.org](http://www.cgdev.org)

This work is made available under  
the terms of the Creative Commons  
Attribution-NonCommercial 4.0  
license.

**Actually Navigating by Judgment:  
Towards a New Paradigm of Donor Accountability Where the  
Current System Doesn't Work**

Dan Honig  
Johns Hopkins SAIS  
Center for Global Development

Many thanks to Özsel Beleli, Alice Evans, Gus Greenstein, George Ingram, David Jacobstein, Doug Johnson, Charles Kenny, Rebecka Kitzing-Ivarsson, Chris Roche, Sarah Rose, Jake Shapiro, Matt Winters, Stacey Young, and Janet Vähämäki for their wise comments on previous versions of this paper. I also would like to thank Paul Skidmore for his wise counsel on the English Premier League analogy in section 2. This paper has also benefited from the wisdom of multiple individuals inside multiple aid agencies who wished to remain anonymous. That a number of earnest folks were excited about this paper but did not want to be publicly associated with it (thus having me claim as my own their great ideas, without attribution—which I feel awkward about, but see no way to avoid without excluding their wisdom)—is to me a symptom of the organizational dysfunction described herein.

Dan Honig, 2020. "Actually Navigating by Judgment: Towards a New Paradigm of Donor Accountability Where the Current System Doesn't Work." CGD Policy Paper 169. Washington, DC: Center for Global Development. <https://www.cgdev.org/publication/actually-navigating-judgment-towards-new-paradigm-donor-accountability-where-current>

**Center for Global Development**  
**2055 L Street NW**  
**Washington, DC 20036**

202.416.4000  
(f) 202.416.4050

**www.cgdev.org**

The Center for Global Development works to reduce global poverty and improve lives through innovative economic research that drives better policy and practice by the world's top decision makers. Use and dissemination of this Policy Paper is encouraged; however, reproduced copies may not be used for commercial purposes. Further usage is permitted under the terms of the Creative Commons License.

The views expressed in CGD Policy Papers are those of the authors and should not be attributed to the board of directors, funders of the Center for Global Development, or the authors' respective organizations.

## Contents

Preface.....	i
Introduction .....	0
Imperfect Choices: The Costs of Both Navigation by Judgment and Top-Down Control .....	1
A Misplaced Sense of What it Means to “Succeed” Encourages Aid Agencies Get the Balance Wrong.....	1
The Cost of “Too Much Control” Varies by the answers to “Where” and “What” .....	2
This Pattern Shows Up in Project Ratings Data .....	3
Qualitative Case Comparisons Show the Same Pattern, and More Nuance.....	4
Where the Evidence Leaves Us: Many Projects with Façade Numbers, and Lots of Value Left on the Table.....	5
Section 2. Principles for How to Make Aid Delivery Better, With a Particular Focus on the Sacred Cow of “Accountability” .....	7
Principle 1: Give Up the Delusion of Top-Down Fiat Control .....	8
Principle 2: Accept That Accurate Knowledge of Results is Superior to the Reductive Façade of Inaccurate Metrics—Even if That Reduces “Success Rates” .....	10
Principle 3: Use Different Accountability Systems for Different Tasks, Environments, and People .....	12
Principle 4: Accept that Changing Agency Practice Requires Constant Tending.....	14
Principle 5: Move Towards a Default of Trust Where it Serves Agency Goals.....	18
The Model Agency .....	24
Section 3. Tuesday Morning at My Desk, and What Do I Do?.....	24
You’re Pretty Junior in an Aid Agency .....	25
You’re Pretty Senior in an Aid Agency .....	26
You’re Somewhere in the Broad Middle of an Aid Agency .....	29
You Have Another Role in or Near the Aid World .....	31
You’re Not in the Aid System .....	33
This Coming Tuesday: Let’s Get Moving.....	33
Conclusion.....	34
References.....	39
Annex I. Some Answers to Questions That Might Arise Reading Section 1, if You Haven’t Read the Book.....	42

## Preface

Policy papers don't normally have a preface. This one does, though you should feel perfectly free to ignore it. But I think it's possible the paper that follows will make more sense with some idea of where it, and I, are coming from.

In 2018, I wrote a book. Called *Navigation by Judgment: Why and When Top-Down Control of Foreign Aid Won't Work*, the book is about how aid agencies manage foreign aid projects and argues that tight top-down controls and a focus on target-setting and metrics often lead aid projects astray.

After publication I spent much of the following year running around talking about the book. In addition to academic institutions, I gave talks at many aid agencies, too. Why? Well, I wrote this book in no small part because I hoped it might contribute to improving how the aid system works. I imagined that if I was successful, after a talk there'd be some people in the crowd who'd say to themselves "OK, I didn't think our management system was a problem, but maybe I need to reexamine that view." And maybe if I was *really* lucky there'd be a few people who'd say "I really always felt this was a problem and wanted to change things. This guy's book might help me do this. Let me see what I can do."

So off I went. I spoke at KfW in Frankfurt; I spoke at the UK's Department for International Development; I spoke at the World Bank multiple times, including to the Executive Directors who sit on the World Bank's board; I spoke at the US Agency for International Development and the Millennium Challenge Corporation; I went to Rome and spoke at the International Fund for Agricultural Development; I spoke in Brussels at the European Commission; I spoke to Sida staff and the organization's leaders in Sweden; I spoke in Canberra to Australian Aid (DFAT). I went to the OECD Development Assistance Committee secretariat in Paris and spoke there. I spoke at the Gates Foundation and the Hewlett Foundation. I spoke at the Carnegie Endowment; at London's Overseas Development Institute; at the Center for Global Development in both the US and the UK; to the US' Society for International Development. I spoke at a number of NGOs, and non-traditional funders, and civil society fora and conferences. This is an incomplete list, even if it's already way too long; suffice to say I did a LOT of speaking, basically to anyone involved in international development assistance that would have me. I joked that "if you can get three people and a dog together, I'll show up." The joke (in addition to not being funny, I grant) was maybe not fully accurate; I probably would have turned up for two people and no dog.

While giving all these talks, I learned a number of things that perhaps I ought have expected but hadn't.

First, I learned I was wrong about where the people who worked in aid agencies were starting in their thinking.<sup>1</sup> That is, lots of people—it felt like an overwhelming majority—were already at step 1 of the Alcoholics Anonymous process: they knew their agency had a problem.

If I'm honest, I wasn't so surprised when field staff or project managers came up to me after a talk to tell me that I'd described what felt like their reality and maybe helped them make sense of that reality.<sup>2</sup> I was really glad to hear I'd gotten it right in their view—but that people in the field thought it would be good if they themselves had more power wasn't a total shock.

What truly did surprise me was how often folks in senior management who set the rules and pulled the strings said much the same. These were the “villains” in my story, to the extent my story had a villain, in the sense that they were the folks at “the top” doing the monitoring and controlling. But many of these folks *also* thought the status quo was a problem, in ways that fit my data and diagnosis. One World Bank executive director said to me “I know that when I set targets I don't get what I want. But it's the only way I can move the system at all, the only way I can steer; if I give up that lever, what do I have? Why am I here?”<sup>3</sup> My data demonstrated they had a problem; but what they wanted to know was what the heck to do about the problem.

And that's the second big thing I learned; that for all the ink I spill in the book I don't really get to something that a potential reformer, working inside their agency, can grab on to and use if they want to forward this agenda. Academics have sometimes described my work as “practitioner oriented,” which frankly is not always meant as a compliment; but it turns out you can be “practical” by academic standards but still land far from actually being useful—a practitioner-academic uncanny valley of sorts.

Third, it seemed to me that most of the time I was the person in the room who knew the least about the organization I was speaking at; but it turned out I might know more about the experience of other aid agencies than many others in the same room.<sup>4</sup> It turns out a very small number of aid personnel, even at senior levels, have ever worked in any other agency than their own. So over and over I had the experience of someone saying “you know, we face this unique problem...” and then saying something that—while it of course had unique features—sure had a heck of a lot of echoes of something I'd heard before, at a different agency. This strengthens my belief that there are general things that can be learned, and

---

<sup>1</sup> Or at least those who showed up for a talk given by me, or who I otherwise interacted with. There could be substantial selection bias here, but some of these talks were quite large, so still a fair number of people.

<sup>2</sup> Michael Woolcock once described this experience—practitioners saying that his work had been useful in sensemaking and charting a way forward—as the moment(s) he was most satisfied with his career. I have to agree, it's pretty lovely.

<sup>3</sup> This is a paraphrase. Here and throughout in my reporting of conversations with aid agency officials I identify the comment but not the speaker, in deference to anonymity/the lack of any formal research process (or indeed any intent/plan by me to quote any of these conversations at the time they happened).

<sup>4</sup> The think tanks, and OECD DAC, being notable exceptions.

general lessons for, aid agencies—recognizing at the same time that none will fit perfectly any specific agency.

I wish that I could have done the “book tour” before I had to finalize the book, though I can’t figure out how that would be possible. I learned so much more by being able to talk about this book, including to senior leaders at aid agencies; but I only got to have the conversations because I’d *already* written the book. Hence this paper; my surely inadequate attempt to put forward what I think I’ve learned, and what that might mean for putting Navigation by Judgment into practice. This paper thus draws not just on the data, analysis, and theory in the book but also on the experience of talking to, and learning from, practitioners in a wide variety of organizations. Many thanks to so many folks for their time, openness, and interest; and for the opportunity to continue to learn from and be inspired by the dedicated professionals I meet in aid agencies.

Many thanks for reading,

Dan

## Introduction

As the idiot-wise American baseball great Yogi Berra once said, “If you don’t know where you’re going, you’ll end up someplace else.” This paper is focused on describing the destination: the principles that can guide transformed aid agencies – and the concrete actions each of us could take to move aid agencies towards more Navigation by Judgment.

Navigation by Judgment is an organizational management strategy where projects are driven primarily by the judgment of field staff, rather than directed by central control or by meeting pre-specified targets.<sup>5</sup> The evidence strongly suggests that more Navigation by Judgment, where appropriate, can assist everyone from the World Bank to bilateral donors to foundations to contractors to small NGOs achieve the impact they seek, while making the job of working for these agencies more rewarding and fulfilling in the process.

## Section 1. Diagnosis: Why, and When, More Navigation by Judgment Is a Good Idea

The “Cliffs notes” version of my 2018 book, *Navigation by Judgment: Why and When Top-Down Control of Foreign Aid Won’t Work* is below.<sup>6</sup> If you read the book, you can probably skip this; if you haven’t, and this doesn’t satisfy you, there’s a little bit more in this paper’s Appendix, or please do take a look at the book itself for the fuller version of this argument.

---

<sup>5</sup> It is not the only effort to focus on internal organization by any means - the aid world is currently in the throes of what in my view is a long overdue focus on internal organization (e.g. Doing Development Differently; Problem-Driven Iterative Adaptation; Thinking and Working Politically; Adaptive Management). There are many similarities between Navigation by Judgment and many other efforts at re-examining internal organization and aid delivery. They all either are or facilitate what Nilima Gulrajani and I (Honig & Gulrajani 2017) have called “contingent ways of working”; ways of working that are linked to what organizational behavior and management scholars would call “contingency theory”, a broad range of literature that stresses the need for adaptation and field control where environments are uncertain and/or monitoring is difficult. Navigation by Judgment is focused on the “facilitate”; on how the organization manages its staff, rather than the details of individual projects (though of course how the organization will manage its staff for maximal effectiveness depends on the kind of project and environment, per the description below).

<sup>6</sup> Honig 2018. Cliffs Notes is a company which puts out very short summaries of books students are likely to read in school, and thus are often used by students who haven’t had time/inclination to read the whole book. As a frequent user myself in high school, I think reading the Cliff’s notes gives you the big themes but not the nuance or detail – usually enough for a B or B- on tests at North Farmington High School in the late 90’s (usually the better the teacher the less sufficient the Cliff’s Notes, in my experience). This section is aimed at conveying that similarly broad, “B” understanding. For more detail, in addition to the book itself there’s both academic (Honig 2019a) and non-academic (Honig 2019b) summaries of the argument that may be of interest.

## **Imperfect Choices: The Costs of Both Navigation by Judgment and Top-Down Control**

The question of how to manage—when to control, and when to let go—isn't an easy one for managers, for politicians, for anyone. There are clear pros and cons to both a navigation from the top approach—one where we let those at the “top” drive action—and a Navigation by Judgment approach.

If we navigate from the top, if we use tight top-down controls, we're going to get more management control, more oversight, and more standardized behavior because we'll be able to get dissimilar people in dissimilar places to do similar things using those controls. We're going to get less flexibility and adaptability when we control from the top; it's going to be much harder for those closer to the ground to adjust things along the way. We might get folks to focus on targets where and when we use pay for performance, but they're also going to focus on those targets when they're no longer appropriate and invest less in parts of the job that can't be measured and thus aren't rewarded.

On the other hand, if we Navigate by Judgment, if we empower those closest to the ground, we're going to get more agent initiative. We're going to get more use of soft information – things that agents can see and learn, but can't verify with numbers. We're going to get more flexibility. We might get more intrinsic motivation because people like having autonomy and feeling like they can make a difference in their jobs. But we're going to have rely on fallible agent judgment.

## **A Misplaced Sense of What it Means to “Succeed” Encourages Aid Agencies Get the Balance Wrong**

So OK, it's a goldilocks situation—you can have both “too much” and “too little” control. But aid agencies weren't created last week; if there is kind of a balance to be had, why is that balance not in alignment already?

In part this has to do with the nature of being a public agency and the nature of political authorizing environments. When an agency creates performance targets to guide projects, it does so not just to improve performance, but also to demonstrate performance. That is, those numbers help justify funding, help prove to politicians that they get something for their dollars and yen and renminbi. When numbers are used for legitimacy-seeking, agencies will employ a top-down, metrics-driven approach even when the agency itself doesn't think that approach is useful in helping projects achieve their goals. More insecure agencies - agencies that spend more time justifying themselves to authorizers – are going to be even *more* likely to make the “too much control” error.

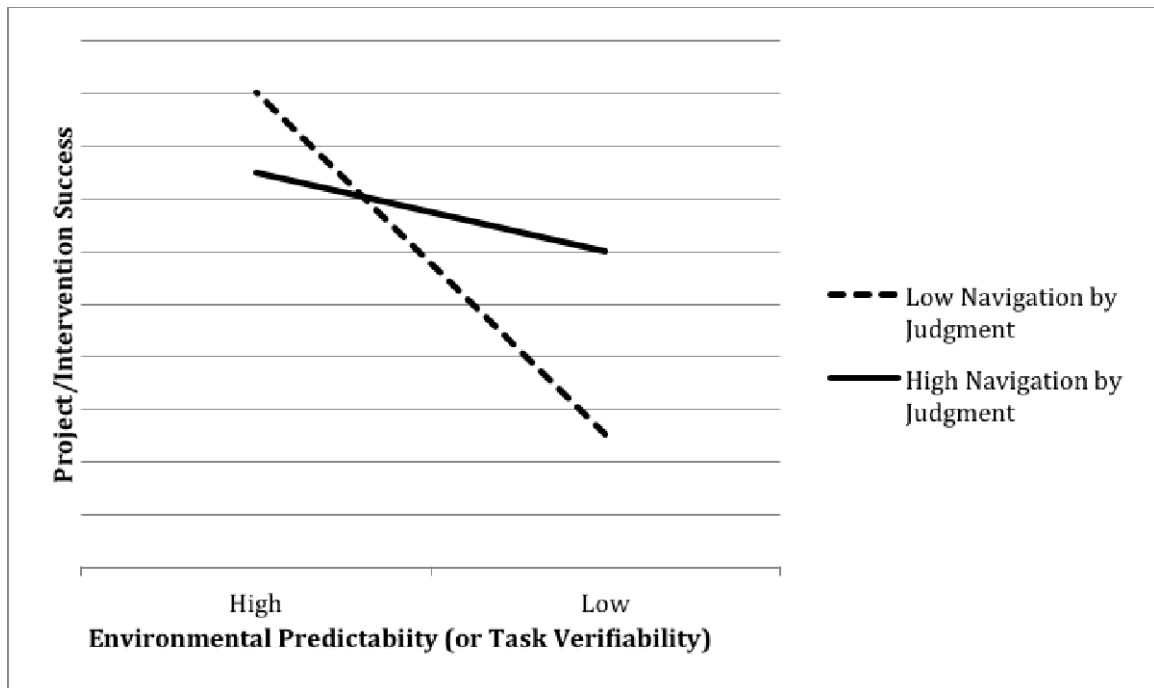


## The Cost of “Too Much Control” Varies by the answers to “Where” and “What”

The returns to Navigation by Judgment (and the costs of too much control) aren't fixed. They depend on the environment and the task - the country you're in and what exactly you're trying to do. The more unpredictable the environment, the harder it is to figure out what's going on using objective and verifiable data (think “building a road” vs. “building a justice system”), the greater the returns are going to be of Navigation by Judgment and the more consequential the error of too much control. The agency that uses more Navigation by Judgment, the agency that puts more control in the field is going to be better able to adjust to those more difficult circumstances. In addition, the less the task can be pre-planned or effectively managed using numbers the higher the returns to Navigation by Judgment. So where aid seeks to deliver clear, countable, verifiable things – e.g. building a road, or dam – Navigation by Judgment is unnecessary, and maybe even costly. But where aid seeks to build institutions (e.g. building a justice system, or strengthening governance), or just deliver things that are difficult to count quickly and accurately (e.g. a quality education), we should expect Navigation by Judgment to be quite useful.

Figure 1 attempts to depict this visually. While Figure 1 is in the book, I made it before I started getting data; it's a graphical description of theory, not results.

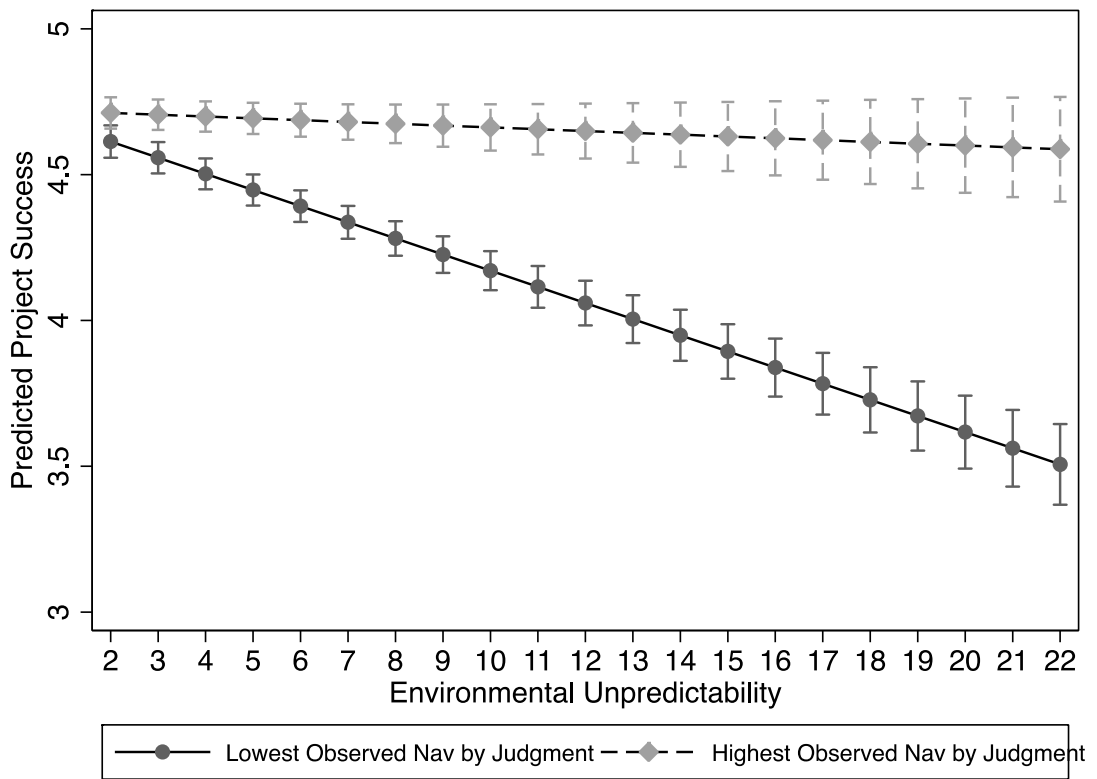
**Figure 1. The theorized relationship between navigation strategy and task/environment**



## This Pattern Shows Up in Project Ratings Data

I set out to see whether these ideas were borne out in the data. For one set of tests I used econometrics—formal quantitative data analysis. I built a data set of 14,000 projects across 9 organizations, across 40 years, across 180 countries to think about the relationship between agency management practice—specifically, how likely an agency is to Navigate by Judgment - and performance. I find (see Figure 2) that agencies that have greater likelihood of navigating by judgment see much less difference in their performance as environments become more unpredictable than do agencies which use more top-down control.<sup>7</sup>

**Figure 2. Econometrics on the interaction between navigation strategy, environmental unpredictability, and predicted project success**

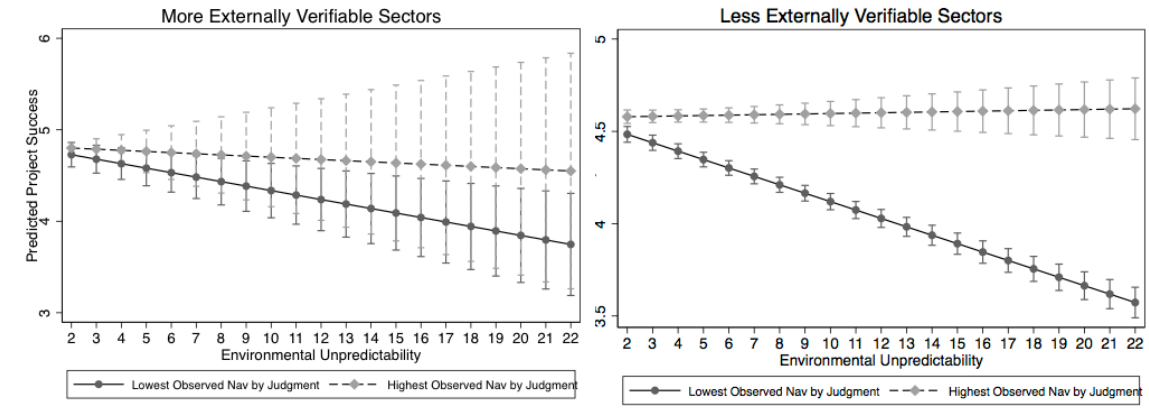


When the road gets tough, those agencies that let the person sitting in the car’s driver’s seat handle the controls are better able to adjust to the rockier road. But those agencies that are

<sup>7</sup> There’s of course a ton of econometric detail sitting behind these figures. In very (too) short, environmental unpredictability is measured using the Polity IV State Fragility Index, and Navigation by Judgment is measured using both measures from the Paris Declaration monitoring surveys and a survey of people who work at developing country aid management units (who can compare donors directly). This figure is drawn from a model with fixed effect by agency (which are collinear to the measure of Navigation by Judgment), so the focus here is on comparative slopes, not intercepts. For more (including full model specifications and details and lots and lots of robustness checks), please see Honig 2018.

trying to control things from afar – that try to steer from a great distance, by remote control - are unable to adjust as effectively to more difficult environments.

**Figure 3. Econometrics on the interaction between navigation strategy, environmental unpredictability, and predicted project success across differentially verifiable sectors**



This pattern—the difference in performance between navigation strategies—is driven by what I call less verifiable sectors, as figure 3 shows. It’s not building the drip irrigation system, it’s agricultural water management, where we see the effect. It’s not in building the road, it’s in administration of the transportation systems. Remote control fails for tasks where we don’t have the ability to collect accurate numbers about what’s going on.

### Qualitative Case Comparisons Show the Same Pattern, and More Nuance

For another set of tests, and to understand what drove the patterns in these data with more nuance, I compared projects which are similar in scope and purpose that are managed differently from one another. I dove deep into eight projects—or four pairs of projects—that vary with regards to how unpredictable is the environment and how verifiable the task. In each pair I compare a project from USAID—a quite constrained agency—and DFID—an agency much more likely to Navigate by Judgment.

In stable environments with verifiable tasks, a tight focus on outcomes is indeed the best strategy. But in three of the four case study pairs (all but the most verifiable task in the most stable environment, as figure 1 would predict), Navigation by Judgment works better. I see evidence of chasing targets and data leading projects astray in most (but not all) settings. But I also see that some of those projects that succeed less look very successful, by focusing on numbers that are unrelated to success. So the *donors themselves* sometimes react as if the more successful project (which is managed in a way such that it reports more accurately) is less successful than the in fact less successful project.

There's more flexibility and adaptability in the more Navigation by Judgment-laden projects. These projects also show the importance for success to incorporate things that field staff perceive but can't verify. There's also more organizational learning, not least about the proper use of judgment. If you are never able to use your judgment, you are never able to figure out where your judgment has led you astray and improve your judgment. Judgment is a muscle; we need to exercise it to make it stronger.

## **Where the Evidence Leaves Us: Many Projects with Façade Numbers, and Lots of Value Left on the Table**

This evidence points to political authorizing environments as part of the problem. That is, legislatures and executive boards may seek reportable “results” with the best of intentions; but the targets and constraints that result induce sub-optimally conservative behavior and constrain field agents. Projects are designed so they can generate numbers; the obsession with metrics also affects what can be done in the field once a design is agreed.

Many projects do worse than they might as a result, but not all—indeed, some projects are better off *without* Navigation by Judgment. Navigation by Judgment is contingently beneficial—it's going to be more useful in some contexts and situations than others. It will be *most* useful—and current management practice is likely to cause the most problems—when environments are less predictable and tasks the kind that are hard to measure accurately.

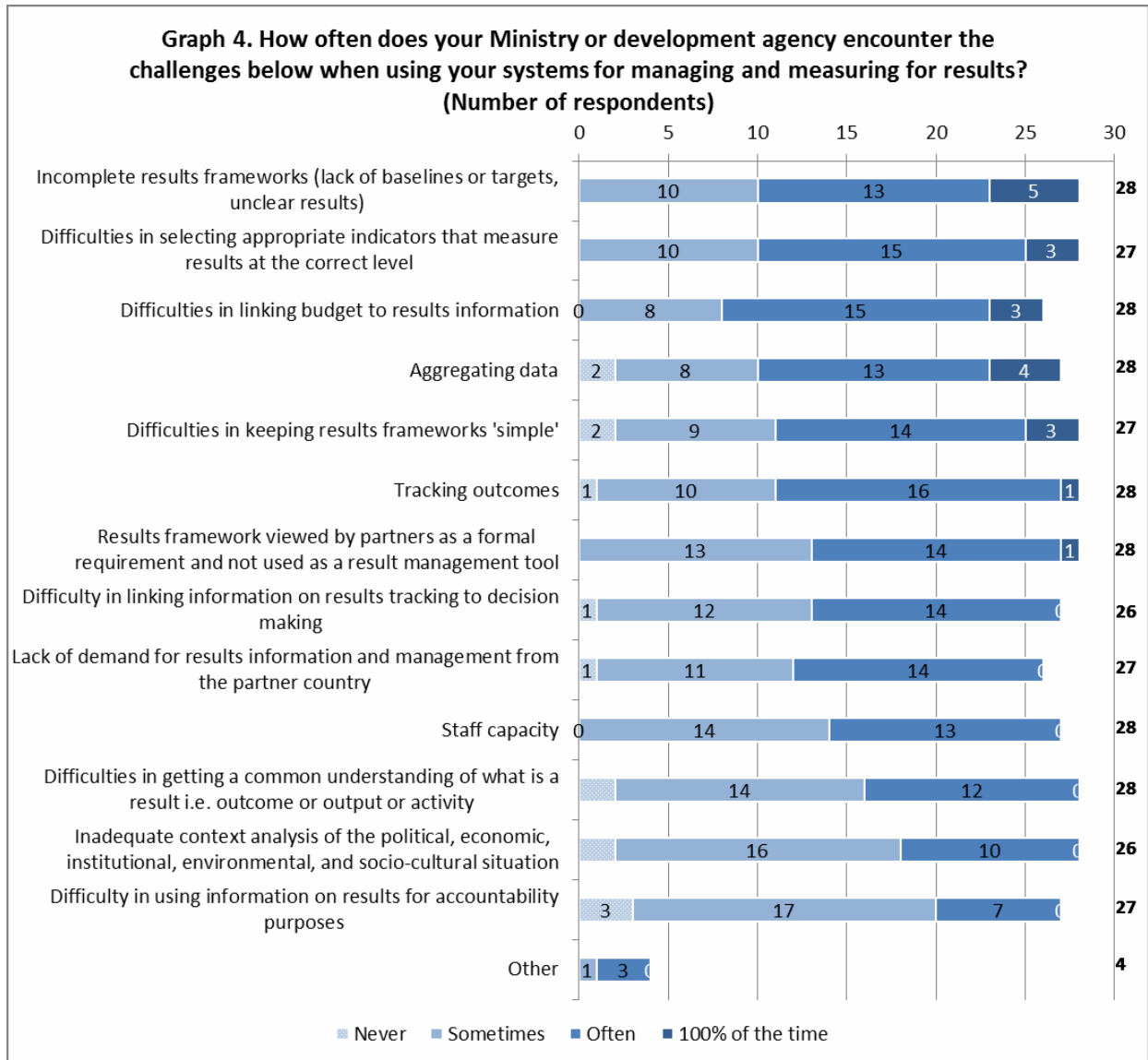
So where donors seek to build physical infrastructure in stable environments, the current system may be pretty good. But where projects have hard-to-measure statebuilding goals and/or the place where the project occurs is fragile, unpredictable, and hard to understand from afar, projects are very likely not realizing their fully potential due simply to management and reporting choices.

As noted in the preface, one of the biggest surprises to me in my conversations with aid agencies discussing this work was how widespread recognition of this problem already is; but perhaps it shouldn't have been such a surprise. Figure 4 is reproduced from OECD Development Assistance Committee survey on “Managing and Measuring for Results” filled out by members—that is, bilateral and multilateral aid agencies.<sup>8</sup>

---

<sup>8</sup> OECD 2013.

Figure 4. Difficulties in “Managing and Measuring for Results”



Source: OECD (2013) (Graph 4)

Figure 4 demonstrates that *every* respondent reported “difficulties in selecting appropriate indicators that measure results at the correct level”, and the majority of agencies said this was “often” or “100% of the time” the case.<sup>9</sup> And that does, indeed, have implications for accountability; 17 of the 27 reporting agencies reported they “often”, and another 7 reported they “100% of the time,” had “difficulty in using information on results for accountability purposes.”

<sup>9</sup> The totals don’t add up, and I cannot identify the underlying data. It seems 3 agencies responded this was a problem “100% of the time” and 15 “often” out of either 27 (the total of the categories displayed) or 28 (the reported number) of respondents.

Agencies all agree that the current management system is at least *sometimes* a problem.<sup>10</sup> This isn't "a problem" where "problem" means "what critics object to" or even "what you the reader think aid agencies should do" or "what citizens want or need." This definition of "problem" is entirely instrumental—that is, aid agencies agree *their own management system interferes with the goals of the agency*. Which is perfectly consistent with the message I heard over and over—that the diagnosis of my book is widely accepted, even if rarely articulated.

The question, then, is what we should do about it. The next section tries to articulate some principles of what "better" might look like. Not "better" in terms of what I or you might want aid agencies to do, if and where that differs with current goals—but rather "better" in terms of more successfully pursuing their own existing goals.

## **Section 2. Principles for How to Make Aid Delivery Better, With a Particular Focus on the Sacred Cow of "Accountability"**

Perhaps the most frequent question I got when I presented this work at aid agencies was about what the implications were of my work for "accountability." Often, the questioner would marry this to a statement in support of my diagnosis, saying something like "I totally agree, metrics are getting in the way and reducing our ability to succeed. And things would work better if we put more control in the hands of field personnel. But isn't this going to reduce accountability? How can that be good?"

It frankly took me a long time to make sense of this question.<sup>11</sup> I think of "accountability" as in service of some objective (impact, or efficiency, etc.)<sup>12</sup> The focus of many aid agencies is explicitly maximizing "value for money." For those agencies, if what we call "accountability" is a practice or process that reduces the numerator (value) without changing the denominator (money), "reducing accountability" is a good thing. Or, to be more precise, reducing a practice some agency or some inside the agency claim or believe is in service of "accountability." The word "accountability" is not a magic potion; sprinkling the term onto something does not, in and of itself, make that thing useful or good or welfare-enhancing.

Lant Pritchett and I have described the metrics obsession as a form of "accounting-based accountability," and argue accounting-based accountability is far from the only, the most helpful, or historically the primary way of conceiving of "accountability."<sup>13</sup> It feels as if somewhere along the way a lot of us started incorrectly using the term "accountability" as if

---

<sup>10</sup> The other 3 agencies not choosing "often" or "100%" choose "sometimes" on this item.

<sup>11</sup> Very helpful in this regard for me was Jerry Muller's *Tyranny of Metrics*, which outlines many varied domains in which "metrics" is confounded with "accountability".

<sup>12</sup> Accountability could also be in service of "deliberative democracy" or "justice" or "transparency" or "the public right to know" or something less 'cost-benefit'-y, of course. While harder to put into an 'equation' of the kind I do here, I still think it's hard to imagine that when the process of "knowing" is not leading to an accurate representation of reality that any of those objectives are being served, either, by the present accountability regime.

<sup>13</sup> Honig & Pritchett 2019.

it was the same thing as “objectively measuring everything we can and sometimes pretending to objectively measure things we can’t.”

So far this is a good description of what accountability shouldn’t be—but it of course begs the question of what an accountability system, or a broader management system, should look like to address the problems my research identifies. There’s a whole community of people and institutions who at this point have spent many years talking about the need to be more adaptive, or have more flexible programs. But for all the talk on the program level, there’s a relatively un-covered space on how agency-level practices need to change.<sup>14</sup> This section is an attempt to move in that direction.

### **Principle 1: Give Up the Delusion of Top-Down Fiat Control**

Modern management, economics, and political science’s understanding of organizations rests substantially on a branch of economics known as contract theory.<sup>15</sup> One of the core foundations of contract theory – and in my view a powerful way of understanding organizations—is to see management as engaged in contracts with their own staff, as principals “contracting” with agents rather than having some unassailable “control” over their agents. As one of the pioneering papers in the field (with over *18,000* citations) puts it, “It is common to see the firm characterized by the power to settle issues by fiat, by authority, or by disciplinary action superior to that available in the conventional market. This is *delusion*.”<sup>16</sup>

In order to get staff to do things, management (1) needs to be able to monitor them well enough to sanction and reward them based on performance; (2) needs some other form of sanction and reward (e.g., social reputation, or embeddedness in professional networks, etc.) to prompt staff to act in the manner management desires; (3) needs staff to *want* to do the job for their own internal reasons. In this, an aid agency is just like every other kind of firm. But unfortunately, aid agencies are places where the “sanction and reward” strategy is unlike to work, at least for most tasks.

Where sanction and reward won’t work, top-down management control isn’t a good idea in practice, attractive as it might seem on the drawing board/planning document/simple model. The idea that these controls and numbers are in fact an accurate reflection of reality

---

<sup>14</sup> This view comes in part from a CGD roundtable in late 2019 called “Toward a New Paradigm of Accountability”. And I think that “towards” – reused with permission in the subtitle of this paper – is exactly right. That is, I was struck at the roundtable that I didn’t think any of us (including me) actually had a very good answer to what “the new accountability paradigm” ought to look like. Thanks to Sarah Rose (the organizer) for the phrase – and of course including me in this event. The event itself was subtitled “The Importance of Locally Led Adaptive Learning in Fragile States”, though I focus here on accountability for more Navigation by Judgment, which I believe is likely to allow more “locally led adaptive learning in fragile states”, amongst other benefits.

<sup>15</sup> This is far from a fringe part of economics; in addition to Tirole’s 2014 Nobel, Oliver Hart and Bengt Holmstrom shared the award in 2016 for their work on contract theory.

<sup>16</sup> Alchian and Demsetz 1972, page 777. Citation count from Google Scholar, November 2019. Italics mine.

is a mirage, a delusion—and most people are aware of this. But just like a mirage may lead a caravan in the wrong direction, the delusion is nonetheless costly. That is, the delusion is dangerous in that those at the top have the power to *keep* those in the field from doing much of anything, by constraining them with tools that undermine the broader goals of projects. Controls can be powerful enough to be counter-productive even where they are not powerful enough to achieve their intended ends.

Some of those at the “top” of organizations are all too aware they only *seem* to be in control; that what actually happens depends on getting a lot of people they can’t (and will never be able) to control very well to do productive things. In the preface I mentioned a World Bank executive director who described top-down target setting as flawed, but as “The only way I can move the system at all, the only way I can steer; if I give up that lever, what do I have? Why am I here?”<sup>17</sup> The model aid agency would accept that most aid simply cannot be delivered by command and control; fiat control is a delusion. And the choice between the *delusion* of control and agent initiative is no choice at all, whatever the (very real) failings of relying on agent initiative may be.

The toothlessness of top-down commands goes for reforming the institution too, of course. In one case, the head of a department at a bilateral donor said something to the effect of “You know, I heard you talk, and read the book, and I’m convinced. So I asked my staff to draft a directive telling those in the field to use their judgment more, and make programs more adaptive.”<sup>18</sup> In some ways, I think this is wonderful. I of course think it would be great if projects were, where appropriate, more adaptive, and staff were trusted to use their judgment. But I don’t think “use your judgment more” is any more likely to work by fiat than is “improve results” or “perform better.” In my view, the department head would have been better served by figuring out why his staff did not currently use their judgment, and what changes to the work environment, or promotion standards, or procurement rules, or decision procedures, etc. might induce use of same.

In moving beyond the illusion of top-down control, the model agency would need to engage with aid staff as individuals; to think “contract” in the sense of “contract theory,” not “command.” This means thinking about the *agency* of agents; their motivation, the management support they receive, the barriers they face.<sup>19</sup> It means thinking about whether field personnel trust managers, and are willing to follow their lead. It means accepting that

---

<sup>17</sup> Relatedly (though less obviously managerial), Emily Springer has written eloquently about what she terms “the paradox of quantified utility” for gender advisors in an aid agency. Springer finds that advisors advocate the use of metrics even when *they themselves* believe the metrics divert attention away from important parts of the job. This is because the advisors believe that while quantitative indicators may not *actually* be superior knowledge, they are perceived by others as such; hence the need to generate numbers to secure continued funding for important work. Springer 2019. I use “aid agency” here, but it is possible gender advisors are working for an implementing partner/contractor rather than the donor themselves; these advisors are part of a “20-country bilateral agricultural development initiative”.

<sup>18</sup> To my knowledge, this directive was never promulgated. This may have just been a kind attempt to complement my work, of course; in any case, the point holds.

<sup>19</sup> See Honig & Gulrajani 2017 for more on these points.



the best way to manage an aid agency is by consent and agreement, not by top-down instruction.

## **Principle 2: Accept That Accurate Knowledge of Results is Superior to the Reductive Façade of Inaccurate Metrics—Even if That Reduces “Success Rates”**

A model agency would be one oriented towards *actually* knowing what its programs do, rather than simply generating reportable metrics to depict a façade of success and impact irrespective of the reality on the ground. This in turn requires accepting that not all aid is successful, particularly that which focuses on systems change.<sup>20</sup> Where donors support institutions, or governance, or systems strengthening—and well beyond—the current metrics of success are not painting an accurate picture, and almost everyone involved in actually implementing projects agrees.

What do I mean by “façade of success”? In the pair of projects in South Africa I briefly introduce in Appendix I, a USAID program focused on municipal government capacity building calls “results” something like “holding training sessions on expenditure management and anti-corruption.”<sup>21</sup> The comparator DFID municipal government capacity building program does a better job of incorporating narrative (from beneficiaries and other stakeholders) into its evaluation. As a result, the DFID evaluation describes mixed success, with good examples of impact but some localities with very little to show for DFID’s time and treasure. The USAID program, on the other hand, delivers all the trainings. It is thus seen as officially “successful.” Based on my review of documents and 35 qualitative interviews about this project pair, I think the official USAID story is in this case a façade.<sup>22</sup> The façade allows USAID to appear more “successful” if one were to compare the official records, but to in fact accomplish less.

A project aiming to help a fragile state transform is first round venture funding, not a place to put your pension savings. Aid projects aren’t baseball’s New York Yankees, or the

---

<sup>20</sup> I find there are very, very few aid professionals who actually disagree with that last statement, rarely uttered publicly though it might be. A number of my conversations at aid agencies suggest that one (but far from the only) common motivation for not pointing out the flawed success metrics of programs that seek to improve systems or institutions is tied to political authorizing environments; that is, aid agency leaders fear that pointing out the current system is a façade will lead inevitably to lower “success rates” and thus to less funding. Roche 2019 posits a range of other reasons for measurement’s distortions in the contexts of INGOs in arguing for “a better and more honest evidence base.”

<sup>21</sup> For more on these projects beyond the appendix, see Chapter 7 of Honig 2018. This project pair is also, more briefly, discussed in Honig 2019c.

<sup>22</sup> Including interviews with key members of the USAID project’s implementation team who had mostly negative things to say about the project. I was also unable to find any South African beneficiary who could actually *remember* the USAID project (this was not true of DFID’s effort). Interview number from Honig 2018, Appendix Table A1.1 (35 interviews for this project pair, of over 135 interviews total).

English Premier League’s Manchester United;<sup>23</sup> they’re the Detroit Tigers, or Leicester City Foxes. The Tigers and Foxes do indeed emerge victorious sometimes; but fans of either squad would be reacting ridiculously were they terribly shocked each and every season their side failed to win the championship.

At best a given externally funded project which aims to catalyze broader, systemic change (rather than build a road or deliver needed services etc.) has a puncher’s chance of success. But as every angel investor knows, longshots are still good bets. From every project the model agency can and ought to learn, leaning forward into the next earnest attempt to facilitate inclusive growth, or reinvigorate the civil service, or strengthen the health service, or support the fair and equitable administration of justice with a greater understanding than in the last attempt. Here the perpetual fan’s cry at season’s end—“We’ll get ‘em next year”—is exactly right. But to in fact better the next project’s chances of succeeding, we need to be able to admit and understand the failings of its predecessor.

The model agency needs to get beyond what I’ve elsewhere called aid’s Obsessive Management and Control Disorder,<sup>24</sup> and increase the scope of what can be mentioned, what can be discussed, regarding the success, failure, and potential for learning and improvement in aid projects. That USAID project I mentioned was actually a phase 2 project—phase 1 had used a very, very similar model. Where agencies don’t gather accurate information on results (as distinct from quantitative metrics masquerading as results), it is much more likely that the next project will repeat the mistakes of its predecessor.

Very few successful founders will tell you they had exactly the right model when they founded their firms. I suspect that were we more honest about development, we would find the same true of aid projects; that the initial design is rarely the right design. More honesty and more flexibility go hand-in-hand, inasmuch as the obsession with top-down control and delivery targets is consistent with the view that if we just get the setting on all the dials right, aid will *work*. But that’s just not the case; aid isn’t an engineering problem. And so more honesty will engender more flexibility, both within programs (per Adaptive Management) and over time in trying to address a problem (per Problem Driven Iterative Adaptation).

There’s a catch, however—the alternatives to the false sense of certainty that metrics provide is not some other source of data that, in addition to being more accurate, will give the same sense of certainty. That is, moving beyond the façade of metrics (where it is a façade—which is sometimes, not always) means relying on other data that may be more accurate, but whose lack of absolute, unquestionable reliability is obvious. It is self-evident that narrative accounts, or qualitative perceptions, or feedback surveys, are not always 100% reliable. In some contexts, they are nonetheless *more* reliable, or, less distortionary, or more compatible with good performance (or all three of these) than controls and success reporting based on

---

<sup>23</sup> Yes, yes, I know that Man U has had some down years recently. The Yankees have had some bad patches, too; and Man U did come top of the table in the *majority* of the first two decades (11 of 20 seasons) after the Premier League formed in 1992. For the record, lest there be any confusion, I’m a Spurs fan – but they’re maybe a bit too good to be the Tigers in this analogy, and surely aren’t the Yankees.

<sup>24</sup> Honig 2019c.

quantitative metrics. But they sure don't *feel* that way. Thus moving past the façade requires not just accepting a different basis for “knowing,” but also accepting that there will be uncertainty in whether we know. The façade is comforting not just because of its reductiveness, but because of the false sense of knowing it provides.

I cannot count the number of times I have heard someone in the aid world tell me “well, you know, what gets measured gets managed.”<sup>25</sup> While not actually necessitated by the phrase, this is usually said with the implication “so we should, or need to, measure everything we care about or think important.” The original quote, possibly from Peter Drucker, is “What gets measured gets managed—even when it's pointless to measure and manage it, and even if it harms the purpose of the organisation to do so.”<sup>26</sup> The quote itself in fact implies the opposite of its common-day usage; that is, we should think about what important things that we should care about but *shouldn't* measure, because sometimes measuring will be counter-productive.

The way in which it's primarily counter-productive in aid agencies, in my view, is by crowding out what we can consider in discussing “success.” I think the more important guiding quotation is the philosophy of American children's television host Mr. Rogers and his longtime collaborator Margaret McFarland—“What is *mentionable* is manageable.”<sup>27</sup> The original context is child psychology—teaching children to be able to share, discuss, and think about their feelings. But I think it works for management of aid agencies, too.

Accepting that accurate knowledge is better than a false sense of certainty and success is an important step in the right direction, critical to moving beyond simply restricting our vision to what we can count, and declaring success when those things are delivered. The current system is not just adding time and reporting effort; it is undermining success, making it less likely that aid will have its intended impact.

### **Principle 3: Use Different Accountability Systems for Different Tasks, Environments, and People**

Navigation by Judgment requires a very different conception of accountability than does the current system. If different kinds of tasks and environments necessitate different management practices that means they need different ways of thinking about accountability, too.

If you were to stop people at random on the street anywhere in the world, give them whatever your preferred definition of “accountability” is, and then ask them “how should I set up accountability routines for aid personnel helping build a road network in Turkey” and “how should I set up accountability routines for aid personnel helping build a justice system

---

<sup>25</sup> The quote is often quoted as from Drucker 1954 (which I can't get a copy of at the moment but is suspiciously never quoted including a page number).

<sup>26</sup> Ibid. Many thanks to Barnett 2015 for alerting me to the longer quote. Italics mine

<sup>27</sup> As reported by Laskas 2019. Italics mine.

in Sudan,” they might not give you terribly well-informed answers to either query. But I strongly suspect they would give you *different* answers to the two questions. Because it’s pretty obvious, frankly, that the nature of the task, and the environment, means you’d probably want to manage these two things differently.

While it would be possible to manage the same person differently when they worked on the Sudanese justice system if they transferred to the Turkish road, it would be much easier to do incorporate “task clustering,” as articulated by leading economists.<sup>28</sup> That is, have some people work on tasks which require a lot of unverifiable judgment and thus necessitate Navigation by Judgment; have others work on tasks that require very little and thus where metrics make quite a bit of sense. This is because it’s hard for people to rapidly rotate between different kinds of accountability and reporting schemes. Accountability routines shape not just the team, but also the individuals’ instincts and ways of working.

The model agency might, then, have departments like “health” and “infrastructure,” but in each of these departments it would have sections that incorporated how much Navigation by Judgment was required into the design of accountability routines.<sup>29</sup> So in health, the “vaccine delivery” team would look very, very different than the “health systems strengthening in fragile states” team.

Having one set of accountability routines for all personnel in an aid agency is the rough analogue of holding James Bond, Q, and the people responsible for having the helicopters ready on the roof of MI6 to the same reporting standards.<sup>30</sup> The helicopter team does incredibly important work—work that has a natural output, well-captured by a set of verifiable metrics (e.g. helicopters in good working order, fully fueled, at all times; no accidents). The helicopter team could indeed labor under an accountability system full of metrics and monitoring and reporting. Q (the R&D/gadget design team) can’t be held to the same kind of performance metrics... but there are still medium-term outcomes (cool things invented) and indicators (staying ahead of other spy agencies) that are entirely reasonable for this work. But Bond? Bond needs a different kind of accountability routine.

---

<sup>28</sup> both discuss clustering in the public sector, though they focus on clustering by the talents of agents and the extent to which managers can learn about same.

<sup>29</sup> While it would be possible to manage the same person differently when they worked on the Sudanese justice system if they transferred to the Turkish road, it would be much easier to do incorporate “task clustering”, as articulated by leading economists. (see e.g. Dewatripont, Jewitt, & Tirole 1999, 215) & Dixit 2002, 707.) That is, have some people work on tasks which require a lot of unverifiable judgment and thus necessitate Navigation by Judgment; have others work on tasks that require very little and thus where metrics make quite a bit of sense. This is because it’s hard for people to rapidly rotate between different kinds of accountability and reporting schemes. The argument below implicitly extends clustering to incorporate the notion of differential mission orientation as discussed in e.g. Besley & Ghatak 2005 and verifiability of task as discussed in e.g. Aghion & Tirole 1997, which would require clustering of people and career concerns based on more than simply agents’ technical abilities. Accountability routines shape not just the team, but also the individuals’ instincts and ways of working.

<sup>30</sup> I thought I came up with the “Bond” bit myself, but apparently I’m not the first to use this analogy – see Roche 2018. As Roche 2018 puts it (quoting a staff member of Coalitions for Change), “James Bond didn’t write reports.”

## Principle 4: Accept that Changing Agency Practice Requires Constant Tending

Many in the aid world are currently talking about “adaptive management” and how to make it work in managing aid projects.<sup>31</sup> Key to the adaptive management conversation is the need to *keep* changing, to *keep* focusing on what’s working and now. The same is true of management practice in aid agencies, including any move to push towards greater Navigation by Judgment.

In a recent paper Bain, Booth, and Wild, in talking about how the World Bank might “do development differently,” talk about the organization’s “plumbing.”<sup>32</sup> As a former apartment complex maintenance man (which included some plumbing work), I find the analogy apt.<sup>33</sup> To fix—much less change—the plumbing, you sometimes need specialized tools and knowledge, but more often you need patience and careful observation. Most of all, you need time—that is, you can’t just close the maintenance ticket in one visit, but with multiple returns over many days. Very often in plumbing, you think you’ve found the source of the problem, but you’re wrong.

In 2014 DFID introduced a new operating framework known as the Smart Rules.<sup>34</sup> The Rules

*Provide the basis for open and flexible development programming, achieving results innovatively and at pace, and are focused on outcomes. We want staff across the organisation to be both empowered and accountable:*

- **Empowered** to create better policy and programme outcomes, through increased innovation, taking well-managed risks with flexibility to adapt to realities on the ground.
- **Accountable** for effective delivery, through clarity of roles and responsibilities to consider choices and make good decisions, underpinned by quality information.<sup>35</sup>

As a vision statement this sounds pretty good to me. If I was lucky enough to get an invite, at the first executive committee meetings to set up this model agency I keep talking about, I might propose this vision statement as the basis for our discussion.

---

<sup>31</sup> See e.g. Hernandez, Ramalingam, & Wild 2019, which addresses when codifying adaptations or the need for same is useful, in an example of the kind of hybrid system potentially adaptable to aid agencies’ reliance on agent judgment.

<sup>32</sup> Bain, Booth, & Wild 2016.

<sup>33</sup> I’m not the only one; 2019 Economics Nobel Laureate Esther Duflo has described economists as plumbers as well, arguing in Duflo 2017 that plumbers like economists should be “mindful that tinkering and adjusting will be necessary”.

<sup>34</sup> See DFID 2019, “Smart Rules: Better Programme Delivery”. While originally issued in 2014, they have been updated repeatedly since.

<sup>35</sup> Ibid, page 8.

When the Smart Rules were first introduced, DFID’s Independent Commission for Aid Impact (ICAI) conducted a review.<sup>36</sup> The review praises DFID for moving to simplify rules and providing more flexibility, in an echo of my concerns about donors’ obsession with control; indeed, ICAI argues that “the rules need to be as lean as possible and could be pared down further” and worries that it’s possible “the number of guides and rules [could] creep back up over time.”<sup>37</sup> Echoing Principle 2, ICAI also notes the need for more honest assessment of agency performance and explicitly links that honest assessment and the organization’s ability to learn and improve.<sup>38</sup>

I remember literally tearing up, ridiculous as that may be, the first time I read in the ICAI review that “Enabling effective discretion, where the **judgment of frontline staff is trusted**, is important and key to the success of adaptive programming and continuous learning.”<sup>39</sup> I’d already done the empirical work for what would become the book, and had already decided that I did indeed want to make the “judgment” part the core of the central concept. And to see not just a major agency, but an independent review commission, explicitly tie staff judgment to good programming in support of an actual agency reform felt... impossible, but here it was. It gave me hope; a hope of more fulfilling jobs and more impactful aid, of a Smart Rules-powered DFID boldly leading other donors towards a Navigation by Judgment-laden future.

From my many conversations about the Smart Rules in the past few years it seems to me that in practice the Smart Rules have led to modest at best increases in the use of judgment in the field.<sup>40</sup> The smart rules are simpler and clearer than the regime they replaced, which is of course a good thing; but even where the rules allow staff to exercise judgment and discretion, staff often choose not to actually make use of what the rules allow. The answer to my “why not?” question almost always involves a 3-letter acronym I’ve come to loathe—“SRO.”

An SRO is a “Senior Responsible Officer”— the person responsible for project oversight. And when I say, “responsibility for oversight,” I mean a LOT of stuff. The current version of the Smart Rules contains the phrase “SRO must” *64 times*, including (and please stop reading this section/skip beyond the bullet points when your eyes start crossing):

---

<sup>36</sup> ICAI 2014. ICAI is a remarkable institution in its own right – led by appointed experts, it reports to parliament, not the executive. It thus is a way for the UK’s International Development Committee, and by extension parliamentarians, to get knowledge that is both independent and informed about these complicated issues. US observers often bemoan the difference between US and UK authorizing environments; in my view the existence of a permanent staff for parliamentary committees and institutions like ICAI that help the legislature get independent, well-informed data and opinion is an important part of bettering authorizing relations.

<sup>37</sup> Ibid, page 4.

<sup>38</sup> Ibid, pages 10-11.

<sup>39</sup> Ibid, page 5. Boldfacing mine.

<sup>40</sup> I’d say I’ve talked to more than 20 but less than 50 people about the Smart Rules, and DFID’s authorizing framework more generally, over the past few years. Of course this is far from a systematic sample, or formal research, etc. – this is my educated but far from expert perception, nothing more, nothing less.

- “The **SRO must** ensure that financial forecasts for programmes are realistic, updated monthly on ARIES, and should explain any variances and how they will be managed.”
- “The **SRO must** ensure that the programme is appropriately monitored throughout the year and that the delivery plan, logframe or similar, risk register and AMP delivery chain map are updated as necessary, with significant changes to the logframe agreed with the Head of Department or delegate.”
- “The **SRO must** ensure that there is a complete, accurate and up-to-date inventory for all programme assets owned by DFID, and that physical checks of the assets take place at least annually.”
- “The **SRO must** ensure that a risk assessment is embedded within the business case and that risk throughout the life of the programme is managed within the risk appetite approved within that business case or that changes to the agreed risk level are subsequently approved by HoD or delegate
- “The **SRO must** seek advice at the concept note stage of the programme, and ensure that at business case stage they have approval from: Financial Accounting in FCPD for any liabilities, guarantees, indemnities, returnable investments or other contentious or novel financial arrangements, including projects that will generate income for DFID; and from the Finance and Performance Department in FCPD for any non-ODA spend; and Management Accounting for any spend beyond the current Spending Review period.”
- “The **SRO must** ensure that all programmes follow DFID’s UK aid branding guidance – including having a completed visibility statement in place and that digital elements of programmes (e.g. text messaging, cash transfers with a digital angle, websites and databases), are reviewed at the earliest possible stage in the process by DFID’s Digital Service Team.”<sup>41</sup>

You’re maybe thinking “huh, that SRO thing sounds like a crummy job. People must really want to avoid that job and instead really want the jobs where they’re empowered and get to use their judgment and such, like you outlined earlier.” But here’s the thing—it’s often *the same job*. The SRO *is* the person who’s supposedly empowered to use their judgment.<sup>42</sup> The person who is responsible for incorporating the nuance of the field into judgment-based decisions is the same person responsible for making sure all the DFID-funded desks and chairs (physical checks of assets) get counted annually, and that the DFID logo’s properly displayed, and that the risk assessment is done, not to mention the financial forecasts and the logframe and much, much, much, much, much (much) more.

That 2014 ICAI report I’m so fond of discusses the SRO role, which first came to DFID via the Smart Rules, as well. The report opens with praise for the role; “We are positive about the introduction of the concept of SROs to DFID... We agree that identifying clearer

---

<sup>41</sup> DFID 2019, various pages. Boldfacing mine.

<sup>42</sup> Where it’s not, the “empowered judgment users” usually report to the person who’s the SRO, which has the same chilling effect on the use of judgment.

accountability is important and the benefit of knowing who can make the key judgments and decisions at a particular point in the project lifecycle is a step forward.”<sup>43</sup> But in the same subsection ICAI articulates a worry which in my view has proved prophetic: “these [SRO] roles will only meet the planned objectives if they are operating within the right overall environment of risk appetite and supporting culture. If they simply represent the focus of potential blame for failure, they will not be attractive or effective.”<sup>44</sup>

What I hear over and over again from DFID staff, contractors, recipient country government officials, etc. is that SROs largely believe—rightly or wrongly—that if they do anything they can’t formally account for, or insert their judgment in any way, they will be sanctioned if anything goes wrong. The SRO has indeed come to “represent the focus of potential blame for failure”, and that has indeed in turn made them not “attractive or effective”. The Smart Rules offer more autonomy, and more judgment, but coupled with more “accounting-based accountability”. It says to staff in effect: “Use your judgment, but also report on everything. And if anything goes wrong, you’re responsible.” Accounting-based accountability has crowded out anything else, coming to constitute all of what “accountability” is in practice. To DFID’s credit, the Smart Rules have been revised to give increasing amounts of discretion to those in the field.<sup>45</sup> But as of yet these changes do not seem to have unlocked the Smart Rules’ full potential.

I can’t help but wonder what might have happened had DFID, rather than rolling out the Smart Rules across the agency in one go, simply piloted them. Road-tested the rules, as car companies do new models to find some of the unintended consequences, with a focus on organizational culture and a close eye on how field personnel responded. Building in rigorous evaluation (where rigorous probably includes something like embedding ethnographers in the organization, not an RCT) might have helped, too. As it stands, I can’t help but feel that DFID’s Smart Rules have yet to achieve their laudable objective.

The DFID experience strongly suggests that simply changing the formal rules to put more decision power in the field isn’t enough. Perhaps the single most important part of the “model agency” is that the focus would be, consistent with the spirit if not the mechanics of a “results orientation” as conventionally practiced, on what is *actually* happening. *Actually* done. *Actually* felt in the field, by implementers. The model agency would recognize that “fixing management practice” isn’t something to do and then stop doing—it’s a continuous, not a discrete, activity.

The agency would continue to pay attention to its own management practice and evaluate where it is appropriate and where not, including through third-party evaluators who could report back to authorizers. Evaluations would assess the impact of Navigation by

---

<sup>43</sup> ICAI 2014, page 5. To be clear, other UK agencies already had introduced SROs; it is their introduction at DFID, not to the UK Government as a whole, which was part of the Smart Rules.

<sup>44</sup> Ibid, section 2.13

<sup>45</sup> This includes changes to whether full program documents need to be approved at the Ministerial level (they no longer do) and an increase in the monetary threshold of spending/programs that can be approved at the sub-ministerial level. See DFID 2019, Table 3.



Judgment—examining where more Navigation by Judgment is leading to better development outcomes in line with the agency’s priorities—but also on the process via which Navigation by Judgment is being achieved, or constrained. That means remembering to be a plumber – to tinker, to adjust, to pay attention to the organization’s *de facto* practices and constraints. That is, paying attention not just whether the organizational pipes are aligned to manufactured specs but also whether water is going where it’s supposed to and how, if possible, to improve the flow.

## **Principle 5: Move Towards a Default of Trust Where it Serves Agency Goals**

What would the model agency’s accountability system look like for tasks and environments that require Navigation by Judgment? How would a model agency with a vision similar to that of the Smart Rules’ authors actually put it into practice?

At the risk of seeming at this point to be actually on the payroll of ICAI (which I promise is not the case), they point nicely to the heart of the problem. That is, per Principle 1 and contract theory, using judgment has to be made “attractive” to be made “effective.” And that is, indeed, both about “the overall environment of risk appetite and the supporting culture.”

When I first started talking about the book I would discuss “coaxing” or “cajoling” or “inducing” people to use their judgment as a way to make actually using judgment attractive. And I continue to think ensuring those who use judgment won’t be punished if their judgment call proves the wrong one, and won’t face future career sanction, are part of the solution. But the core of that solution, the core of what a new accountability paradigm, is something so simple as to seem naïve: Trust.

Jane Mansbridge calls for more “trust-based accountability” where appropriate, and notes this is “most appropriate in contexts of justified trust”—that is, where you think the folks you’re trusting are likely to do, on the whole, a pretty good job.<sup>46</sup> Based on my experience interacting with aid agency employees I believe every aid agency I’ve ever seen is indeed a context of justified trust, populated largely (though of course not entirely) by people who actually *want* to do the thing the agency seeks.<sup>47</sup> A model agency would start with a default of trust—assuming its people in the field were competent, honest, and earnestly pursuing their mission until proven otherwise—and be based around figuring out what HQ could do to support those in the field. This support would be both technical (e.g. providing the agent with specialized expertise when needed) and institutional (e.g. having the agent’s back and ensuring that they feel supported in making decisions). This means that the arrows of authority would in some relevant sense flow “up” from the field; and the arrows feeding and

---

<sup>46</sup> Mansbridge 2014, 55.

<sup>47</sup> That is, who have a high degree of mission match in the sense of Besley & Ghatak 2005.

supporting the decision would flow “down” from headquarters, in a reversal of the de facto environment today.<sup>48</sup>

If “start with trust” sounds crazy, then let’s remember what the current default is: a reporting and control regime that asks for so much verifiable justification that it undermines the very value it seeks. Maybe no one ever thought “well, let’s start with distrust” in building this system; but let me say, if someone *had* thought that, I’m hard pressed to come up with how the current system would need to change to match up with the “start with distrust” starting point.

In my book I talk about the effect of this system on staff. High-quality staff who want to make a difference are disproportionately likely to exit. Many high-quality folks who truly care about a low Navigation by Judgment aid agency’s goals never even apply, as they don’t think working at the agency is a good way to pursue the agency’s own goals. And those who stay are demotivated. As Henry Stimson, US President of Truman’s Secretary of War, put it in a letter to Truman: “The chief lesson I have learned in a long life is that the one way you can make a man trustworthy is to trust him; and the surest way to make him untrustworthy is to distrust him and show your distrust.”<sup>49</sup> The management and accountability system has been showing aid personnel distrust. That so many remain aid staff remain dedicated and motivated is despite, and not because of, the current performance and project management and reporting systems. As Mansbridge puts it, “sanction-based accountability” (e.g., accountability based on agent rewards and punishment tied to following rules and delivering on pre-specified targets) “not only stems from distrust; it creates distrust.”<sup>50</sup>

Sweden has taken this to heart, and has a public “Delegation for Trust-Based Public Management” (which I’m going to call “the Delegation” here) to examine public sector management practices.<sup>51</sup> The Delegation starts by thinking about where value is created – the field. As the Delegation puts it, “The starting point we use in discussions of management is the meeting between the citizen and the civil servant. This meeting is the ‘moment of truth’; it is where value and quality arise.”<sup>52</sup> The Delegation notes that top-down control from a distance won’t work; that “management should be conducted not at a long distance

---

<sup>48</sup> Many thanks to USAID’s Stacey Young for the imagery here, at the CGD roundtable on accountability mentioned above (used and attributed with permission). Young’s use of the image extends further “down” through implementers to communities (to whom she feels aid projects can and should be ultimately the key judges, and guiders, of what aid agencies do); that is, her use of the term is more expansive than my use here.

<sup>49</sup> Stimson 1945

<sup>50</sup> Mansbridge 2014, 55.

<sup>51</sup> Tillitsdelegationen; <https://tillitsdelegationen.se/in-english/>. The Swedish delegation, to be clear, is *not* about aid, but rather focused on delivery to Swedish citizens – though there are moves to explore Trust-based management in aid as well. My awareness of this work stems in part from being on a reference group (a kind of steering committee) for a study on “Trust and Trust-based Management in Aid” commissioned by Sweden’s Expert Group for Aid Studies (EBA).

<sup>52</sup> Tillitsdelegationen 2018 (English version). This is very, very close to how James Q. Wilson put it in what remains in my view the best book on bureaucracy ever written: “The work of the operators is the place to begin because it is their efforts that determine whether the agencies’ clients are satisfied.” (Wilson 1989, p. 34)

‘by remote control,’ and that “Our ambition for more trust-based public management is to reduce detailed management so that staff and managers in health care, schools and social care will have more freedom to act.”<sup>53</sup> But right in their definition of trust-based management the Swedish the Delegation note that formally changing the rules isn’t enough. As they put it:

*The basic idea of trust-based public management is to increase and retain the quality of publicly financed services through more trust in management that also brings more freedom for action to the meeting between the citizen and the employee. We define trust-based public management as management, culture and ways of working that focus on the purpose of the service and the needs of the citizen, in which every decision-making level takes active steps to stimulate collaboration and a holistic perspective, build trusting relationships and ensure that the employee is able to—**and wants and dares to**—help the citizen.*<sup>54</sup>

As hard as it may be to change the formal rules to allow Navigation by Judgment, it’s even harder to actually get it used – to have those in the field “want and dare to” rely on their judgment. Hence the need to begin, simply put, with trust.

In praising the Swedish Trust Delegation (which is Government-wide and not focused on foreign aid, to be clear) I don’t mean to imply that a good vision is sufficient or that Sweden’s Sida is perfect—not even close. While Sida’s policies are amongst the most Navigation by Judgment-incorporating aid agency I’ve encountered, my conversations at Sida suggest that current promotion systems and procurement rules may undermine agent use of judgment – creating a substantial gap, in an echo of DFID’s Smart Rules, between what is “allowed” formally and what is actually done.<sup>55</sup>

In an echo of the overall argument, the degree of Sida’s Navigation by Judgment has also been subject to repeated rollbacks due to external pressure to generate measurable results.<sup>56</sup> As Vähämäki describes it, “Lack of trust... leads to a search for solutions in the form of results initiatives that can increase (the feeling of) objectivity and bring back trust.”<sup>57</sup> But in

---

<sup>53</sup> Tillitsdelegationen 2018 (English version).

<sup>54</sup> Ibid. Boldfacing mine.

<sup>55</sup> Underscoring this, when I spoke at Sida more than one person said something to the effect of “oh, it’s good that the folks from Foreign Affairs are coming to your public talk, they’re always trying to force us to put in more controls.” Even the existing degree of Navigation by Judgment at Sida is hard-fought and in need of constant protection. See Gulrajani & Mills 2019 for a fuller description of current tensions at Sida through a risk management lens.

<sup>56</sup> Vähämäki 2017 is the fullest description – this is where you can find chapters covering Sweden’s results monitoring systems from 1971 onwards. See Vähämäki 2015 (Chapter 8 in Eyben et. al.2015) for a shorter version of the argument, and Vähämäki & Verger 2019 for a broader overview of results-based management reviews from across OECD DAC.

<sup>57</sup> Vähämäki 2017, 236.

an echo of Principle 2's concern Vähämäki finds that Sweden's results information has in many ways not reflected reality, and thus distorted effort and action.<sup>58</sup>

So how to make it work in practice? What would “accountability” look like in this system? Pritchett and I in our work talk about alternative forms of “account,” of justification, of explanation.<sup>59</sup> We note that many—perhaps most—successful organizations allow narrative accounts to form part of their accountability systems. Doctors are not judged solely on patient outcomes, as every case is different (and doctors may face differentially ill/difficult/complicated cases). Difficult cases are discussed, with the doctor's narrative of the case playing a substantial role. Same with military after-action reports; not every retreat is a bad one, nor is every retreat good. The question of good or bad relies on more than simply observing the retreat, or even the position of the enemy; the question rests on a judgment *of* judgment—on whether the judgment of the commanding officer was reasonable given the circumstances. And that means narrative accountability is embedded in accountability to other human beings (peers, supervisors, a professional community), and to their judgment of the narrative-providers judgment.

Susanna Campbell uses the phrase “accompaniment-based accountability” in her work, suggesting that ‘accountability’ works best when staff and even management feel responsible to, and part of a community with, program beneficiaries and other staff placed in the field. Campbell stresses that this requires repeated contact, in-person dialogue.<sup>60</sup> This is also what I see in aid agencies; staff who care about doing a good job and their reputation as professionals.<sup>61</sup>

Campbell quotes an interview with a donor representative in DRC that nicely depicts much of this vision:

*Our organization, at the moment, has a very strong emphasis on having an open relationship and expressing when things aren't going right. You see this in how we spend the funding. It is a difficult environment and no one knows what will and will not work. We say, let's be realistic and let's have a partnership and dialogue around it. When our partners see that this space is created then they have a dialogue around it. We don't just want a glossy report. We will not condition their funding on whether they admit failure or not.<sup>62</sup>*

---

<sup>58</sup> See Vähämäki 2017, 244-245. She uses the metaphor of “The Matrix” and Neo's choice of red or blue pill as an analogy, arguing that results information can “actually lead us to miss what is happening in the ‘real world’”. (245)

<sup>59</sup> Honig & Pritchett 2018.

<sup>60</sup> Campbell 2018, pgs. 242-244.

<sup>61</sup> In Honig & Weaver 2019 we find that professional reputation (rather than sanction) is also the primary driver of responsiveness to the Aid Transparency Index (ATI); so in the ATI case the metrics “work” not via their use as a control, but rather by codifying a norm which dedicated professionals are eager to achieve.

<sup>62</sup> Campbell 2018, p. 244

In the most effective use of Navigation by Judgment by USAID I examine in my book, USAID effectively co-locates staff with the implementing contractor.<sup>63</sup> These staff—USAID, the head of the contractor’s team, and local Liberian staff—worked together in a true sense of the term, with judgments made collectively and with a relatively flat hierarchy. One Liberian team member said there was “room for thinking on your own, for taking decisions.” Multiple interviewees talked about working harder, for the same or less pay, than on any other project they’d been a part of—and also called it their *favorite* project they’d been a part of, the one they were proudest of. While the reports for the formal accountability system were all filled out, in the case of this project—and no other USAID project I examined—the formal system was somewhat of an afterthought. The actual “teeth” of the accountability regime were in the dialogue between the people on the project team, in making account—not in accounting.

Accountability for staff in the model agency who work on tasks that require Navigation by Judgment, then, would involve a fair bit of dialogue; of field staff with their supervisors, of supervisors and staff with colleagues and beneficiaries. Accountability would be about understanding, cultivating, supporting, and sharpening judgment while earnestly and honestly addressing successes and failures. It would be aimed towards accompanying and supporting staff, not monitoring and controlling them. It would also attempt where practicable to draw on the judgment of recipients—in some cases e.g., to replace output metrics with satisfaction and use data.

In collecting more honest information about program success and failure, it would also help the model agency understand where more (or less) Navigation by Judgment was needed, and over what (finance, personnel, program direction, etc.) That is, not only do we have good reason to think this accountability system, by allowing more Navigation by Judgment, would improve aid intervention success. We also have good reason to think that such a system would lead to continued improvement, and thus even greater Value for Money, the longer it went on.

This doesn’t mean trust can, or will, never be abused. In that USAID Liberia project, there was an incident of procurement fraud; of an individual buying goods soliciting bribes from suppliers. In no other project does any interviewee talk about any corruption; and I suppose it is *possible* that this is the only case of malfeasance in any of the projects I examine. But I find this unlikely; it seems more likely that the unique feature here is that the accountability regime of the project made the reporting of this to an outsider (me, the interviewer) seem like an appropriate thing to do. And what happened as a result of this fraud? Other local staff, acting as a group, confronted the fraud committer in a kind of intervention. One

---

<sup>63</sup> This is Building Reform and Recovery through Democratic Governance (BRDG) Liberia, 2006-2008. BRDG was an Office of Transitional Initiatives (OTI) program, which gave it the operating flexibility to do as described here, which likely would not have been possible for most USAID programs. The contractor was Development Alternatives International (DAI) See Honig 2018, pgs. 95-101, for more on this case.

described the central message as telling the fraudster they were “ruining **it** for us... how dare you embarrass us this way.”<sup>64</sup>

The “it” in the quote is, for me, quite telling. “It” was what currently is, in aid, far too rare of a thing; being treated with trust, feeling empowered, having one’s judgment valued. Being treated with trust induced staff to want to live up to that trust. Thus discursive, metric-less, non-accounting-based accountability was able to accomplish what I suspect a top-down system never would have: an end to the fraud. Uncovering petty procurement fraud would have been *very* unlikely for anyone to identify but other local staff, who heard about it through their social networks.<sup>65</sup>

A trust-based accountability system would empower professional networks and others with experience to primarily judge the rightness or wrongness of action; it would focus primarily on horizontal channels, not reporting to those “above.” Trust-based accountability “down” from HQ to the field is itself facilitated by trusting relationships between agencies and authorizers. Where possible, such a system would bring authorizers (e.g. members of congress or executive board members) further into the decision process; with more authorizer visits to the field, more active communication, more collaboration and less defensiveness.

None of this is to suggest that a trust-based accountability system would involve no sanctions, or no reporting to those at senior levels.<sup>66</sup> What would change would be the primacy of reporting up as a control (rather than a learning) mechanism, which means necessarily untying reports from sanctions (by management, or by authorizers) and the nature of that reporting (from sending data to engaging in dialogue/narrative accounts).

As Charles Kenny has put it, foreign aid’s goal is “Results, not Receipts,” and thus reporting and management systems should focus on the “receipts” part only inasmuch as it’s in service of the “results” part.<sup>67</sup> Starting from a default of trust doesn’t mean no controls, or e.g. no sanction for those who engage in fraud; but it does mean not *only* focusing on minimizing fraud without paying any attention to what the rules that do so mean for performance. In their work on administrative burdens, Pamela Herd and Don Moynihan point out that in setting the rules to tolerate as little fraud as possible, Government social welfare programs inevitably reject many deserving applicants who can’t prove themselves beyond a shadow of a doubt.<sup>68</sup> A default towards trust in management practice means not operating on a “shadow of a doubt” standard—and that does, inevitably, mean there will be instances

---

<sup>64</sup> Honig 2018, p. 97; boldfacing mine, of course.

<sup>65</sup> In this case the fraudster reportedly changed behavior and became a productive member of the team; had the actions recurred, I strongly suspect they would have been reported, and fired. Personally, I think achieving a “within-person” change – altering the person’s way of working – is even more impressive, and suggestive of the potential positive effects of trust-based management, than his firing would have been.

<sup>66</sup> “Even systems that have evolved to work largely on public-spirited internal motivation, selection, and justified trust need some sanction-based accountability at the edges of the system”, as Mansbridge puts it. (2014, 62)

<sup>67</sup> Kenny 2017.

<sup>68</sup> Herd & Moynihan 2019.

where trust is abused. This is hard for agencies to talk about—and they often seem to believe their authorizers will balk at any risk whatsoever, and never accept that. But I hear from authorizers that often they want little more than to better understand what’s going on, to be brought “in” to decision making. The risks and benefits of trust are a potentially excellent opportunity for just such engagement. In any consideration of the costs of trust-based accountability it’s necessary to compare it to the current state of affairs – which I’d describe as distrust and control in many agencies. That is, a fair evaluation would compare the comparative gains (from e.g. reduced fraud) from a default of distrust against the foregone performance benefits of the current system. A tight system of control may prevent more fraud - but in so doing it will ensure many actions in service of an agency’s mission also are go untaken, and thus authorizers’ objectives also remain unfulfilled.

### **The Model Agency**

The model agency, then, would be one where the agency and its authorizing environment (the politicians or board members the agency reports to):

- Accepted the impotence of top-down control for some tasks and in some environments;
- Accepted that very different management, control, and accountability systems were appropriate for different tasks and environments;
- Focused relentlessly on outcomes, meaning;
- Learning what actually has happened in a project and improving accordingly;
- Using a top-down measurement and control system where that was appropriate for achieving outcomes;
- Using Navigation by Judgment where it was appropriate for achieving outcomes
- Started from a default of trust – with the field in the lead, HQ supporting - where Navigation by Judgment is appropriate.
- Didn’t think the job was done at the initial policy change, but kept paying attention, evaluating, iterating. That treated management and agent’s *de facto* behavior and use of judgment as something to continue to pay attention to, to nurture.

So how do we move in the direction of that imagined ideal? In my view the first step is recognizing that we *all* have oars we can put in the water and start rowing in the right direction.

### **Section 3. Tuesday Morning at My Desk, and What Do I Do?**

At this point many of my students (usually in my career advising capacity) and coworkers (in policy roles, usually when hiring staff) have heard me invoke the “Tuesday morning” test. That is, it’s all well and good to be devoted to the vision and impact of an organization. But at some point, you will have worked there many months—or many years. And it will be

Tuesday morning, at 10AM.<sup>69</sup> When you think about the job (as an applicant) do you think that you will enjoy, or want to do, the things that you're doing?<sup>70</sup>

I think the same goes for organizational change. Below I lay out what I think might be done by people who work for or with aid agencies or are members of the public this coming Tuesday at 10AM to increase Navigation by Judgment.<sup>71</sup> I recognize I have none of these roles. And while each and every one of these recommendations has been said, or agreed would be useful, by *someone* (sometimes put a little more gently than I do here)—and often many people—in my conversations with and at aid agencies, does that make it right for you and your agency? No. Do I know what's right for you and your agency? Of course not. But I'd rather err here on the side of hubris than the side of milquetoast general calls to action. We've had more than enough of the latter in the discourse on aid reform, in my view.<sup>72</sup>

Few, if any, individuals can unilaterally change the organizations they work for in big ways. But they don't need to. I have often heard leaders complain that organizations are “ships that turn slowly”; but the key is they *do* turn. And while the captain needs the oarsmen to actually make the turn, the oarsmen actually don't need the captain. If one oarsman changes his stroke, and then others follow, that turns the ship, too. In other words, I think there are actions almost all of us can take that have the chance of – and in some cases are highly likely to – contribute to organizational change.

## **You're Pretty Junior in an Aid Agency**

Maybe you just graduated; maybe this is your first job. Congrats! You may think you have no power—but in some sense you have the *most* power. And that's why I'm starting with you—because you and your peers are, in my experience, most likely to (in my view) falsely believe you are powerless. You may feel like you have a lot to lose—but actually, you have the least, at least in some sense. They can fire you... but that's about it. And would you rather do

---

<sup>69</sup> I'm apparently not the only one with a Tuesday focus; a list of Tuesday-linked songs here:

<https://www.ranker.com/list/best-songs-about-tuesday/ranker-music>

<sup>70</sup>As an employer, of course, the question is whether the applicant will continue to enjoy/do a good job at the work. Once you start thinking in “contract theory” terms, employers and employees have a joint incentive to match employees and job roles where good performance is hard to monitor externally; this is part of the brilliance of “mission match” (Besley & Ghatak 2005). Both those on the “supply” and “demand” side of the labor market have an incentive to maximize the degree of “match”.

<sup>71</sup> After writing the first draft of this paper I discovered – thanks to Chris Roche and Janet Vähämäki – the concluding chapter of Eyben et al. 2015, which attempt something similar for aid professionals attempting to cope with the results agenda as practiced in their agencies (rather than change the practices of their agencies per se). It is striking how much overlap there is in tone/spirit, and even in a few cases in concrete suggestions – which further suggests to me that my experience talking to those in aid agencies reflects a more widely shared reality.

<sup>72</sup> Special thanks to Jeremy Weinstein for pointing out that I needed to say something more actionable somewhere, if I wanted anyone to actually take this on board (to be clear, Jeremy's responsible for the *idea* of proposing more actionable things, not the specifics here) and Alice Evans who encouraged me to appear on her podcast *Rocking Our Priors* and discuss my early ideas for same in December 2018. Episode here: <https://soundcloud.com/user-845572280/how-can-aid-agencies-foster-navigation-by-judgement-dr-dan-honig> if of interest.



something else and have the impact that brought you to the field, or stay in this agency? And wouldn't you rather figure it out *now*? And if you *do* stay, you're going to be in positions of increasing influence as the years go by. Your ability to alter your own work environment—and eventually that of others who work for and with you—will only grow with time.

- **Propose New Things You Think are Right.** You're also part of the conversation now about what gets done (how cool is that, eh?) Raise your hand, and speak your mind (with humility, when appropriate, and recognizing the long experience of others, of course). Also ask whether that metric we all think isn't-so-good is better than none at all. Ask questions that your relative youth and inexperience make it *easier* for you to ask; e.g. "I know that I'm the new person, and there's probably a good reason for this, but why can't we just use a different kind of accountability framework here?"
- **Learn the *Actual* Rules.** The next time someone says "you can't do it that way", maybe say "I'm sorry, I'm new here. Why not? What's the rule that governs?" It's a bureaucracy; if it's not written down it's a custom, not a rule.
- **Find a Champion.** Have a lot of coffees. Talk to other staff. Find someone who cares about the same things you do – if that's Navigation by Judgment awesome, but if not, the same advice holds. Find the person you admire most, watch, and learn from them how to navigate this agency, how to be the kind of person you aspire to be and make the agency work for you. And if you can't find one – if there are literally no people you admire and find inspirational in any more senior roles—get out now, before future-you becomes the kind of person present-you doesn't want you to be.
- **Take Some Risks.** All you've got to lose is this job. It's not a career yet; and if you lose this job, I suspect you can find another way to fulfillment. Better that than a life of paperwork and no impact, eh?

## **You're Pretty Senior in an Aid Agency**

Per Spiderman (actually the French Revolution as near as I can figure, but via Spiderman for most of us), "With great power comes great responsibility."<sup>73</sup> And if you're in a position to and want to, you can, of course, do some big things. But there are smaller ways to move this agenda forward, too.

- **Start a Pilot of Navigation by Judgment, or Form a Committee/Explore Same.** That is, basically: *try* Navigation by Judgment. Turn the reform lens your agency shines on others towards the mirror, and pilot a reform of organizational practice on the agency itself. Maybe pick an area where you think judgment is important and programs aren't doing well—civil service reform, or your work in a particular country. You pilot things all the time at your agency—why not pilot new

---

<sup>73</sup> <https://quoteinvestigator.com/2015/07/23/great-power/> for the background on the quote.

management and accountability practices? You can sell it the same way you do RCTs, I think—we're trying something we have good reason to think *might* work, to find out if it will. Do marry the pilot with rigorous (possibly 3<sup>rd</sup>-party) evaluation; the knowledge you generate will help not just your programs, but global knowledge.<sup>74</sup>

- **Commission Analytic or Diagnostic Work on Where Metrics Aren't Giving an Accurate Picture of Success, and Where/When/How Rules and Reporting Are Getting in the Way of Impact at Your Agency.** Find out if indeed your agency has a problem. Commission an external firm to come in and create an agency-specific diagnosis of your reporting and accountability systems, and to explore if indeed impact is undermined (that is, these constraints are precluding behavior you think is in fact useful/good).
- **Talk to Authorizers (Politicians, Board Members) About When Pursuing Measurable and Reportable Results is Undermining Your Ability to Deliver Results:** Authorizers (MPs, Congress, Executive Board Members, etc.) are, in fact, human beings. Many—I think probably most—of these human beings *also* would like to see the money they spend translate into actual impact, not its façade. I recognize that you think I sound naïve here. And maybe I am. But I have yet to meet anyone on the authorizer side—even the most devoted aid skeptic—who would rather, contingent on a given project being funded, that it *fails*. Maybe they would have preferred it not happen; but if it happens, who precisely wishes it not to succeed? I may favor aid because of its welfare impacts; but even someone who thinks aid is only good because it 'wins hearts and minds' is likely to agree it will win *more* hearts and minds if it *actually accomplishes the stuff it sets out to do*. Build trust, or at least understanding, through dialogue. Is this easy? No. But if things are going to change, you and your authorizers need to both be part of the conversation. If you want to run a pilot, talk it through with them. If you commission diagnostic work, use it as a basis for conversation. Get talking.
- **Create "Track II" Channels Between Authorizers and Your Staff:**<sup>75</sup> I have yet to encounter an agency that I think is using all the flexibility its authorizers intended it to have; if authorizers intend X amount of flexibility/scope for autonomous action, it seems to me most agencies are using somewhere between 70 and 90% of that X. Agencies are afraid to get close to the boundary; after all, the risk is asymmetric. If you misunderstand and go *beyond* X, that will work out poorly. Better to create some buffer. But how much? And why not get clarity where you can? I have occasionally served as a go-between for both bilateral and multilateral agencies between staff and authorizers; that is, someone at a middle level in an agency calls me (always a call) and says "we're not sure if we're allowed to do this". So I ask

---

<sup>74</sup> Particularly if you do indeed do an RCT of a Navigation by Judgment pilot, which I think is entirely possible – pick an eligible population of countries/projects etc., and randomly assign them to conventional management or Navigation by Judgment. Yes, it will take a while to see if Navigation by Judgment is in fact helping – but you run RCTs already that take a long time to show results.

<sup>75</sup> A brief explanation of the term and its evolution here: <https://foreignpolicy.com/2011/06/20/track-ii-diplomacy-a-short-history/>

someone on the authorizer side; and as often as not, the answer is “yes, sure, that’s not what we intended to preclude in the regulation/instruction/etc.” *This is an absolutely crazy way to run an aid agency – or any agency.* The basic problem is that for someone in the agency to talk to the authorizer to get clarity, they have to escalate the query to a senior level; that requires making it a *big* deal, and probably using some capital. Better to play it safe, they think. Which is entirely reasonable; and also means that things agency staff *want* to do, that authorizers *also want them to do*, don’t get done – simply because of a lack of information. Establish backchannels; allow your agency staff and authorizers’ staff (ED offices, or MP/Congressional Staff, etc.) to talk to each other – to clarify information asymmetries. If the Soviets and US could make this work during the Cold War, surely it’s not impossible for you and your agency.

- **Identify What the Rules Allow and Highlight This for Staff:** I’ve never seen a donor who didn’t have unused flexibility in their own *existing* rules, sometimes unbeknownst to their own senior management. Find the flexibility in your own rules – or hire an intern to ask the questions that find it. Once you’ve found it, make it public. Point it out.
- **Build Knowledge Management on Surprising Success and Failures of Attempts to Navigate by Judgment, and Use Experiences Both Inside and Outside the Agency.**<sup>76</sup> The USAID/OTI Liberia story with the co-location of USAID and contractor staff, which I see as a marked and surprising success, is written up nowhere in USAID, to my knowledge. This seems like really low-hanging fruit; task someone in M&E with the job of codifying surprising successes, positive deviants of management practice. If you can, make it two sided – think about failures too. The SRO/DFID/Smart Rules story is, as far as I know, also written down nowhere in official documents, and I think also has much to teach us.
- **Figure Out How to Provide Cover for Judgment, Where Appropriate:** Of course, pointing out what the formal rules allow is just the beginning. How do you encourage their use – making that use, to crib from ICAI one last time, “attractive”? I have literally never had a senior manager at an agency disagree with the statement “A big part of any manager’s job is shielding their direct reports from things coming from ‘above’ that could get in the way of good work.” That is, good management means protecting your team from requests from your boss that could screw things up. So what do you need to do to similarly protect the whole agency? What about if you go away, or the Government changes, etc. – how can you overcome the natural (and appropriate) risk aversion of civil servants, who have their careers to protect? Maybe you can figure out a way to “require” a new practice, in a way that makes its use more defensible even if the tide rolls out. Maybe performance evaluation can shift from “project” to “portfolio”, encouraging folks to try out things they’re not sure will work in any given case but are likely to make things better overall. The answer is surely agency specific.

---

<sup>76</sup> Thanks to Sarah Rose for this suggestion.

- **Find and Celebrate Navigation by Judgment Where it Exists:** One straightforward way of encouraging the behavior you want is to find those who are *already* acting that way. Maybe they’ve been doing it in the shadows; bring it into the light. Find the kinds of practices you want to encourage more of – the ‘positive deviants’ who have been pushing the envelope. If you don’t know where they are, ask them to self-identify; follow the USAID model in their Collaborating, Learning, and Adapting (CLA) case competition, have people apply.<sup>77</sup> The awards will encourage those already doing good things; but maybe even more importantly, they will send an important signal to others to do similarly.
- **Think about the “Fuzzy Stuff” and Trust your Own Judgment About It:** You probably know the folks who you’d want on your team if you were still running projects. How do you get more folks like that? How do you encourage your staff to be *more* like that? I don’t know the answer, but you do. How do you manage so as to create an environment that encourages the staff you want, retains the brilliant folks you already have, and attracts the next generation of earnest, humble practitioners who you want in the “foxhole” with you? Don’t just think about today’s staff and today’s structure. Start managing differently, and system will change over time, as people change.
- **Don’t Allow the Problematic to be Framed as “Real” Accountability vs. None.** Finally, per the theme of Section 2 of this paper, push back on the idea that only numbers can create “real” accountability. The sense of accountability those metrics are currently providing is sometimes a pyrrhic victory. The certainty may be false. Recognize reality: we often don’t know, and can’t judge, what’s going on. Navigation by Judgment is a step towards, not away from, understanding reality. That requires thinking about alternate – probably trust-based, and more narrative – accountability.

## **You’re Somewhere in the Broad Middle of an Aid Agency**

If this applies to you, and you’re still reading (thanks! I promise the next few sections are shorter), you likely read the last section thinking in part “great, but I don’t have that kind of power.” Maybe not; but I bet some of what’s true of more senior or more junior staff could *still* apply to you. You may *also* be able to commission diagnostic work or propose same (in your program, or country, etc.) In your interactions with implementers there may be a need for a “Track II” in a way that parallels the agency-authorizer divide. You may *also* be able to protect your direct reports a bit more, to think about what they may need. And there are additional things you’re actually maybe better placed to do. For example:

- **Challenge Assumptions You Think May Not Be Accurate.** Learn your organization’s rules; the next time someone says in a design meeting “Well, we need to have a measurable output”, push back on that if you think that’s not accurate. Or

---

<sup>77</sup> <https://usaidlearninglab.org/cla-case-competition>

if one of your colleagues says “It’s not a great measure but it’s the best we’ve got, let’s go for it”, you could suggest that maybe sometimes if we *don’t* have a great measure, the measure might end up hurting the project.

- **Start Some Dialogue about Navigation by Judgment, or “the Rules,” or Metrics.** With your bosses; with your staff; with your peers. Start the conversation; open space for dialogue and thinking. I bet you already think at least some folks in your organization see this the same way you do, from my experience talking with folks in agencies; and I suspect however frequent you think that view is, it might be an under-estimate. Maybe propose a lunch conversation about when rules and metrics are getting in the way, and see who turns up and what strategies (with the existing rules) they have for mitigating if not eliminating these challenges. It doesn’t need to be a movement (though great if it is); it can just be lunch. As my first boss in DC taught me, “everyone needs to eat lunch. Why not with you, about this?”<sup>78</sup>
- **Get Involved in an Existing Forum or Community of Practice.** Many agencies now have internal groups thinking about adaptive management, or thinking and working politically, or doing development differently, etc. If not, fora like the Global Delivery Initiative<sup>79</sup> bring together aid professionals from donors and implementers to discuss common challenges. Get involved in one of these, and think about how to apply the community’s insights and energy to encourage positive change in your agency’s management practices.
- **Think Hard about When—and How—to Contract.** Think about when the “contract” itself serves your purpose. Contract theory suggests that the more uncertainty in what needs to be done, the more likely contracting out is likely to cause problems rather than prove net useful. This is because writing a contract requires specifying a whole lot of things; this may prove inappropriate in unpredictable settings, or where “what you need to get done” might change because of the nature of the task. What can be left flexible, or out of the contract? When are there alternatives to traditional contracts?
- **Pay Attention to “Pass Through” to Implementers.** Flexibility at the top allows, but doesn’t guarantee, flexibility at the bottom. Think about where the flexibility (or constraint) inside your agency is affecting implementers, and what you may be able to do within the existing rules to give implementers – either your staff or contractors – more wiggle room, more ability to use their judgment and incorporate their perceptions into their work where appropriate. I’ve heard at least one story where a simple change of frame accomplished much – with substantial effects flowing from a change for a USAID field officer in beginning quarterly meetings with “are you on target?” to “what have you seen changing, and how are you

---

<sup>78</sup> Many thanks, Mark Schoeff (CSIS), wherever you are. The case in point was getting press releases written, circa 1999.

<sup>79</sup> <http://www.globaldeliveryinitiative.org/about-us>

adjusting to that?"<sup>80</sup> Your frames and words have more power to affect the implementers you interact with than you may realize.

- **Talk to the Implementers You Work With.** It is sometimes (and in some agencies often) your partners rather than your agency's staff who are actually on the ground observing the projects – and thus it is from them that you can learn what's working, what isn't, and where the rules are getting in the way. That is, you may play a critical role—as the link to implementers – in understanding what changes in agency management or reporting or accountability practice as it relates to implementers might both help unlock more of the gains of Navigation by Judgment *and* be acceptable/feasible for your agency.

## **You Have Another Role in or Near the Aid World**

Maybe you don't work at an aid agency. If you have a desk, and also have a Tuesday morning, I think there are probably things you can do, too.

### **Contractors/Implementers**

People at various levels inside large aid contractors can of course borrow from the respective lists above as regards your internal management, how you manage subcontractors, and how you relate to funders. But there are other, contractor-specific things, you're actually *better* placed to do, too. That is, you're where the work actually gets done, and your thoughtfulness and judgment are already built into the system – but also are limited (to the program's detriment, as almost every implementer I've ever spoken to agrees) by constraints that flow from your funders. Point them out, individually and collectively – that is, in concert/coordination with other implementing organizations. And, of course implementer coordination bodies (by which I mean Australia's ACFID, US' Interaction, UK's Bond, etc.) can *also* commission analytic work and identify problems, and in so doing provide some buffer to any pushback from the agency whose contracts you implement to anything that report might find (that is, lowering the chance of negative consequences for your firm).

If you're involved with writing bids for new work, see if you can find a bit you think you're *unlikely* to win, sit down with your team, and imagine a new way of working that's more consonant with *Navigation by Judgment*. You might just win the bid; but I bet even if you don't, the act of coming together and talking about what new ways of working would look like will be time well spent, and might lead to some other awesome things.

Don't neglect your own power in individual project management, either. As Pedro Prieto Martin recently put it in the context of managing adaptive programs, "We [implementers] need to find ways to tweak and leverage the current donor's system, their existing tools and processes, so that we can progress. One programme, one logframe, one risk matrix at a time. Until it becomes a trend. The system will not change by just being proved that it does not

---

<sup>80</sup> Many thanks to David Jacobstein for the anecdote.

work, that it is wrong. It is required to demonstrate that it can work better... by making it work.”<sup>81</sup>

### **Authorizers**

If you are, or work for, an authorizer – a member of the legislature, or a committee, or an executive director – you should build the bridge in the “other” direction. Reach out to staff in the agencies. Propose ‘track II’. Ask questions (of your principal, or of the agency) like “are controls getting in the way here?” or “Is this ‘real’ accountability or a mirage? or “are these ‘real’ results or just a façade? Because no numbers at all are better than fake ones that undermine the work.” Provide some more space for the agency, and see what flourishes. You can also, of course, actually request a pilot of *Navigation by Judgment*, or analytic work on the constraints the agency faces; indeed, if you provide the cover for this, I have no doubt this is one request where the agency will be very, very happy to comply.

### **Freelance Consultants**

You’re critical to sharing knowledge about what other NGOs, agencies, etc. are doing. You’re the ‘blood’ of the aid system; you circulate most seamlessly between the parts. Share your observations and information.<sup>82</sup> Connect people you meet at different organizations who might be allies to each other. Help the system overcome pluralistic ignorance.

### **Other Observers (Scholars, Private Sector, Many Others)**

Economists and econometrically oriented political scientists tend to still like this project despite the fact that it deals with pretty messy things and uses observational data. Why? In part, I think, because if you’re an economist focused on development, you’ve almost certainly taken money from, evaluated the programs of, and/or interacted with donor field agents. And that means many scholars have seen, with their own eyes, the obsessive measurement and control of aid staff in ways that undermine programs. As Princeton’s Jake Shapiro, who as a scholar of conflict has dealt with lots of donors, said when I first explained the book to him something like “You’re obviously right – I mean, this is obviously true. What data do you use?” Jake’s priors were strongly in favor of his claim – because of what he’s seen with his own eyes.

I believe this view is widely shared; and thus while I believe that my analysis is indeed pretty rigorous, the fact that some of those in a seminar room had dealt with aid agencies probably helped shape their reactions to this book. But the system itself would *also* benefit from this collective, if subjective/perceptual, data. Jake, and other scholars who have interacted with aid agencies, are collectively an important potential source of influence. And it’s not only researchers; private sector actors in contexts with a lot of aid, members of the developing world public, developing world Government officials, and many many others have enough exposure to aid agencies to provide observations on this, to raise their voices, to help change

---

<sup>81</sup> Email to the “adaptdev” google group, December 18, 2019. Used with author’s permission.

<sup>82</sup> Information you’re not contractually barred from sharing, of course. I am not advocating ignoring NDAs and other confidentiality agreements.

the system. Close observers can help by speaking out, by convening “Track II” channels where they see clear information problems, and by realizing their own agency in changing aid agencies for the better.

## **You’re Not in the Aid System**

You don’t work for, or around (e.g., as or for an authorizer, or at a think tank, or as a consultant, etc.), aid. Your office is a coffee shop because you work there on your laptop, doing something else. Or you’re a barista. Or a corporate lawyer. Or an electrician. In any case, you have no role, you say, in the aid system, but you’re seeing this text because you have an interest in global development.

First, you still *are* part of the system in some sense. It’s almost certainly your tax money being spent, or your government doing the giving, and/ or your government on the receiving end. What you say, think, and do doesn’t just matter—on some margins, it matters *more*.

If you’re a voter, for example, hearing from constituents *does* change minds. And I think there’s a strong case to be made by members of the donor public in this direction, along the lines of: “What you’re asking aid agencies to report is lower the value of MY tax dollars. I don’t want the report; I want the impact. Please get out of the way.” Write a letter to your representative; or better yet, start a campaign. Schedule a Meetup; bring this up at organizations you’re already a part of; please, help foreign aid meets its laudable goals and escape the suffocating controls that are put on it by invoking “your” name, in service of the façade of accountability.

## **This Coming Tuesday: Let’s Get Moving**

The things in this section are a hodge-podge of “big” and “small.” Even for folks who have the power to do so, proposing a big change still requires a lot of capital, and some risk. But this isn’t a “go big or go home” situation; there are smaller things that can also move agencies in the right direction. Agencies change because of what’s done Tuesday at 10AM; indeed, there are few other ways to change them, in my view.<sup>83</sup>

If my wanderings are any guide, you might find some surprising allies inside the building your desk is in if you *do* do something this Tuesday. More people likely agree with you than you think. Alice Evans’ work highlights the importance of “pluralistic ignorance”, finding that for activists thinking about activism; for marginalized groups thinking about mobilization; and for women thinking about men sharing care work a lack of realization about (1) how widely shared one’s views are and that (2) change is possible limit the pace

---

<sup>83</sup> And if you want to do something on the bigger side, and I can help with the design etc., please do holler if I can be of use thinking it through. That’s why I wrote this piece – and the book more generally – after all.



and extent of social change.<sup>84</sup> Evans frames this as a “despondency trap”; individuals don’t see anyone else voicing their views or pushing in the direction they think appropriate, so they assume others don’t share their views.<sup>85</sup>

From the very top to the very bottom of every aid organization I’ve spoken to, I hear consonant messages. I hear shared worries. And I hear over and over again from the speaker that most people don’t agree with them, that nothing can change. And I think that’s just got to be wrong, given the number of people who have voiced it, and the shared view across so many organizational layers. That is, I haven’t spoken to enough people to know that “most” is empirically false; it may still be that a majority of people disagree. But (1) I don’t think so and (2) it is at the *very* least a very substantial minority of people who feels this way. Far more than I think are likely a sufficient critical mass to bring about change.

I believe the aid world is awash with “pluralistic ignorance”—those in the aid world are unaware of how many others agree with them. “Others” here means staff; authorizers; management; implementers. And this does indeed lead to a “despondency trap.” Aid agencies should, can, and will change—if we work to change them.

I strongly suspect if we all start pushing, the current accountability and management regime will, where it does not serve results, crumble. The status quo serves no one; not authorizers (who are being ‘lied to’ by numbers, and often know it); not the agencies (who cannot implement programs they think would be beneficial); not staff (who are demotivated by rules that serve no productive purpose); not the beneficiaries (who see less benefit than could be the case); and not the taxpayers and donors who supply the money (whose money is put to less effect than could be the case). It is time for us all to start rowing, to point out when and where the metrics-based accountability emperor has no clothes and what a better agency might look like.

## Conclusion

The currently dominant accountability system in use in aid agencies gives a false picture of value for a wide range of projects where what is countable is not an accurate summary of the project as a whole. If this led *only* to inaccurate perceptions of success, however, that would in fact be an improvement. Theory and evidence strongly suggest that arguably the primary way donors are asked to demonstrate “Value for Money”—with reportable, countable, accounting-based accountability reports of success—can in fact *undermine* the “Value” part of that equation.

In my talks at aid agencies, I have many times been told some version of “Well, you know, we have a phrase here: We need to make every dollar/pound/Swedish crown/euro/penny

---

<sup>84</sup> These pieces are Evans 2019; Evans 2018; and Evans 2016, respectively. See Bursztyn, Gonzalez, and Yanagizawa-Drott 2018 for an empirical econometric example, and Duque 2017 for a more general modeling of related problems.

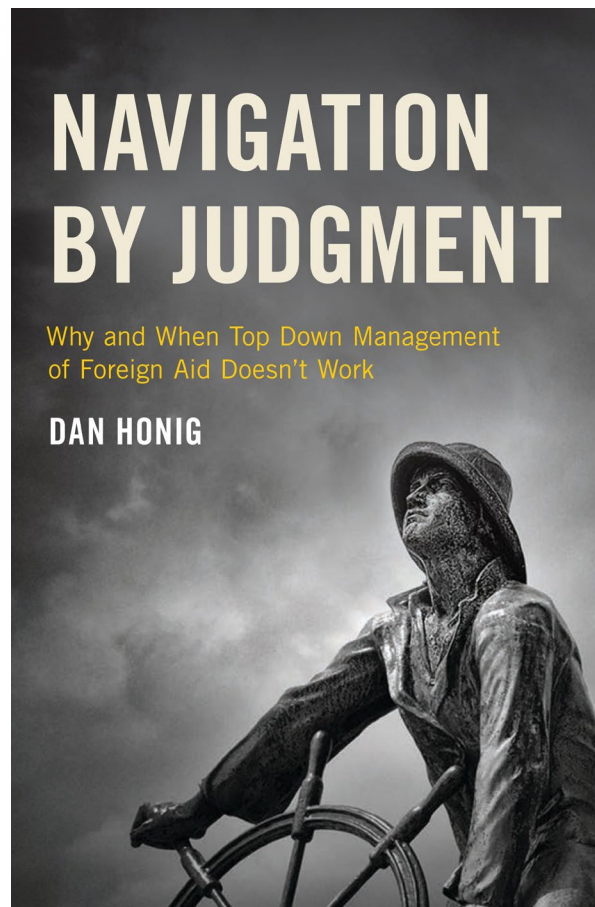
<sup>85</sup> See Evans 2019. This is a kind of collective action problem in which asymmetric information regarding the preferences of others plays a large role in the failure to mobilize.

count.” It is perfectly natural to think that the best way to make every penny count is to always in fact count every penny, to try to make everything monitorable and measurable. It is perfectly natural—but it is absolutely wrong, and for many projects counter-productive. Making every penny count is *not* coterminous with counting every penny.

Making every penny count requires, at least sometimes, letting go of a bit of control and building a system that trusts, empowers, and supports the people within. And I think building such a system in turn probably means accepting something like the five principles articulated in section II. Agencies need to give up the delusion of fiat control; accept that accurate knowledge of results is superior to the reductive façade of inaccurate metrics; recognize the need for different accountability systems for different tasks and environments; start from a default of trust in building those alternate accountability systems; and keep tweaking to make the new system work.

At aid agencies I often closed my book talks with a slide of the cover of the book (Figure 5).

**Figure 5. Cover of *Navigation by Judgment*<sup>86</sup>**



---

<sup>86</sup> Honig 2018; Photo credit Marty Luster.

I'd usually make a joke about people saying "don't judge a book by its cover"—but me having no problem with you judging this book, and this argument, via this cover.

The image is of a memorial in a town called Gloucester, Massachusetts in the US. It's where the book (and movie) *The Perfect Storm* is set, if that means anything to you.<sup>87</sup> It's a fishing community—one where people go out to sea, and very tragically some of them perish. As a result, the community mourns them and builds statues to them. But what they, and we, don't do is say that because they exercised judgment and failed, we should remove the capacity to make judgments from that whole class of actors. No one thinks that because sometimes fishing boats capsize, the right way to steer the fleet is to have a headquarters on the land, GPS trackers on the boat, and treat it like a kind of drone video game. We know that the best way to run a fishing fleet is to have the people who have their hands on the wheel be the people who are actually on the water.

Judgment fails. It fails all the time. It fails in myriad ways. If we rely on human judgment, there will be failures. But "Will judgment always work" is the wrong question. The question should be "Fallible judgment as compared to what?" Fallible judgment is not properly thought of in this comparison as an alternative to perfect metrics, perfect top-down controls. Because top-down controls and metrics fail, too. All the time. In myriad ways.

Fallible judgment ought to be compared to also-fallible, and sometimes-delusional, top-down control. The right question is, in my view, when each of these two ultimately fallible broad strategies is going to lead to better results.

I see, as I move through aid agencies, so so many dedicated professionals who genuinely care about their organization's goals, who want to achieve the welfare impacts their organizations seek. Many of these folks are thwarted by their *own organization's rules and practices* in executing on their organization's mission. This paper aims to present some ideas for what people working for aid agencies, in the broader aid industry, and beyond may be able to do to make things better for themselves and their colleagues in pursuit of the impact they seek.

I spent years working on a research project, and ultimately a book, because I thought the world needed evidence that there was a problem here. I now think I was wrong about that; the problem is widely recognized by insiders. What's needed are solutions. This paper attempts to provide the outline of some. I think there's much more to be done, though, in thinking through that "new paradigm of accountability" in practice,<sup>88</sup> and how to ensure that any changes in organization vision or rules actually achieve their object. There's also much more to be said and thought on relations with contractors, procurement standards, HR, and risk management, and a unique set of problems and rules and complexities in each and every

---

<sup>87</sup> Junger 1997. The movie is a 2000 release from Warner Brothers.

<sup>88</sup> Including but not limited to how to build horizontal accountability channels between peers, or if, where, and when "reversing the arrows" so aid agencies can be made accountable to recipients, amongst many other important questions. That is, this paper leaves to the side quite important questions about how one might build successful accountability alternatives. Roche & Denney 2019 develops suggestions in these regards, as do various chapters of Eyben et. Al. 2015.

agency. Real change needs to at least incorporate, if not flow from, the details and specifics of each case.

In my view, most of what I've written here is true not because aid is a special snowflake in the organizational world, but because it's the same. That is, organizations—public, private, nonprofit, developing world, developing, etc.—that is, any time lots of people get together to try to do something—face a trade-off between navigation from the top and Navigation by Judgment; between.<sup>89</sup> In this sense, aid is the rare domain where we can see effects across country, sector, etc.—just because aid agencies do so many things, in so many places. Aid is the potential canary in the coal mine, reminding us that for all their benefits, technology and data can also give us the illusion of control, resulting in the pyrrhic victory of more targets and dashboards, but less actual success.

The top of Maslow's hierarchy of needs is self-actualization. Action that helps agencies allow more Navigation by Judgment is doubly “actualizing”—helping motivated individuals do more of what they are capable of, as these individuals in doing so also help their agencies do more of what they are capable of. The world has rightly become more skeptical of supposed “win-wins” in recent years; of arguments that suggest that all may gain. But where Navigation by Judgment is more appropriate, more fulfilling jobs are indeed totally consistent with—indeed, help forward—the ends of the organization's managers.

To get there, organizations need to change. And—to state the obvious, but perhaps oft ignored—for organizations to change *something* about the organization needs to change. Organizations change when their personnel change, or culture changes, or HR practices change, or operating procedures (*de facto*, not just what's in the manual) change. They change when people feel more comfortable taking risks, or buy into the mission of the organization more (or less), or feel they can have more impact. Not on this list is “when they announce they're going to change”. Telling an organization to “be more adaptive” or “use more judgment” is like instructing it to “be more successful” or “do better.”

When my five-year-old struggles hitting a baseball, I don't tell him to “swing better” in an attempt to help him achieve the performance he seeks. I help him remember to square up; I help him with his stance, and his swinging motion; I think about whether he's got the appropriate bat. These details aren't part of his swing—they *are* his swing.

In the relevant sense, an organization is also composed of its details. All of us in the aid world just need to move each of our elbows a little; or pay attention to the pitcher's release point; or turn our back foot just a little bit. Small adjustments can yield big results, if they're the right small adjustments. And if one small change begets more success and that in turn begets still more... well, we might just end up with aid agencies that fulfill more of their objectives, and people at those agencies who feel more fulfilled working for them.

---

<sup>89</sup> I mean this to be roughly parallel to Aghion & Tirole's (1997) tension between principal control and agent initiative. See Honig 2018 for a fuller discussion.

Will change be easy? No, of course not. But most people working in foreign aid have already chosen something other than the easiest path for their lives, because they care about the organization's mission. We owe it not just to these professionals, but to the welfare of beneficiaries, to do everything we can to build aid agencies worthy of the people who populate them.

## References

- Alchian, Armen, & Harold Demsetz. 1972. "Production, Information Costs, and Economic organization." *The American Economic Review* 72:5, 777-795.
- Aghion, Philippe, & Jean Tirole. 1997. "Formal and Real Authority in Organizations." *Journal of Political Economy* 105(1):1-29.
- Bain, Katherine, David Booth, and Leni Wild. 2016. *Doing Development Differently at the World Bank*. London: Overseas Development Institute.
- Barnett, Paul. 2015. "If What Gets Measured Gets Managed, Measuring the Wrong Thing Matters." *Corporate Finance Review*, January/February: 5-10.
- Besley, Timothy, and Maitreesh Ghatak. 2005. "Competition and Incentives with Motivated Agents." *American Economic Review*, 95 (3): 616-636.
- Birdsall, Nancy, Ayah Mahgoub, and William D. Savedoff. 2010. "Cash on Delivery: A New Approach to Foreign Aid." *CGD Brief, November 2010*. Center for Global Development.
- Birdsall, Nancy, and William D. Savedoff (with Ayah Mahgoub and Katherine Vyborny). 2010. *Cash on Delivery: A New Approach to Foreign Aid with an Application to Primary Schooling*. Washington, DC: Center for Global Development.
- Bursztyn, Leonardo, Alessandra L. González, and David Yanagizawa-Drott, 2018. "Misperceived Social Norms: Female Labor Force Participation in Saudi Arabia," REPEC Working Paper 2018-042, Human Capital and Economic Opportunity Working Group.
- Campbell, Susanna. 2018. *Global Governance and Local Peace*. New York: Cambridge University Press.
- Chenoweth, Erica & Maria Stephan. 2011. *Why Civil Resistance Works: The Strategic Logic of Nonviolent Conflict*. New York: Columbia University Press.
- Dewatripont, Mathias, Ian Jewitt, and Jean Tirole. 1999. "The Economics of Career Concerns, Part II: Application to Missions and Accountability of Government Agencies." *The Review of Economic Studies*, 66:1 199–217.
- DFID 2019. "Smart Rules: Better Program Delivery." Version XII. Available at: [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/840802/Smart-Rules-External-Oct19.pdf.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/840802/Smart-Rules-External-Oct19.pdf.pdf)
- Dixit, Avinash. 2002. "Incentives and Organizations in the Public Sector: An Interpretative Review." *The Journal of Human Resources*, 37(4), 696-727.
- Drucker, Peter. 1954. *The Practice of Management*. New York: Harper.
- Duflo, Esther. 2017. The Economist as Plumber. *NBER Working Paper 23213*.
- Duque, Mauricio Fernandez. 2017. "Pluralistic Ignorance and Social Change: A Model of Conformity to the Perceived Majority." Working paper, available at: [https://scholar.harvard.edu/files/duque/files/pluralistic\\_ignorance\\_and\\_social\\_change\\_new\\_1027.pdf](https://scholar.harvard.edu/files/duque/files/pluralistic_ignorance_and_social_change_new_1027.pdf)
- Evans, Alice. 2019. "Overcoming the global despondency trap: strengthening corporate accountability in supply chains." *Review of International Political Economy* (advance view).
- Evans, Alice. 2018. "Politicizing Inequality: The Power of Ideas." *World Development* 110, 360-372.

- Evans, Alice. 2016. "The Decline of the Male Breadwinner and Persistence of the Female Career: Exposure, Interests, and Micro–Macro Interactions." *Annals of the American Association of Geographers* 106:5, 1135-1151.
- Eyben, Rosalind, Irene Guijt, Chris Roche, & Cathy Shutt (eds). 2015. *The Politics of Evidence and Results in International Development: Playing the Game to Change the Rules?* Rugby, UK: Practical Action Publishing.
- Gulrajani, Nilima and Linnea Mills. 2019. Fit for Fragility? An Exploration of Risk Stakeholders and Systems Inside SIDA. EBA Report 2019:02, *Expert Group for Aid Studies (EBA)*, Sweden.
- Herd, Pamela, and Don Moynihan. 2019. *Administrative Burden: Policymaking by Other Means*. New York: Russell Sage.
- Hernandez, Kevin, Ben Ramalingam, and Leni Wild. 2019. Towards Evidence-Informed Adaptive Management. Working Paper #565, *Overseas Development Institute*.
- Holmstrom, Bengt, and Paul Milgrom. 1991. Multitask Principal- Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design. *Journal of Law, Economics, & Organization* 7: 24.
- Honig, Dan, and Nilima Gulrajani. 2017. "Making Good on Donors' Desire to Do Development Differently." *Third World Quarterly*: 1– 17. doi:10.1080/01436597.2017.1369030.
- Honig, Dan, Ranjit Lall, and Brad Parks. 2019. "When Does Transparency Improve Performance? Evidence from A Database of 23,000 Foreign Aid Projects." Working Paper.
- Honig, Dan, and Catherine Weaver. 2019. "A Race to the Top?: The Aid Transparency Index and the Social Power of Global Performance Indicators." *International Organization* 73:3, 579-610.
- Honig, Dan. 2018. *Navigation by Judgment: Why and When Top-Down Control of Foreign Aid Won't Work*. New York: Oxford University Press.
- Honig, Dan. 2019a. "When Reporting Undermines Performance: The Costs of Politically Constrained Organizational Autonomy in Foreign Aid Implementation". *International Organization* 73:1, 171-201.
- Honig, Dan. 2019b. The Power of Letting Go. *Stanford Social Innovation Review*, Winter 2019.
- Honig, Dan. 2019c. Let Local Leaders Lead: Why Donors Should Create More Space for Local Leadership. CGD Blog, September 27. Available online at <https://www.cgdev.org/blog/let-local-leaders-lead-why-donors-should-create-more-space-local-leadership>
- Independent Commission for Aid Impact (ICAI). 2014. Rapid Review of DFID's Smart Rules. Report 39: December 2014.
- Junger, Sebastian. 1997. *The Perfect Storm: A True Story of Men Against the Sea*. New York: Norton.
- Kenny, Charles. 2017. *Results Not Receipts: Counting the Right Things in Aid and Corruption*. Washington, DC: Center for Global Development.
- Laskas, Jeanne Marie. 2019. "The Mister Rogers No One Saw." *The New York Times Magazine*, November 24: 46.

- Mansbridge, Jane. 2014. A Contingency Theory of Accountability. In *The Oxford Handbook of Public Accountability*, edited by Mark Bovens, Robert E. Goodin, and Thomas Schillemans, 55–68. Oxford University Press.
- Muller, Jerry. 2018. *The Tyranny of Metrics*. Princeton, NJ: Princeton University Press.
- OECD Development Assistance Committee. 2013. *Managing and Measuring for Results: Survey Highlights*. Paris: OECD. Online at: <https://www.oecd.org/dac/results-development/docs/Managing%20and%20Measuring%20for%20Results%20-%20Survey%20Highlights%20-%20Final.pdf>
- Roche, Chris, and Lisa Denney. 2019. How Can Developmental Leadership be Supported? Foundational Paper 4, *Developmental Leadership Program*.
- Roche, Chris. 2018. Simplicity, Accountability, and Relationships: Three ways to ensure MEL supports Adaptive Management. *From Politics to Power (Oxfam Blog)*, July 3. <https://oxfamblogs.org/fp2p/simplicity-accountability-and-relationships-three-ways-to-ensure-mel-supports-adaptive-management/>
- Sida. 2002. *Action Programme: Vision for a Strengthened Field Orientation*.
- Silverman, Rachel, Mead Over, and Sebastian Bauhoff. 2015. Aligning Incentives, Accelerating Impact: Next Generation Financing Models for Global Health. *Center for Global Development*.
- Springer, Emily. 2019. Bureaucratic Tools in (Gendered) Organizations: Metrics and Gender Advisors in International Development. *Gender & Society*.
- Stimson, Henry L. 1945. Memorandum on the Effect of Atomic Bomb.
- Tillitsdelegationen. 2018. October 2018 Report Summary (English). Available online at <https://tillitsdelegationen.se/wp-content/uploads/2018/10/english-summary.pdf>
- Vähämäki, Janet, and Chantal Verger. 2019. Learning from Results-Based Management Evaluations and Reviews. OECD Development Co-Operation Working Paper 53, *OECD*.
- Vähämäki, Janet. 2017. *Matrixing Aid: The Rise and Fall of 'Results Initiatives' in Swedish Development Aid*. Dissertation, Stockholm University.
- Vähämäki, Janet. 2015. The results agenda in Swedish development cooperation: cycles of failure or reform success?. In Eyben, Rosalind, Irene Guijt, Chris Roche, & Cathy Shutt (eds). *The Politics of Evidence and Results in International Development*, Chapter 8, 135-154.
- Wilson, James Q. 1989. *Bureaucracy: What Government Agencies Do and Why They Do It*. New York: Basic Books.



## **Annex I. Some Answers to Questions That Might Arise Reading Section 1, if You Haven't Read the Book**

### *A Bit More on the Downsides of Performance Measurement*

I mean in part that aid has what economics would call a multitask problem; you have somebody and you want them to do some things which you can measure well and some things which you cannot measure well.<sup>90</sup> If you measure the things you can measure well, you will get more of those things, but you will get less of the other things. People will invest in the observable and not invest in the unobservable.

In addition, Nobel laureate Jean Tirole and Philippe Aghion tell us that if we don't give people close to the ground the power to use what they learn as they're engaging in practice, they will not invest in gathering the information the organization needs.<sup>91</sup> That is, if I have a job, even a job I care about, but I know that making the costly investment to figure out what's going on in nuanced ways is a waste of my time – because I can't make use of whatever I learn – I have no incentive to learn anything.

### *A Bit More on Fallible Judgment*

Well, in development there's often this image that the reason we might not want to trust agents is because we might have bad agents who want to do bad things. I don't think that's very frequent in the world, but human judgment is nonetheless inherently fallible. We *all* make mistaken judgments. Me; you; everybody. Behavioral economics is arguably an entire discipline focused on all the ways that well-intentioned human judgment often goes awry. Human judgment is fallible and if we rely on it, we're going to have to accept the fallibility of the judgment that they steer with.

### *I Also Think There's Another Big Reason for the Error of "Too Much Control" Dominating.*

The error of too little control means that agents have discretion, have autonomy that they use and sometimes make mistakes. We're going to see those mistakes and we're going to be able to adjust against them. The error of too much control means that the rules are going to prevent people from doing good things they would have done but could not do because the rules got in the way. We can't see good actions that could have been taken but weren't – we don't see the project gains that go unrealized because people were unable to take action that would have improved the projects. As a result, then, organizations are going to react to the problem they see – they're going to focus more on the error of too little control, and respond to any examples of mistaken judgments by further tightening control. This makes the problem of "too much" control worse, of course – but this will go unobserved, and thus

---

<sup>90</sup> Holmstrom & Milgrom 1991

<sup>91</sup> Aghion & Tirole 1997

it's very unlikely a politician or critic will ever ask them why they're being so controlling. Most organizations are going to err on the side of too much control more often than too little – responding to the problem they can see, rather than the countervailing problem they can't.

### Qualitative Evidence in VERY Brief

*On political authorizers constraining agents:*

Interviewees implementing projects in the field for USAID often talk about Congress and reporting up the chain, much more so than do those implementing for DFID. USAID interviewees talk about arbitrary targets set at the top, and one interviewee describes USAID as “an agency under siege” which “make[s] you cautious”. Throughout the interviews, it is clear that political authorizing environments and legitimacy-seeking loom much larger for USAID than they do for DFID.

*A brief comparison of Two Illustrative Projects, and what I mean by “the less successful project looks more successful to the donor themselves”*

Two projects – one from DFID, one from USAID - in South Africa sought to help municipal Governments work more effectively in expenditure management, control of corruption, revenue systems, revenue planning etc. Despite these projects being almost exactly the same size, in the same broad geographic area, over the same time period and having essentially interchangeable goals, they went about implementing in very different ways (see Figure A1).

**Figure A1. Comparison of USAID and DFID municipal government capacity-building projects in South Africa**

	USAID Local Governance Support Program Phase II (LGSP) 2004-2010	DFID Consolidation of Municipal Transformation Program (CMTP) 2003-2009
<i>Goal</i>	“Improve municipal planning skills and operating systems to increase effectiveness, transparency, and accountability.”	“Promote th development of effective and efficient municipalities.”
<i>Primary Implementation Modality</i>	Training visits	Resident Advisers
<i>Outputs (Official Language)</i>	“Train all staff on implementation of policy”	Advisers “advise/mentor [municipal] managers in achieving targets, plan and budget, unblock delivery obstacles and achieve institutional coherence.”
<i>How to Begin Delivering?</i>	“Hands-on Mentoring Dates” for training on e.g. “Credit control policies and	Advisers “conduct an assessment of status quo and prepare a report” (which included a workplan)

<i>(Official Language)</i>	implementation, and debt policy implementation.”	
<i>Reporting (Official Language)</i>	Success Indicator: “All staff trained in Finance Dept and Municipal Secretariat” (in practice, count people at training)	Advisers “implement their workplans and report on progress monthly and quarterly.”

USAID ran its project through training visits. Trainers traveled to project municipalities to provide day-long or two day-long sessions. DFID, on the other hand, hired advisors who were embedded in communities for a number of years, who worked with the local government structure day in and day out.

I filed freedom of information requests with both agencies. As a result, for USAID I can tell you quite a bit about the project’s details. In some cases I can see expense reports, attendance rosters, training materials, etc. But I can’t say whether anyone learned anything, or whether anyone did anything useful with the things that they learnt.

On the other hand for DFID, I can’t tell you much. DFID’s embedded advisors did inception reports and then were asked to report against those inception plans. So I have some of the quarterly reports; but these reports just show some advisor’s view on their success in accomplishing a plan they designed. As a form of control, it’s the rough equivalent to me as a professor asking the students to compose the test questions; then take the test; then assess how good their answers were on the test questions they wrote. Bottom line: this is not a very tight form of control.

Which is the point, from an illustrative standpoint. That is, I’m not at all worried that the USAID project has too little control. And I’m not at all worried that the DFID project has too much. So in comparing them, we can see how relative extremes fared.

So how did these projects do? The USAID project, according to the Deputy Chief of Party—the deputy head of the project—“might not have made the most dent or impact”. One of the trainers talks about the training not contributing much; “you go there, you don’t have any authority, if people don’t co-operate you can’t say anything”.

The DFID project, on the other hand, was sometimes – but by no means always - effective. Relying on people’s judgment means that people make different judgments. It’s also possible in some circumstances it would’ve been difficult for anyone to make impact. But when I talk to those in the communities in which these projects happened, over and over, I had to remind folks of the USAID’s project existence. On the DFID project that never happened because everyone remembers their advisor and the time they spent in the community and can talk about what they did or in some cases didn’t do effectively in their time there. DFID’s assessment noted that successes were linked to good judgment calls and that “although uneven in parts, [the project] included some highly positive examples in selected municipalities”; and I agree. So comparatively, the DFID project was clearly more successful.

But it is, in some sense, a funny kind of “more successful.” That is, it’s not that the DFID project has no failures; it’s that the USAID project has very little to show by way of success. *But the USAID project was set up in such a way that it never found this out.* That is, the USAID project counted outputs, and reported those outputs; as seen from DC, things happened in the field. Almost more interesting, more informative, than the USAID project’s lack of success is that it set up in such a way as to not find out that it didn’t succeed. So this might mean that *as seen by their respective political authorizers*, it might well be that it is the more successful DFID project that needed to explain ‘failure’, not the less successful USAID project.

### *How This Relates to Cash-on-Delivery Aid*

This being a Center for Global Development (CGD) policy paper, let me also say a few words about how this relates to Cash-on-Delivery (COD) aid.

CGD has long championed Cash-on-Delivery aid, the notion of “paying in proportion to progress that is independently verified” on an important development outcome, e.g. completion of primary school.<sup>92</sup> And I think this argument is perfectly consistent with COD being a brilliant idea... under certain conditions – basically, consistent with being far towards the “left” of figure 1, with pretty verifiable tasks and pretty certain environments. This is totally consistent with Birdsall & Savedoff (2010), who stress the need for “a clear enforceable contract”, “independent verification”, and of course outcomes that aren’t going to change/be distortionary. I agree completely with a CGD publication which argues that COD “is only feasible where program goals can be captured by one or more contractible indicators”.<sup>93</sup> The paper does on to note that those indicators must be “measurable and verifiable *in that context*”; under the recipient’s control; and “difficult or impossible to manipulate or game”.<sup>94</sup>

I strongly support a move towards more COD when all these things are true. Figure A2 adds a line to figure 1 for COD aid, pointing out that where “task verifiability in context” is really high - by which I mean all the conditions above are met – COD aid is better than any other strategy. That said, the range where Navigation by Judgment is the superior strategy to Navigation from the Top is for a range of environments, and tasks, where these conditions are very unlikely to be met in practice.

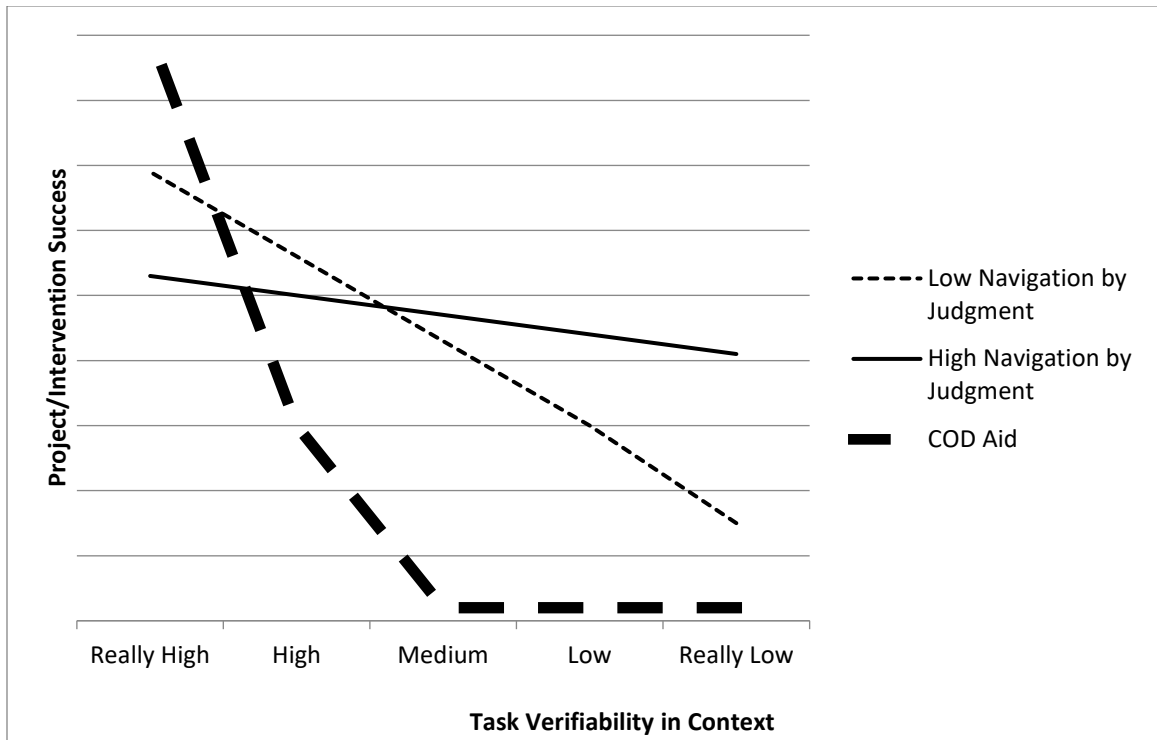
---

<sup>92</sup> Quote and example from Birdsall, Maghoub, & Savedoff 2010, which is drawn from Birdsall & Savedoff 2010.

<sup>93</sup> Silverman, Over, & Bauhoff 2015, 13.

<sup>94</sup> Ibid, pgs. 14 & 17.

Figure A2. Comparing COD and Navigation by Judgment



A world with *both* more Navigation by Judgment *and* more COD aid (where each is appropriate) is, in my view, an excellent idea. But where conditions aren't amenable to COD the answer is not to get as close to outcome-based aid as possible, but with less-than-perfect triggers. It's to *reduce* monitoring, increase discretion and autonomy - to Navigate by Judgment much more than at present.

### Where to Go for More Questions

This isn't close to a full answer to what the data says, or the argument is; I also haven't given you any detail at all on methods and empirical models etc. Please do refer to my book and other writings from this work stream for all the detail, or if you have a specific query just drop me an email ([dhonig@jhu.edu](mailto:dhonig@jhu.edu)) and I'll do my best to answer, or point you in the right direction.