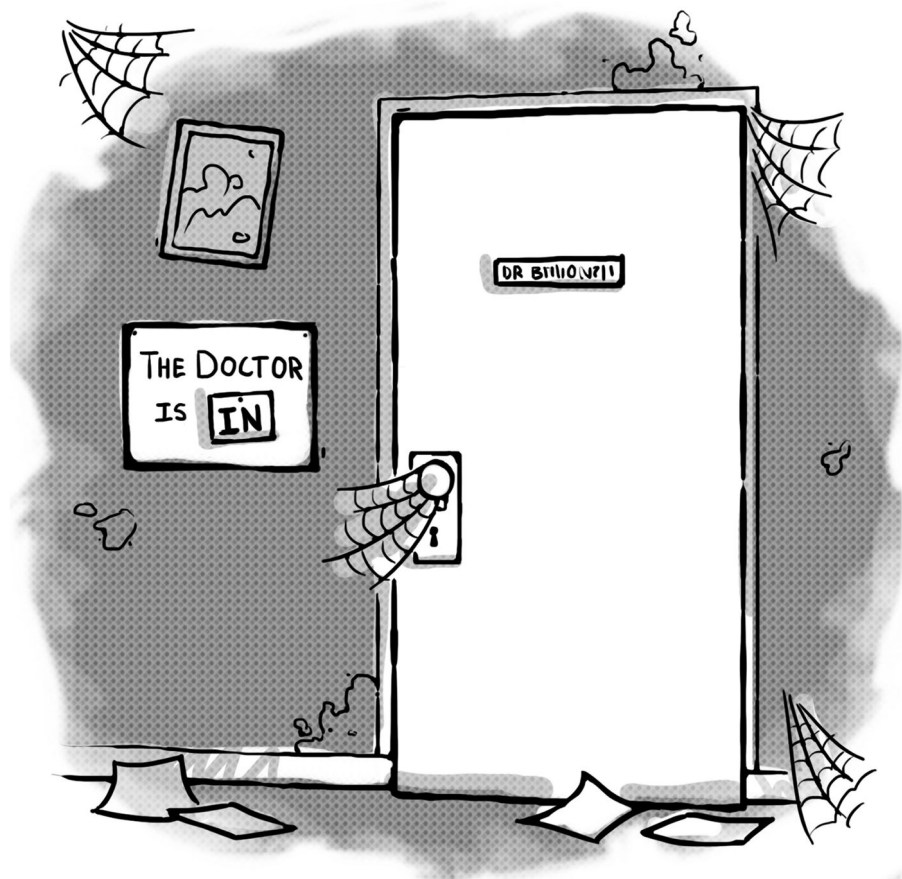


# 4

## Capability for policy implementation



When one of us (Lant) was working in Indonesia for the World Bank he was tasked with verifying the accuracy of government reports detailing which households were receiving subsidized rice, a program provided to mitigate the impact of a major economic crisis. As he traveled from the state capital to the village he was reassured by officials at each level (state, then district, then local) that all households that reported getting rice were actually getting the total amount, and that only those households were receiving the subsidy. Once in a village, however, it took all of about fifteen minutes to ascertain that, once the village head had received the allotment of rice from the logistics agency, the rice was being spread among many more people than just those on the eligibility list. Lant already had good reason to suspect that this spreading of benefits was happening, as it had been widely reported for months. This was both perfectly understandable (given the village dynamics) and perhaps even desirable in some ways—which is why he was traveling to the village to see for himself. The real insight, though, came when he turned to the officials who were accompanying him on the trip and said: “Why did you keep telling me all was exactly according to the reports?” After some furtive glancing back and forth, one of them said: “Well, you were from the World Bank. None of you has ever wanted to know the truth before.”

## Defining Organization Capability for Policy Implementation

Chapter 1 documented the low and stagnant levels of state capability using primarily country-level indicators like “rule of law” or “bureaucratic quality” or “government effectiveness.” In this chapter we zoom down to specific organizations and ask: what does it mean for an organization to have capability for policy implementation? For that we need to articulate what we mean by “policy” and by “organizational capability.”

A study of getting a driver’s license in Delhi, India, in 2004 (Bertrand et al. 2007) helps illustrate the key concepts. Researchers solicited participation from people arriving to get a driver’s license and documented how the “control” group in their experiment got their license (or not). The official or formal or *de jure* policy for getting a driver’s license in New Delhi looks pretty much like anywhere else: one goes to a government office, proves various personal facts about eligibility (like identity, age, and residence), shows the physical capacities associated with driving (like adequate vision), and then demonstrates driving ability through a practical test. In principle, those that meet the requirements get a license and those that don’t, don’t.

You might guess that that is not at all what happened. Fully 70 percent of the control group who successfully obtained a license hired a tout (also known as a fixer or facilitator) to help them with the transaction. The touts did more

than just facilitate the interaction with the bureaucracy. Only 12 percent of those in the control group who hired a tout actually took the *legally required* driving examination. In contrast, 94 percent of those that did not hire a tout had to follow the law and take the practical road test, and two-thirds of those who took the test without hiring a tout failed the driving test and did not get a driver's license (at least in their first attempt; most of them wised up and just hired a tout in the next round). The intervention of the tout did not just speed the process along, it actually subverted the purpose. The study tested the driving ability of those that hired a tout and got a license—and two-thirds of them could not drive either and, if the policy were actually implemented, should not have had a license. All else equal, those that hired a tout and had a license were 38 percentage points more likely to fail an independent driving exam than those who got a license without hiring a tout.

Knowing the results of this study, what is the “policy” for getting a license in Delhi? One could recite the formal rules or policy formula but equally persuasively one could say the actual policy is “hire a tout, get a license.” In our working definition a “policy” has four elements: a *formula* that maps from actions to facts, *processes* for determining the policy-relevant facts, a set of *objectives*, and a *causal model*.

A policy *formula* is a mapping from facts to actions by agents of an organization. This *formula* from facts or conditions or “states of the world” to actions by agents is often what is described as a policy. A fire insurance policy says “if the fact is that your house burned down, here are the actions we the company, via its agents, will take” (though it may say this in a few hundred pages). We call this a policy formula because in mathematics class we all learned that a function maps from a domain to a range; a policy formula is a mapping where the domain is “facts” and the range is “actions by an agent of an organization.”

Discussions often conflate the policy *outcome* and the *formula*. For instance, a tariff policy is a mapping from different types and value of imports (the policy relevant facts) to authorized actions of agents in collecting revenue. But the total tariff revenue collected is not the policy formula; it is the outcome of an application of the policy formula to a set of facts. The exact same policy formula can produce very different outcomes: two countries could have exactly the same tariff code and yet different tariff revenue if the composition of their imports varied.

As a policy formula is a mapping from facts to actions, a policy must have a specification of how the *administratively relevant* facts of the formula are to be determined and, if necessary, adjudicated. The driver's license formula says “if the *administrative fact* is that you can drive (and other conditions are met), the action is that you are issued a license.” An integral part of the policy is the specification of which organization and which agents have the authority to declare what the *administrative* facts are. These need not have anything to do with common-or-

garden variety facts. A property tax policy formula applies a tax rate to the *administratively relevant taxable* value of a property. The value of the property as determined by its market price or its value as collateral can be completely irrelevant to the policy implementation relevant fact of its value for tax purposes (either de jure or de facto) and there is a process whereby that value is determined.

The combination of a policy *formula* as a mapping from administratively relevant facts to actions by agents and the *process* of determining the facts implies that policies and organizations are inextricably linked. Integral to a public policy is a designated organizational mechanism for implementation. Conversely, most public sector organizations are defined by the policies they are authorized to implement.

The emphasis on the organization authorized for policy implementation as an integral part of a policy helps distinguish organizations and institutions. Institutions are commonly defined as “norms or rules or human devices for affecting the behavior of individuals so as to structure the interactions of groups of people.” This definition of institutions would include an incredibly broad array of human practices, from those associated with the “institution” of marriage to an “institution” of private property to the “institution” of religion. Some formal institutions may be enforced with official policies and legally constituted organizations responsible for implementation. But institutions can be also be “informal,” with no written policy formula and no organization responsible for implementation. The distinction between organizations and institutions and between formal and informal “rules or norms” is crucial because formal organizations often lack capability for implementation because there are informal norms that have more traction on the behavior of implementing agents than formal rules and processes. As we will show in several cases, this leads to policy dysfunction.

While one could regard the specification of the policy formula (mapping from facts to actions) and the organizational processes of implementation (how facts are determined and adjudicated and actions taken) as complete, in our approach a policy has two more elements: *objectives* and a *causal model*.

We define *objectives* as an intrinsic component of a policy. Many policies may lack an explicit declaration of objectives or purposes but, whether these are implicit or explicit, a policy exists to do something—educate a child, limit environmental damage, prevent corruption, resolve disputes to avoid violence. We focus on a policy’s *normative* objectives, which may differ from the actual purposes to which the organization or policy is being used. That is, the *normative* objective of a tax is to collect revenue for the government. The actual organization purposes may be to use the authorization to tax to collect some tax and also extract some revenue that flows to others illicitly, but one can still consider a policy and implementing organization relative to its normative objectives.

A policy also has a *causal model*. A causal model is what relates the policy formula (mapping from facts to actions of agents) to the policy objectives (what the actions of the organization implementing the policy are meant to achieve). While the causal model is almost never made explicit by organizations, it is nevertheless a critically important part of the policy as it ultimately serves as part of the organization's claim to legitimacy, both externally to its "authorizing environment" (Moore 1995) and internally to its own agents.

The delineation of a policy into the four elements of policy formula, organizational process for determining facts, normative objectives, and causal model (as illustrated in Table 4.1) highlights two distinct ways in which a policy could fail to achieve its normative objectives. The *policy formula* could be based on a *causal model* that is wrong about the connections between the fact-contingent actions of agents and the normative objectives. In this case, even if the policy formula was faithfully implemented—the policy relevant facts correctly assessed and policy formula stipulated facts taken—this would not achieve (or perhaps even promote) the policy objectives.

The other possibility is that—as in the case of the post office and international mail in Chapter 1, the driver's licenses in Delhi, the examples about healthcare below, or any of hundreds of examples around the world—policy is just not implemented. The driving test may or may not reduce traffic accidents—but people are getting licenses without it so it doesn't matter. Having nurse-midwives in clinics providing antenatal care may or may not decrease child mortality at birth—but if they are not there the question is moot.

Much—almost certainly most, and quite possibly nearly all—analysis of public policy focuses on which policy formula are based on correct causal models such that *if they were implemented* they would produce better outcomes. Economic analysis, for instance, often has very different predictions about the impacts of the expansion of government supply of commodities because it assumes consumers are already acting to secure what they demand. This can mean that policies based on a naïve model of impact of supply will fail to predict actual outcomes even if the policy were perfectly implemented. We are all for more and better scientific analysis of policy impacts. But "black box" approaches to policy/program/project impact cannot distinguish between a failure to produce desired outcomes because of an incorrect causal model linking outputs to outcomes, and a failure of policy implementation—namely, that implementing agents just didn't do what the policy formulation stipulated they would do. Either will lead to the inputs provided not producing the desired outcome; not making these necessary distinctions can thus lead to inaccurate conclusions as to why a given "policy" (or program or project) failed.

Building State Capability

**Table 4.1.** The elements of a policy: formula, administrative facts, normative objectives, and a causal model

	Policy formula		Organizational process for determining administrative facts	Normative objectives	Causal model
	Facts	Authorized actions			
Imposition of obligations					
An 8% sales tax	The firms taxable sales	Collect 8% of total	Tax authority through records and/or audits	Collect revenue to fund government	
Driver’s license	Age, eligibility, adequate sight, driving ability	Issue legal authorization to operate motor vehicle	A agency/ bureau that approves, including testing driving skills	Reduce road accidents, injuries, fatalities	Allowing only people capable of driving (ex ante and ex post) to legally drive will reduce the risks of traffic accidents
[Reader’s example]					
Delivery of services					
Immunizations	Child of appropriate age, vaccination history	Give child vaccination	Variety—use of healthcare providers or facilities or vertical programs	Reduce child illness/ death from preventable causes	Vaccinated children will be at less risk themselves and less risk of transmitting diseases to others
[Reader’s example]					
Operation of the state					
Procurement	Is the bid the least cost qualified bid?	Sign contract with bidder	Procurement unit of organization	Get most benefit to citizens from use of resources, prevent corruption	The procurement process is capable of generating competition among alternative suppliers that reveals lowest costs

Organizational Capability for Policy Implementation

Sorry for the last section. It was kind of like a predatory big cat sneaking up on prey with slow stealthy moves, no one of which seemed particularly threatening, in fact, kind of boring. While it may have seemed tedious, the definition of “policy” as not just formula but also organizational process, objective(s), and a causal model enables us to define strong and weak organizational capability for policy implementation. Strong capability organizations are those in which agents take those actions that promote the organization’s normative objectives.

Organizations with weak capability for policy implementation are those that cannot equip their agents with the capacity, resources, and motivation to take actions that promote the organization's stated objectives.

There are two elements embedded in this definition of capability for policy implementation that we need for our overall approach but which we wish to highlight are unusual: the seemingly sudden pounce after the boring stalking.

It might be a big surprise that our definition of organizational capability for policy implementation does not refer at all to the policy formula. One very popular older approach was to *define* the objective of public sector management or public administration as policy compliance. In that frame, an organization with strong capability for policy implementation would be one that implements the policy formula: it ascertains the facts and applies the policy formula to those facts with fidelity. In our view this approach is very attractive—indeed it may seem like common sense, if not definitional, that high organizational capability for policy implementation should be measured by policy implementation. But we feel this approach is deeply wrong about what capability is, indeed that it leads to misguided and counterproductive approaches to achieving capability.

Embedded in our definition of organization capability is that organizations discover and act on a workably correct causal model of achieving the policy's normative objectives. Take an extreme example. Suppose there was a society that believed that the sun would only come up in the morning if during the night a crank was turned. Given the importance of the sun coming up, this society may create an elite organization responsible for the nightly crank turning. This organization may achieve perfection in complying with the policy formula and the crank is turned every night. Does this crank turning organization have high capability? Certainly it has high capability for achieving compliance with the policy formula. But since no reader of this book sincerely shares the causal model that the sun's rising is determined by crank turning the organization has no capability at all for achieving the normative objective of raising the sun.<sup>1</sup> Enabled by the prevalence of isomorphic mimicry (Chapter 2) and overambitious agendas (Chapter 3), the developing world is full of excellent policies (indeed often "best practice" policies) and lousy outcomes. At least partly responsible for this state of affairs, however, is that definitions of organizational capability and its construction have been separated from achieving objectives and instead reduced to compliance.

<sup>1</sup> And, as modern academics, we can go "postmodern" and define other social objectives that the crank turning may achieve as defined from within a social context in which all participants do in fact believe a false causal model about the sunrise and that the organization also has capability for those, potentially important, objectives. But it doesn't have capability to raise the sun.

We are defining capability relative to *normative* objectives. This is not a reprisal of the “functionalist” approach, in which an organization’s capability would be defined relative to the function it actually served in the overall system. This definition allows an organization engaged entirely in isomorphism to assume the status of being “capable” if it was fulfilling a functional role through that isomorphism.

Using our definition of state capability we delineate five levels.

*Ideal capability*, in which the agent takes the best possible action available and hence produces the best achievable policy outcome. We assume agents are maximizing the normative objective of the organization. This can produce outcomes better (perhaps much better) than policy compliance. This assumes the agent has a perfect ability to determine the relevant “facts of the world” and has perfect causal knowledge of what action will produce the best outputs and outcomes—which, in our imperfect world, is completely impossible. Even so, ideal capability is the standard to which the best organizations aspire and in reality closely approximate.

*Policy-compliant capability* means that agents do exactly and only what the policy formula dictates. Agents give drivers’ licenses when, and only when, the fact of the world meets the policy formula conditions for a driver’s license. The case of the Delhi drivers’ licenses, of course, was less than policy compliant. But even policy compliance can be much less than ideal, if either (a) the policy formula is less than ideal (or just plain wrong), or (b) success requires actions that cannot be fully specified in a written policy (see Chapter 5 for a typology). In education it is hard to believe that a policy could dictate exactly what teachers should do such that a “policy-compliant” outcome would actually be an ideal educational experience.

*Actual capability* is what happens in practice when agents make their own decisions. In Delhi, agents colluded with touts and gave licenses to drivers who had not passed the formally required driving test. In this case, actual capability was less optimal than policy compliance. This is the typical case of “actual capability” in the developing world: agents choose to maximize their own wellbeing, with the objective function that is inclusive of intrinsic and extrinsic motivations and with the incentives presented by their social and organizational context.

But in cases of high-capability organizations, actual capability is preferable to pure policy compliance because agents can take actions to improve outcomes. (In such organizations, “work to rule” is a threat because doing so *lowers* effectiveness, whereas in low-capability environments “work to rule” would be a massive improvement.) Thus, actual capability could be more than policy-compliant capability and nearer ideal, or could be (and often is) much less policy compliant and actually near zero.

Weak capability for implementation manifests in organizational inputs, outputs, and outcomes. The agents of organizations do not do what they are supposed to do—they are absent, they do not put in effort, they take bribes, they are ineffective or even counterproductive in their actions. Weak



capability results in low organizational outputs from policy implementation—regulations are not enforced, infrastructure is not maintained, mail is not delivered. The result is teachers who do not teach, police who do not police, tax collectors who do not collect taxes.

*Zero capability* is what would happen if there were no organization at all. Actual capability can be this low—or, as we will see, lower.

*Negative capability* is a possibility because the state, by the very definition of being the state, has the ability to coerce. Organizations of the state can use power to exploit their own citizens and, through the imposition of obligations with no corresponding benefits, make them absolutely worse off.<sup>2</sup>

### *Capability: More Than Individuals, Less Than Countries*

A key task as we move toward a pragmatic approach to building state capability is to shed two common misconceptions that implicitly or explicitly guide efforts to build capability. One misconception is to not distinguish between the *capacity* of individuals and the *capability* of organizations. Perhaps the most common response to low capability, particularly when external agents get involved, is to propose more technical training (“capacity building”) on the view that organizational capability is limited by individual capacities. The second misconception is that state capability is completely determined by broad nation-state (or perhaps state or provincial) level conditions and hence what is needed to build capability is broad “reform” that affects all state organizations.

*Organizational capability versus individual capacity.* Given the overwhelming importance given to “training” in discussions of building capability, one might imagine that the *capacity* of individuals in an organization was (nearly) everywhere and always a key constraint to the *capability* of organizations. We define the *technical capacity* of individuals as their ability to recognize and act on a correct causal model.<sup>3</sup> But in many instances it is obvious that the capacity of individuals is not the key constraint—they know what to do, they just don’t do it.

One simple illustration of this is absenteeism. In this case the relevant policy formula maps from the fact of the world that is date and time to the action of the agent of being there. A study of teachers in India (again, a middle

<sup>2</sup> Leeson (2007), for instance, argues that the typical Somali may well have been better off without any state than with the predatory state they had. Scott (2009) discusses how various communities in Southeast Asia have actively avoided “being governed,” as anarchy was deemed preferable to the predatory states that were available.

<sup>3</sup> We could just as easily refer to organization “capacity” and individual “capability”—the only point is that it is helpful to distinguish the two, in particular the one is not the simple sum of the other. Using a different word for each helps.

capability country overall) found that teachers were not present in the school 26 percent of the time. A follow-up study, a decade later after much attention to this issue, found it had declined, but only to 23 percent. A recent study of eight African countries found average absence was 20 percent. These rates of absence result in weak organizational capability—achieving the normative goal of student learning is clearly inhibited by teacher absence—but it is ridiculous to imagine that any of this absence is because the teachers either don't understand the policy formula ("be there on Tuesday") or do not know the true facts ("it's Tuesday").

An excellent illustration of the distinction between technical capacity and organizational capability on a more complex implementation issue comes from two different studies of healthcare providers in India. One study in Delhi assessed the technical capacity of medical care providers by analyzing their ability to respond correctly about how to diagnose and act on conditions presented in vignettes (Das and Hammer 2007). They then also observed the same providers in their actual practice. The public sector employed only trained doctors as providers, so their *technical capacity* on the vignettes was much higher than the typical private sector provider (many private sector providers of first line medical care were "less than fully qualified"—some might say "quacks," i.e. people offering medical advice and services with very little or no training at all). But, when examined in their public sector primary health center (PHC) settings the trained doctors did only a small fraction of what they had demonstrated they knew how to do while the private sector knew little but did what they knew.

A follow-up study in rural Madhya Pradesh assessed healthcare providers by training research collaborators to present as patients and report specific symptoms. Some presented with symptoms of myocardial infarction (heart attack), complaining of chest pains. Of the public providers, very few asked even the most basic diagnostic questions: only 45 percent asked about the location of the pain, only 19 percent about its severity, and only 10 percent whether the pain was radiating. (We as middle-aged men are a biased sample, but even we know that location, severity, and radiating pain are key symptoms for recognizing a heart attack.) The "policy formula" when faced with a "fact of the world" of a patient presenting with symptoms of myocardial infarction in rural settings is very simple: (1) aspirin, (2) nitroglycerine, (3) ECG, and (4) referral to a hospital. Fifty-eight percent of formally trained (MBBS) doctors in public primary clinics in Madhya Pradesh did *none* of those things for patients presenting with symptoms of a heart attack—not an aspirin, not an ECG, not a referral.

This outcome was not the result of a lack of technical capacity on the part of those individuals to diagnose; it was the *organizational* setting that determined these outcomes. The distinction between technical capacity and

organizational capability was made clear because the study also had the “patients” present themselves with symptoms at the practices of the exact same MBBS-trained doctors who worked in the private health clinics when they were working on the side in their own practices. The result was that, when judged by either the likelihood of checklist completion of the protocol for the disease or by a standardized checklist score, the *worst* medical care in the study was provided by the trained doctors in their public clinics and the *best* medical care was those exact same doctors in their private practices. The differences are astounding, as checklist completion is less than 3 percent in their public practice—which is therefore the *organizational capability*—while in their private practice it is 27 percent—which reveals the *technical capacity* of the doctors as individuals far outstrips their actions when embedded in the organization.

By combining the efforts of individuals in productive ways, organizations can have capabilities much, much higher than individuals alone. Indeed, it might be said one of the very foundations of “modern” economic and political life is the rise of organizations (public and private) that have vastly higher productivity than that of individuals. In such organizations, the whole truly is greater than the sum of its parts. Yet it is also the case that organizations can be so dysfunctional that they become “value subtracting”—i.e. the productivity of the individual when inside such an organization is *lower* than it is outside and the whole is much *less* than the sum of the parts.

Illustrations of this unhappy phenomenon of value subtraction come from studies of contract versus civil service teachers. In an experiment in Kenya, a new teacher was added to early grades to reduce class size. When the teacher added was hired as a civil service teacher the additional teacher had no impact on improving child test scores. When nothing else was different but that the new teacher was hired on a contract renewable at will by the school (and hence with performance and parental input), student test scores improved substantially (Duflo et al. 2007). An observational study in Uttar Pradesh found students learned twice as much from a contract teacher versus a civil service teacher—in spite of the fact that the salary of the civil service teacher was many times higher than that of the contract teacher (Atherton and Kingdon 2010). It appears that a person with exactly the same technical capacity has their absolute level of productivity *reduced* by being inside the standard civil service-type organization—the organization is value-subtracting.<sup>4</sup>

<sup>4</sup> One should not conclude from such a study, of course, that the solution to raising student test scores around the world is to hire lower-cost contract teachers; our point here (and that of the researchers) is that it is the capability of organizations that powerfully influences the performance of its constituent members, and that in the worst situations dysfunctional organizations can literally subtract from, rather than add to, the technical capacity of their members.

*Capability varies across organizations in the same country.* While in Chapter 1 we use national measures of state capability, these are broad aggregates and of course hide massive variations in capability across the same country. For instance, many of our examples to illustrate low capability in this book come from studies in various states of India and in Chapter 1 we saw that India was an average developing country in aggregate capability. Yet many of the top-tier organizations in India exhibit very high capability. In September 2014 India successfully put a satellite into orbit around Mars. The graduates of India's institutes of technology are highly recruited globally. So the fundamental issue is not that India is a “failing” state with no state capability, rather it is a “flailing” state with highly capable elite organizations and yet very poorly performing organizations in other aspects (e.g. policing, basic education, health).<sup>5</sup>

A study in Bolivia surveyed over 1,000 public officials and asked them to rate the performance and characteristics of other Bolivian public agencies.<sup>6</sup> The results revealed large and consistent patterns in differences across organizations, even within a country with very low country-level measures of capability. For instance, the Ombudsman, Electoral Court, and National Comptroller had service performance ratings by agents of other public agencies over 80 whereas the police in Santa Cruz had a service rating of below 30. On an index of bribery, organizations like the Ombudsman are rated near zero while the worst (again the police in Santa Cruz) are rated over 80.<sup>7</sup>

Any adequate account of organizational capability has to be able to explain both differences across countries in aggregated or average organizational capability as well as differences within countries across organizations. A World Bank study on health and education in the Middle East North Africa region shows this same logic can be deployed to explore and explain how public agencies within the same sector within the same country enacting the same policies can nonetheless generate considerable performance variation—in Yemen, for example, staff absenteeism in health clinics ranges from 8 to 83 percent. Learning from this variation in performance can reveal how the capability can be strengthened.<sup>8</sup>

### How Organizational Capability for Policy Implementation Matters

The effectiveness of policy is mediated by the quality of implementation. The example of the (non)returning of misaddressed international mail in Chapter 1 shows that the exact same policy can lead to completely different outcomes—from zero to 100 percent. All countries have policies against

<sup>5</sup> See Pritchett (2009). <sup>6</sup> Kaufmann et al. (2002).

<sup>7</sup> Since the scores are all normed by the mean and standard deviation within Bolivia, it is impossible to compare how large these variations are compared to international differences.

<sup>8</sup> See Brixi et al. (2015).

corruption in public procurement and yet corruption is nearly absent in some cases and ubiquitous in others. When implementation is weak the converse can be true: we show below a case in which completely different *de jure* policies lead to roughly similar outcomes.

Weak organizational capability for policy implementation leads to two practical consequences: administrative fact becomes fiction; and the consequences of *de jure* “policy reform,” particularly a change in the policy formula, are completely unpredictable. Note that the key point here is not whether these matter for final outcomes; rather, this is *how* weak capability affects implementation. We address each in turn.

### 1. *Facts Can Be Fiction*

A policy formula is a mapping from facts to actions. This makes implementation sound easy. But the sad fact is facts are (often) not facts. Public sector organizations do not operate on regular garden variety facts like that the sky is blue, rain is wet, and Tuesday is a workday but on the *administrative* facts. Policy includes the designation of which agents have the authority to declare administrative facts and who and how disputes about administrative facts are adjudicated. One of the ways in which implementation fails is not that the real facts are agreed as administrative facts and implementing agents fail to act on those facts, but rather that the administrative facts stipulated in the policy formula are manipulated to create policy compliance which is a complete fiction.

Living in parallel worlds of administrative facts and actual facts is part of life in most developing countries.<sup>9</sup> A friend of ours was interviewing a girl in rural India about her schooling. Since government schools provide benefits like free uniforms and mid-day meals there are incentives to be enrolled in public school. But learning conditions are perceived to be better in low-cost private schools. The girl regularly attended a private school. But when asked where she went to school she said she went to the government school. Our friend pointed out that she was actually sitting in class in a private school during school time. The girl thought about this for a few minutes then responded: “My *name* goes to the

<sup>9</sup> Nearly any ethnographic work (or just work that actually asks people what is going on) finds that “working misunderstandings” (Watkins and Swidler 2013) between the official policy formula, front-line agents, brokers (who often have no formal existence), and “beneficiaries” are rife (as Watkins and Swidler 2013 did in their work on NGOs and AIDS work in Malawi). Robert Wade’s analysis of irrigation in India revealed that what was supposedly a bureaucratic agency was really a collection of markets (Wade 1982). Diane Singerman (1995) in Egypt discusses the key roles of networks in securing public services. As pointed out in different contexts by Scott (1998), Ferguson (1994), and Mosse (2005)—among many others and the simple vignette that opens the chapter—it actually takes a special kind of training and mindset to “stabilize the epistemic framing” (Mosse 2005) and *not* see the gap.

government school, but *I go to private school.*” In low-capability organizations, even seemingly routine administrative facts like someone’s age are often opaque and potentially open to abuse. Gupta (2012) documents in anthropological detail how certain front-line implementers of social programs in rural India exploit widespread uncertainty about people’s actual age to their personal advantage, demanding sexual and financial favors from citizens in return for declaring them age-eligible (i.e. over 50 years old) for these programs. In such circumstances, one is as old as the implementer deems you to be; the administrative fact is effectively arbitrary on an issue that in high-capability organizations and contexts is precise and readily verifiable in seconds.

Let us give four other quick examples of where administrative fact is fiction.

*Regulation of pollution in Gujarat.* The regulation of industrial emissions of pollutants in Gujarat, India, required private firms to hire other private firms as “auditors” to assess their level of emissions. But these environmental emissions auditing firms were chosen, hired, and paid by the emitting firms. A recent experiment looked at what happened to reported emissions before and after the incentives changed such that auditors were not dependent on the goodwill of emitting firms for business.<sup>10</sup>

Not surprisingly, when firms hired the auditors to declare the administrative facts about their admissions, the facts were a complete fiction. The reported facts were that nearly all firms had emissions just below the legal threshold. Again not surprisingly, the actual facts were that many firms had emissions two or three times higher than those reported by the auditor. Perhaps surprisingly, however, many firms had emissions much *lower* than those that were reported by the environmental auditor. One might think that reporting pollution *higher* than your true level makes no sense. But, once it was widely acknowledged that the administrative reports were a complete fiction the only objective was to be cheap (why even visit the plant?) and not attract regulatory attention (so report a value clustered where it seems not in violation but also doesn’t seem suspiciously low).

*Community development in Kenya.* The World Bank has financed “community driven development” (CDD) projects in Kenya.<sup>11</sup> One element of these projects is to create “livelihoods” by providing poor beneficiaries with assets, often livestock like goats or cattle or chickens. Given that this was a World Bank-financed project there were both activities like local meetings intended to provide accountability to the beneficiaries and hence reports on those meetings, as well as the standard reports on procurements. The World Bank had, on the basis of the reported administrative facts, rated the project performance as “Satisfactory” from 2003 right up until a forensic audit in 2010 revealed the facts were fiction and forced the project to be suspended.<sup>12</sup> A forensic audit of seven districts found that in the CDD component of the project, 84 percent of all expenditures were “suspected fraudulent or questionable.” The records of “community

<sup>10</sup> Duflo et al. (2012).

<sup>11</sup> Ensminger (2013).

<sup>12</sup> See World Bank (2012).

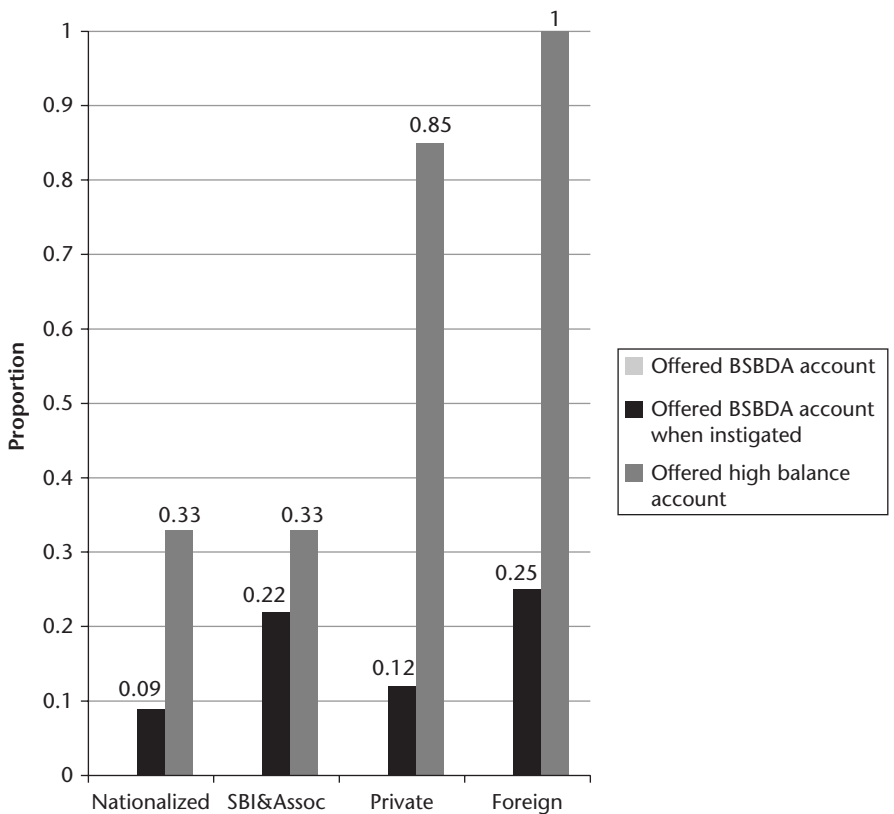
participation” were fabricated and the names of villagers attending were just produced without their actual participation or consent. Ethnographic research found that the project implementers at the district level were able to combine with agents at the village level to almost completely capture the benefits of the project in fact, while producing documents and records creating the fiction that all was well.

*Doing Business.* The gap between the de jure administrative fact and the on-the-ground reality is evident in many studies of particular issues in particular countries and also in cross-national comparisons. As part of the Doing Business indicators the World Bank ask experts to estimate how long it would take a firm or person to obtain a construction permit to build a new building of a specified type *if they followed the law*. In many of the same countries the World Bank also does an Enterprise Survey of a sample of firms. This survey asks of those firms that have recently constructed a building how long it took the firm to get the license to do so. While not a perfect comparison—firms construct many kinds of buildings in many different cities or regions within a country—this provides at least a rough-and-ready comparison between de jure and de facto policy implementation.

The comparison of the two measures of regulatory compliance is revealing. While one might think that it would take longer in countries with stiffer formal regulation, in reality there is almost zero correlation across countries between the Doing Business time and the average or median of what firms actually report. If you wanted to predict how long it would take a firm to get a construction permit, knowing the country estimate of the legal time to obtain a license would have no little or predictive power. Figure 3.4 (see Chapter 3) showed that in countries where the Doing Business measure was fewer than 200 days the average firm reported it took 58 days to get a permit. In countries where Doing Business reported, it took more than 300 days (and in the median country in this category it was 381 days—more than a year), the average firm that got a permit reported it took them *less* time—only 47 days. So an *increase* in 230 days in the de jure time to get a construction permit is associated with a nine day *decrease* in the time the average firm reported it actually took (Hallward-Driemeier and Pritchett 2015).

This isn’t to say the law did not matter at all. In the Enterprise Survey data there are responses from many firms so one can compare the times reported by “fast” firms (the 25th percentile) and the “slowest” firms (the 95th percentile). There was a big gap between the “fast” and “slow” firms, such that the difference between “fast” and “slow” firms in the same country was much, much larger than the gap in legal times across countries. In countries where the Doing Business reported de jure compliance times to 200 to 300 days the “fast” firms reported 15 days and the slow firms reported 165 days—a gap of 150 days. One might say that for how much time it will really take a firm to get a permit it matters less where you are than who you are, and that what really matters for firms it not the de jure regulation but the implementation gap between policy and action.

*Bank accounts in India.* A naïve “public–private partnership” approach that relies on the private sector for implementation to achieve public purposes does not solve the problem of state capability, as it just pushes the question off onto private organizations—which may have capability for some purposes but not for pursuing a policy’s normative objective. One illustration is that in order to promote financial inclusion the government, via the Reserve Bank of India, mandated that all banks (both parastatal and private) had to offer a low-cost, low-balance account. One might think that since the law mandated their availability these accounts would be available. A study in the Indian state of Tamil Nadu in 2014 sent “mystery shoppers” into various banks to see if the banks would in fact tell potential customers about the availability of these accounts or open them if asked. Zero were offered the low-balance account at any type of bank. Even when the “mystery shopper” asked specifically about the type of account, only between 10 and 25 percent would admit to offering the account they were legally required to offer. And, as is often the case, the private and foreign providers



**Figure 4.1.** Changing law, changing behavior? A law in India mandating banks offer a basic savings account didn’t lead them to offer it—even when asked directly—even in public sector banks

Source: Adapted from Mowl and Boudot (2014: table 2)



were better at doing what was in their interest—85 percent got offered a high-balance account versus only one-third in public sector banks—but less likely to do what is not in their interest (see Figure 4.1). So capability for implementation is not solved by pushing implementation into the private sector, which can maintain the same fact-fiction gap in the absence of capable regulation enforcement.

### *2. Weak Capability Makes It Impossible to Predict the Impact of Changing a Policy Formula on Policy Actions, Outputs, or Outcomes*

Since weak capability for policy implementation often implies both that the policy formula is not being followed (and will have little or no traction on the behavior of the organization's agents) and that the normative objectives of the organization are being undermined, this means that the impact of "policy reform"—particularly of the type that changes the policy formula—has completely unpredictable impacts. Sensible sound policies—and even policies that have been rigorously "proven" to work in other organizational settings—may produce zero, or even perverse, results.

Some examples illustrate this point. Three different studies of attempts to reduce front-line worker absences by introducing technology to track attendance and incentives produced three different results. Working in partnership with a local NGO, researchers looked at the impact of using date-stamped cameras to verify the attendance of teachers at the NGO's schools. They found that the improved technology to verify attendance increased teacher attendance and that increased teacher attendance improved child learning (Duflo et al. 2012). So one might conclude that better technology to monitor attendance improved attendance, and thus that better technology improves the quality of service delivery. But no.

One of the same researchers worked with the same NGO to attempt to improve attendance of auxiliary nurse midwives (ANMs) at local clinics. This program introduced new technology to monitor attendance of the ANMs, introduced the possibility that their pay would be docked if they were present less than half the time,<sup>13</sup> clarified responsibility for attendance on a "clinic day" that ANMs should not have other field duties, and utilized the NGO in spot checks to "ground truth" the reliability of the technological monitoring of attendance (to check incentives to damage the new machines, etc.). Moreover, with a realistic nod to the difficult politics of changing the behavior of existing staff, this new policy applied only to newly hired ANMs.

<sup>13</sup> Previous extensive fieldwork by the researchers had revealed that the average absence in sub-centers and aid posts was 45 per cent so this was only a moderately ambitious target (Banerjee et al. 2004).

The outcome of this wonderful policy reform that drew on the rigorous knowledge from the previous paper (“Monitoring Works”) is aptly summarized by the title of the new paper: “Putting a Band-Aid on a Corpse.”<sup>14</sup> What is interesting is *how* the program failed. Eighteen months into implementation the rate of ANM *administratively recorded* absence had *fallen* in the treatment versus control group. Unfortunately the actual physical *presence* rate of the ANMs in the treatment group also fell. The program actually, if anything, increased *actual* absence while decreasing *administrative* absence. How? The category “exempted from duty”—ANMs not in the clinic but not counted as absent—rose dramatically. This attempt at improving health through better attendance of health workers failed because organizational capability for policy implementation was so low that putting increasing pressure on the recorded absence merely increased the manipulation of the administrative facts.

Another example comes from the Indian state of Karnataka, where the government introduced biometric monitoring of attendance with the threat that healthcare workers would be docked their leave days if they were excessively absent. This experiment had even more curious implementation and results (Dhaliwal and Hanna 2014). For one, the “treatment” was never fully implemented as, while the biometric machines were installed and the data reported, it was never actually the case that this data was used to discipline any worker (nor did it appear it could be implemented given the internal political objections and legal challenges). But birth outcomes improved in the “treatment” areas where the PHC introduced biometric recording of attendance—but not at all for the reasons hoped. The introduction of biometrics did not change doctor attendance at all, but did raise the attendance of other workers (e.g. nurses, pharmacists) at the clinic. Even so, this outcome was actually associated with *worse* perceptions of clinic quality by users, which in turn led *fewer* people to use the biometric treatment PHCs and instead they switched into higher-quality facilities—bypassing the PHCs for larger hospitals. Hence the better birth outcomes was the result of *lower* utilization of the PHCs in favor of facilities with better birth outcomes.

So, three rigorous experiments, all in India, each introduced some form of improved technology for tracking attendance into a low capability for implementation environment. The result is that pretty much anything that could happen, did happen: in one—which was an NGO provider—attendance went up and outcomes got better; in another, attendance didn’t change (or, if anything, got worse) as the policy was completely undermined; and in yet another the policy wasn’t implemented, the impact on attendance was mixed, patient-perceived quality got worse, but outcomes got better—because they used the clinics *less*.

<sup>14</sup> See Banerjee et al. (2007).

*Contract teachers.* A policy formula that works when implemented by one agency may fail when implemented by another agency, even when the “policy reform” seems an exact replica. A randomized experiment in western Kenya showed that reducing class size by hiring contract teachers, whose contracts might not be renewed at the discretion of the local community and school, improved children’s learning.<sup>15</sup> In that same setting, reducing class size by the same amount by adding additional civil service teachers did not improve student learning. Not that surprisingly, contractual status affected teacher performance which improved child learning. This policy reform had been tested in the most rigorous way and proven cost-effective.

When Kenya went to take this policy to scale nationwide, other researchers measured the impact of the scaled program. Fortuitously for social science, neither a major NGO nor the ministry of education had the ability to take the program to scale nationwide, so in part of the country an NGO was responsible for implementation and in other parts of the country the ministry was. The new researchers<sup>16</sup> found that when the new contract teachers’ policy was implemented by an NGO it had exactly the impact the previous research had found. However, then the *exact same policy formula* was scaled by the ministry of education, reducing class size by hiring contract teachers had the same impact as reducing class size with civil service teachers—zero. That the policy was “proven” to work with one organization’s capability was not evidence the same policy would work when implemented by an organization with different capability.

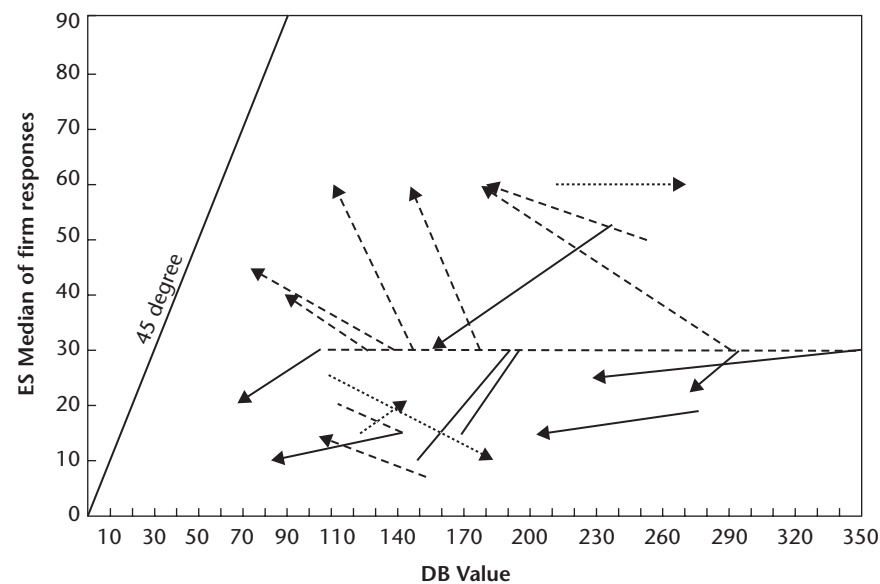
*Improving the “Doing Business” Indicators.* As we argued above, there was very little association between the rules on the books as recorded by the Doing Business indicators and the responses firms gave. Many countries have pursued reforms to aggressively reduce the times to compliance—as measured by the Doing Business indicator. What effect does that have on firms? Figure 4.2 shows that two-thirds of countries that reduced their de jure times to get a construction permit saw the time reported by firms either stay the same or increase. The impact of policy formula reform with low initial compliance is unpredictable (Hallward-Driemeier and Pritchett 2015).

\* \* \*

Organizational capability for policy implementation is not the achievement of policy compliance. Organizational capability is the ability of an organization to equip, enable, and induce their agents to do the right thing at the right time to achieve a normative policy objective. Reductionist approaches to organizational capability often attempt to reduce this to compliance with policy formula, which easily leads to isomorphism (see Chapter 2) or to

<sup>15</sup> Duflo et al. (2007).

<sup>16</sup> Bold et al. (2013).



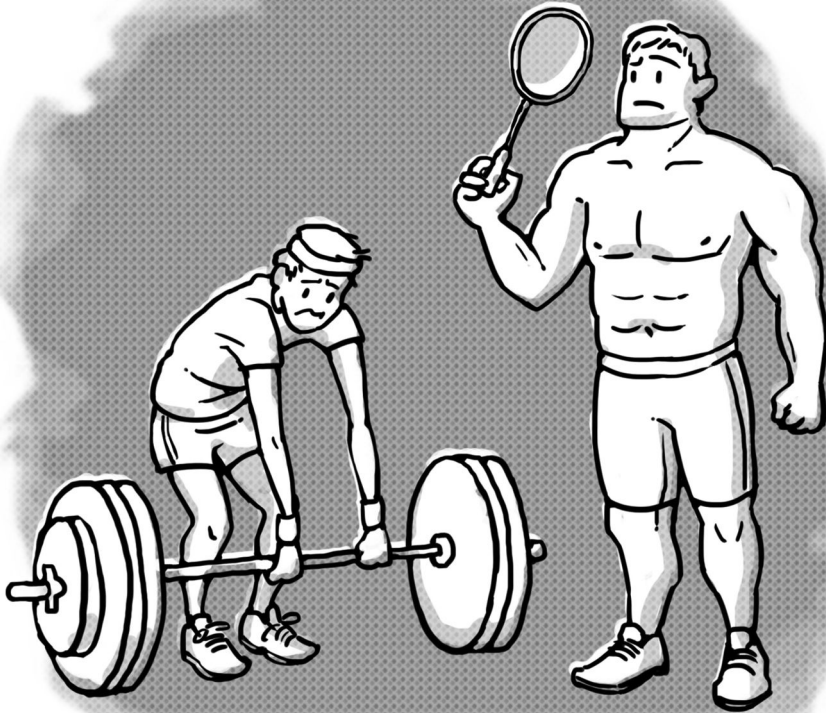
**Figure 4.2.** Evolution of days to get a construction permit: Doing Business and Enterprise Survey results  
*Source:* Hallward-Driemeier and Pritchett (2015: figure 2b)

emphasis on the inputs deployed by the organization rather than the outputs and outcomes achieved, or reduce organization capability to individual capacity, which leads to an over-emphasis on technical training. Conversely, the conflation of state capability with country-level legal or institutional features, like laws against corruption or good looking civil service legislation, assume that creating functional organizations begins with country-level action.

Achieving better outcomes requires better organizational capability for implementation. Before moving on to describe a pragmatic approach to building capability, we first need to examine the different kinds of capability various types of activities require. It is to this that we now turn.

**5**

**What type of organization capability is needed?**



Socrates: Suppose someone came to your friend [who is a doctor] and said “I know treatments to raise or lower (whichever I prefer) the temperature of people’s bodies; if I decide to, I can make them vomit or make their bowels move, and all sorts of things. On the basis of this knowledge I claim to be a physician; and I claim to be able to make other physicians as well by imparting to them.” What do you think they would say when they heard that?

Phaedrus: What could they say? They would ask him if he also knows to whom he should apply such treatments, when and to what extent.

Socrates: What if he replied, “I have no idea. My claim is that whoever learns from me will manage to do what you ask on his own”?

Phaedrus: I think they’d say the man’s mad if he thinks he’s a doctor just because he read a book or happened to come across a few potions; he knows nothing of the art.

(Plato, *Phaedrus*)

## Capability Matching

Imagine you are an athletic trainer and someone comes to you and says: “I want to build my athletic capability to compete successfully in a sport.” The first question you would ask is: “What sport?” If a person wants to be a badminton champion then quickness, agility, and flexibility are key capabilities. If a person wants to be a long-distance runner then cardio-vascular conditioning is a prime concern. A weightlifter’s capability is single repetition maximum power. Capability needs to be matched to the task at hand.

In the private sector high-capability organizations take a variety of shapes and sizes. In some domains high-capability organizations grow very rapidly and become very large while in others high-capability organizations stay very small. In others, organizational reach is large but not through direct expansion but through relationships. The company Facebook was founded in 2004 and ten years later had 1.2 billion active users. Walmart was founded in 1962 and fifty years later had 2.1 million employees—it is surpassed as an employer only by the US Department of Defense and China’s People’s Liberation Army. Yet Harvard University was founded in 1636 and 380 years later still has only 6,722 undergraduate students out of a US total undergraduate enrollment of over 28 million. McDonald’s serves 68 million customers *a day* in over 35,000 restaurants but it only owns and operates about 20 percent of those; the rest are franchises. Clearly in the private sector, where organizational form can freely adapt to function, there are a wide variety of ways, types, and scales at which organizations build capability.

Discussing how to build “state capability” independently of the answer to the question “capability to do what?” is bound to end in disappointment. The art of building the capability of state organizations has to begin with a taxonomy of the types of activities to be accomplished and the capabilities those activities need. Is an organizational building capability to deliver the mail? Set monetary policy? Deliver first contact curative care? Regulate point source pollution? This chapter delineates an analytical typology that uses four questions to classify tasks or activities into five types of organizational capability it requires. The five types are: *policymaking/elite services*, *logistics*, *implementation-intensive delivery of services*, *implementation-intensive imposition of obligations*, and *wicked hard*.

### A Basic Framework of Accountability

What motivates a teacher to teach well, or a doctor to give his best effort in treating patients? What is the difference between a tax collector who performs his job effectively, and a tax collector who takes bribes? Organizational capability often boils down to a functional system of accountability. There are two important dimensions of accountability: direct formal accountability to the organization, and indirect and informal accountability to a broader social and associational (e.g. professional, religious) norms.

Formal accountability is a relationship between two entities (person to person, organization to organization, many people as collective to organization leadership, organization to person). Formal accountability is embedded in an ongoing relationship that creates set of norms and expectations for both parties. Economists have used one type of accountability analysis, “principal-agent” models, to examine features of organizational size, scope, and incentive design as problems of contracting. In a purely market organization there are principal-agent problems that deal with resources (what does the agent work with?), information (how does the principal observe agent effort and outcomes?), decision-making (which decisions are made by the agent, which by the principal?), delivery mechanisms (who does the agent interact with?), and incentives (to what extent do payoffs to the employed agent depend on his/her performance?).<sup>1</sup>

Within any formal accountability relationship, there are four elements that structure agents’ choices. The World Bank’s World Development Report 2004 (World Bank 2004) calls these the “design elements” of an accountability relationship. Based on these, the agent chooses actions and hence the

<sup>1</sup> This is not to say, of course, that a principal-agent analysis exhausts the complexity of the service provision problem.

performance of the agent is endogenous to (a function of) the design elements but cannot be directly controlled.

The four elements of any formal accountability relationship are:

- **Delegation:** A specification of what is wanted from principals to agents.
- **Finance/support:** A flow of resources from principals to agents.
- **Information:** Once the agent carries out the required task some information is created that is available to the principal—although the essence of a principal–agent problem is that the information is necessarily incomplete as many other factors determine success or failure at the observable output/outcome than just the agent’s effort.
- **Motivation:** Based on the information the principal takes actions that affect the agent, which can affect the agent’s intrinsic and extrinsic motivation.

Life is full of garden variety accountability relationships. When your sink is clogged and you contract a plumber, you delegate and finance the plumber (by telling him to fix the sink with the promise of paying him if he does), he chooses his own preferred level of performance (by either fixing the sink well or not well), thereby providing you information (was the sink fixed?), and you are left with some control over motivation through the power of enforceability (to call the same plumber next time your sink clogs, give a tip for exceptional service, spread negative reviews if performance was bad, sue the plumber, or just to call a different plumber the next time). Every time you go to the doctor you become a principal in a potentially fraught principal–agent relationship, as many things could go wrong with each of the elements of the relationship.

*Delegation.* We go to a doctor for treatment when we experience symptoms. But as doctors have specialized knowledge and expertise we cannot tell them exactly what to do: which tests to run, how to interpret the results, and what treatments to give. Rather, we *delegate* in a way that gives broad discretion to the doctor: “Make me feel better.”

*Finance.* A doctor has to be compensated adequately to make her effort worth the time (and repay the years of training) but the structure of the financing arrangement creates different incentives. In a “fee for service” arrangement the doctor gets paid depending on the actions taken (diagnostics done, treatments given): this creates incentives for doctors to over-treat patients, and in turn creates a tension between the interests of the patient as principal (make me feel better at reasonable cost) and doctor.

*Information.* After whatever the doctor does, you as principal now ask how you feel. But it may well be the doctor does the best he or she can and your condition doesn’t respond. Alternatively, many visits to doctors are for “self-limiting” conditions that would have gotten better whether the doctor did any treatment or not. Hence the course of your own perceived health is not a good signal at all about whether the doctor



treated you appropriately or not. Das and Hammer (2007) found massive amounts of over-treatment by private sector health care providers, particularly the provision of quack treatments like steroid drips that have temporary “feel good” benefits to make patients *think* the provider was responsible when in fact the provider’s service is of no real medical value.

*Motivations.* Based on the information from the doctor’s visit the principal may take actions intended to either enhance or reduce the doctor’s wellbeing. These can be either extrinsic or pecuniary motivations—like repeat business or referring the doctor to others that increases the doctor’s income—or extrinsic motivators like direct praise of the doctor’s behavior.

Economists and other social scientists have used analysis of principal–agent relationships to examine how for-profit firms behave, as there are generically three principal–agent problems. One is between owners of firms (as principals) and those who manage the firm on their behalf (as agents). Owners have to design incentive mechanisms that deter managers from utilizing the assets of the firm to reward managers rather than the shareholders. This is complex because in modern corporations ownership is often quite diffuse and so many principals must coordinate to motivate few executives. The other generic principal–agent issue for a large private firm is how the management (now acting as principals) structures the employment relationship and compensation structure to motivate workers (as agents). Finally, firms must generate revenues and this is by the firm (as agent) providing a service demanded by another, with the firm’s clients now acting as principals.

The issues of accountability facing public sector organizations are considerably more complex than for private firms. When the public sector acts it has four continuously operating relationships of accountability between different numbers and types of actors. Each of these relationships of accountability has the four accountability elements of delegation, finance, information, and motivation.

- **Politics:** Citizens, as principals, act to hold politicians, as agents, accountable for how they exercise sovereign power.
- **Compact:** The executive/legislative powers of the state, as principals, act to induce public sector organizations (central banks, police forces, environmental regulators, teachers, courts) to provide functions.
- **Management:** The top management of public sector organizations, as principals, act to induce front-line workers, as agents, in the organization to carry out their functions.
- **Client Power:** Citizens, as principals, act directly on front-line providers and organizations to hold them, as agents, accountable for delivery.

Weak organizational capability can be the result of weakness or incoherence in accountability relationships *within* the organization, in particular the *management* relationship. But just because a tire is flat does not mean the hole is on the bottom. Weak performance of organizations can be symptomatic of weak elements of the *system* of accountability relationships into which an organization is embedded. Weaknesses in state organizations can start from weakness in *politics*, such that politicians and policy makers are not concerned with functional organizations, or from weak *compact*, in which the executive apparatus of the state does not provide the conditions for organizations to succeed.

There are four typical ways in which accountability in state organizations is incoherent.

*Mismatch of what is asked (delegation) and resources (finance).* This mismatch happens at all levels. As we say in the discussion of “premature load bearing” in Chapter 3, often the goals articulated by the state for the organization, the *delegation* element of the *compact* relationship, are far beyond what is possible with the *finance* actually provided. Many developing country governments just have control over far too few resources to do all of the functions as well and as universally as they claim to (and as they are pressured to by outside support). Thomas (2015) describes the situation of Afghanistan after the US invasion in which the Afghan state was expected to provide a wide array of services—from security to health to education to infrastructure—with a tax base per person that was a small fraction of what the USA had *even in 1900* when the US federal government took on very few tasks. This mismatch sets up governments and organizations for failure, as they cannot possibly be held accountable to do the impossible.

*Mismatch of delegation and information.* Another common accountability incoherence is that the *delegation* is at least nominally oriented to normative objectives but *information* is only collected (at best) on input utilization and process compliance. This is a common feature both of the relationship of the state to organizations (*compact*) and inside state organizations (*management*). For instance, a study of regulation of labor safety in Brazil found that the agency’s goal was safer work places but that their only information was about inspector visits and citations to firms about violations. For years, they never actually tracked—and hence could not motivate workers to pursue—workplace safety (until they did; more on this example later). Anyone who has worked in a public bureaucracy knows that at times all that matters is that what gets measured gets done—even if everyone knows that what is being measured doesn’t really matter. As discussed in Chapter 2, when delegation is vague or just inconsistent with the information collected then organizations can—and in many instances must—rely on isomorphism rather than performance as performance isn’t measured.

*Mismatch between delegation and motivation.* Another common failing is that even with *delegation* expressing laudable normative objectives, neither the organization (in the *compact* relationship with the state) nor front-line workers (in the *management*

relationship with the organization) are given the latitude and scope of autonomy to act, nor is there alignment of extrinsic and intrinsic motivation. That is, often organizations will get the same resources year after year whether they perform well at achieving their normative objective or badly.

*Mismatch in objectives across actors.* Even if there is one strong and coherent relationship of accountability the organization can nevertheless lack capability if there is incoherence between the true accountability relationships across the different accountability relationships. For instance, the leadership of an organization might attempt to strengthen the *management* relationship by collecting better information on outcomes and output and attempting to motivate (with carrots and sticks) providers to do a better job. However, this may conflict with other motivations of politicians in the *delegation* function. Politicians may want to use public sector organizations as a means to give patronage jobs to political supporters. This is clearly incompatible with removal of dysfunctional workers. This is incoherence across the rows of Table 5.1—the different actors in their role as “principal” to “agents” really have very different objectives. Again, one can expect failure out of a public sector system in which the citizens, politicians and policy makers, leaders of public sector organizations and front-line workers all have completely different notions of what “success” would look like.

But before one can discuss in detail how to construct effective and coherent relationships of accountability within organizations, governments, and in broader systems, there first has to be a clear analytic of what kind of capability is required, and how that capability aligns with accountability.

**Table 5.1.** Four relationships of accountability (columns) by four elements of each relationship of accountability between Principals(s) (P) and Agent(s) (A) (rows) as a diagnostic for the systems of accountability within which state organizations operate

Four design elements of each relationship of accountability (Principal (P) to Agent (A))	Principal–agent relationships			
	<i>Politics:</i> Citizens to “the state”/politicians (many P to one A)	<i>Compact:</i> “The state” to organizations (one P to one A or one P to many A with non-state providers)	<i>Management:</i> Organizations to front-line providers (FLP) (one P to many A)	<i>Voice/Client power:</i> Service recipients (parents/children) direct to FLP/organizations (many P to one A)
<i>Delegation:</i> Specification of what P wants from A				
<i>Finance:</i> Resources that P provides to A (either in advance or contingent)				
<i>Information:</i> P collects information on performance of A				
<i>Motivation:</i> How is A’s wellbeing contingent on performance? Change to motivation?				
<ul style="list-style-type: none"> <li>• Intrinsic</li> <li>• Extrinsic</li> <li>• Exit (force out)</li> </ul>				
Performance of agent (endogenous)				

## Classifying the Type of Organizational Capability Needed: Four Questions

We ask the reader to think of any concrete public policy objective. The more specific the task and the more specific the context specified, the better. “Education” is too broad, whereas “remediating reading proficiency deficits in Bihar, India” or “vocational training in rural Sindh, Pakistan” is the desired level of granularity. “Public financial management” is too broad, whereas “management of procurement of medium-sized goods and services in Mozambique” is fine. “Microfinance” is too broad whereas “micro-savings programs for urban informal workers in Durban, South Africa” or “providing finance to promote entrepreneurial finance medium-sized enterprises in Saudi Arabia” is better.

It will help if you, as a reader, take time to write down a policy objective that interests you before proceeding. (We’ll wait while you find a pen and paper. Back with pen and paper? OK.)

We want you to answer four questions about what it will take to accomplish your policy objective. The goal is to classify the type of capability an organization would need to be successful. This classification scheme cuts across sectors as within each sector (education, regulation, justice, infrastructure, health) there are analytically very different types of tasks.

Each question begins: “Does the successful accomplishment of your policy objective require actions or activities that are . . . ?”

1. Transaction intensive? The first question is whether the accomplishment of the task is going to require many people or few people (or at least many transactions). For instance, a central bank can set some macroeconomic and monetary policies with decisions of a few individuals that are, more or less, self-implementing. So even though the USA’s \$20 trillion economy is unfathomably complex, key elements of monetary policy are made by a dozen or so individuals who themselves draw on remarkably few people. This is not transaction intensive. In contrast, primary schooling requires that lots of teachers work with lots of students every day. Teaching in primary schooling is transaction intensive. There are also elements of primary schooling, like setting the curriculum or creating textbooks, which may involve relatively few experts and hence are not transaction intensive.

Policing is transaction intensive. Passing laws is not transaction intensive. Dispute resolution is transaction intensive. Appellate courts are not transaction intensive. Procurement and spending budgets are transaction intensive. Setting a budget is not transaction intensive.

*“Does the successful accomplishment of your policy objective require actions or activities that are transaction intensive?”* Write down the answer. (And yes, we realize many readers’ answer will always be “But it is more complicated than

that”; even so, we ask the reader to shake that impulse off for now and just write down yes or no.)

2. Discretionary? Services are *discretionary* to the extent that their delivery requires decisions by the agents responsible for implementation to be made on the basis of information that is important to success but inherently imperfectly specified and incomplete, thereby rendering them unable to be mechanized. Returning to Chapter 4’s definition of a “policy formula” as a mapping between “facts” and “actions of an agent,” whether or not achieving the policy objective requires agents to exercise discretion depends on three aspects of the policy formula:

- Does successful implementation require agents to use professional training, experience and judgment, or are the relevant facts of the policy formula obvious or easily ascertainable? Can policy implementation be reduced to a script that relies nearly exclusively on “hard” or “thin” information?
- How costly is it for a third party to verify and adjudicate in a contractually enforceable way what the “true” facts of a given situation are?
- How sensitive is the link between facts, actions of the agents, and outcomes?

Vaccinations and ambulatory curative care illustrate the difference in “discretionary.” Both are transaction intensive, as they involve a face to face meeting between an agent (health care provider of some type) and the person receiving the service in order to be successful. But ambulatory curative care requires that the action taken be tailored to each patient so that a diagnostic process arrives at the right treatment (if any) can be discerned. A person presenting with severe pain radiating from their chest must be treated differently from a person presenting with pain in their knee for the curative care to be effective. In contrast, nearly every child gets the same vaccinations for childhood diseases. The relevant policy formula fact for vaccinations is the age of the child and their vaccination history, neither of which involves information which is difficult to ascertain or hard to verify.

Nearly all sectors and activities involve some elements that do and do not require local discretion. Policing requires that agents go into complex, often dangerous and tense situations and make hard, sometimes life-and-death decisions. No matter how finely specified the law, policemen operate with discretion. In contrast, giving traffic tickets is transaction intensive but need not involve discretion.

In primary schooling, getting textbooks delivered to each school on time and one to each child is an important task, but not one that requires discretion. In contrast, quality classroom instruction and teaching requires an ongoing interaction in which teachers tailor their actions to the students,

individually and as a class, on a near-continuous basis. As such, these locally discretionary decisions usually entail extensive professional (gained via training and/or experience) or informal context-specific knowledge.<sup>2</sup>

*“Does the successful accomplishment of your policy objective require actions or activities that require implementing agents to exercise local discretion?”* Write it down. If some elements of accomplishing the policy objective do and others do not require discretion, specify which. We’ll wait. OK.

3. Service or obligation? When the government’s agents interact with citizens in the course of implementation they are either *providing a service* or *imposing an obligation*. Taxes, for instance, are the price of civilization and in democracies “the people” collectively agree to be taxed (Pritchett and Aiyar 2015). But in the act of collecting a property tax or sales tax or income tax the agents responsible for tax implementation are imposing an obligation. Similarly, the police necessarily interact with criminals. While this is a service to the society at large, to those who seek to avoid the law the role of the police is to impose obligations.

This distinction of whether the implementing agents are providing a service or imposing an obligation in their typical interaction is key for two reasons. One, it structures the possibilities for how the “client” interaction can be used for accountability. When “service delivery” is the goal then incorporating the feedback of direct users (of water, of schools, of health services, of roads) into the accountability of agents expands the range of inputs and information available to assess performance. In contrast, it is much more difficult to survey criminals about how the police treated them or put too much emphasis on “customer satisfaction” for tax auditors or environmental regulators (Chapter 4 reported how putting “clients” in charge of contracting for the reports of their own emissions lead to the predictable result of biased reporting). Two, in the imposition of obligations the decisions made in implementation can be high stakes and hence the pressure brought to bear on the agents of implementation to mis-declare the “facts” of the policy formula in order to produce an outcome desirable for the citizen but which thwarts the policy objective are high. Corruption is the ever-present risk in the imposition of obligations. Chapter 4 showed that even compliance with very mild obligations like demonstrating driving skill to get a driver’s license can be undermined by payments to implementing agents.

<sup>2</sup> Forgive us the potential confusion, as “discretionary” more appropriately refers to the *mode* of the arrangement of an activity (which, at some level, is an endogenous choice) while we are using the term to refer to the underlying characteristics of the activity that lead it to be provided in a discretionary manner (or suffer losses from *not* being provided with arrangements that provide for discretion).

*“In their routine activities are the implementing agents providing a service or imposing an obligation?”* Write it down.

4. Based on known technology? Many tasks, like ambulatory curative care, are complicated and require agents to exercise discretion. But doctors can rely on bodies of knowledge and training and handbooks and even protocols to follow for diagnosis and treatment. Running a central bank is not an easy task, but there is a body of knowledge and empirical evidence and a strong professional consensus about many components of the decision making that central bank leaders and staff can rely on (or ignore at their peril). But often success requires that the agents of an organization go beyond following established protocols; they must actually innovate and move beyond the frontier of known technology and accepted practice to achieve success.


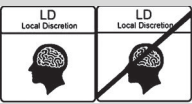

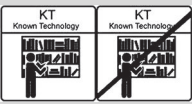
This need to go beyond the known technology and actually innovate can arise for a variety of reasons. One is that new situations and new technological shifts may mean that what was the known technology no longer applies but no one is (yet) sure what does. The other reason is that human beings are just enormously complex and how to motivate them to do certain things and not others cannot be reduced to a formula. So, while the technology of weight loss is relatively well known there are very few successful *programs* to induce weight loss in others. This isn't to say nothing is known but just that, for instance, ambulatory curative care, or treating specific disease conditions that patients present with at facilities, is based on a known technology while inducing populations to reduce risk has proven enormously more complex. A final reason an activity might be wicked hard is that one is promoting something like “entrepreneurship” that itself means individuals need to innovate.

*“Does successful implementation require innovation from agents as opposed to reliance on an agreed upon technology?”* Write it down.

## A Typology of Tasks by Capability Required

Based on these four questions (illustrated in Figure 5.1) we create a taxonomy with five principal types of tasks based on the type of organizational capability and how the task facilitates or complicates building this capability: *policy-making and/or concentrated (elite) services*, *logistics*, *implementation-intensive service delivery*, *implementation-intensive imposition of obligations*, and *wicked hard* (illustrated in Figure 5.2)

*Policy formulation (and elite concentrated services)*. The first category is distinguished from all of the others by its nature of not being transaction intensive. This is really a combination of categories of those for which implementation requires relatively few people. The task of policymaking itself—of articulating the policy formula, objectives,

Is your activity...		Does producing successful outcomes from your activity....
Transaction intensive?		Require many agents to act or few
Locally discretionary?		Require that the implementing agents make finely based distinctions about the “state of the world”? Are these distinctions difficult for a third party to assess?
Service or imposition of obligation		Do the people in direct contact with your agents want or not want the agent to succeed?
Based on a known technology		Is there an accepted handbook or body of knowledge for doing what you are trying to do or will this require innovation (not just context)

**Figure 5.1.** Four key analytic questions about an activity to classify the capability needed




















and causal model—is nearly always possible (if not desirable) to do with relatively few people.

This category can also include apex or elite institutions in many sectors as, in the larger scheme of things, these require very few agents. Nearly every sector has apex institutions—the tertiary hospital, the research university, the highest appellate court, the central bank, for example—that may only involve a few hundred core professionals and hence are not transaction intensive. Even within organizations there are often “elite” units, as in most militaries. When these are in separate organizations or distinct within an organization this creates a different dynamic for organizational capability as their concentrated and apex nature makes peer monitoring and *esprit de corps* the primary accountability mechanisms. Heart surgeons care about what other heart surgeons think of them, Navy SEALs about what other SEALs think of them. We mention this because many countries maintain impressively strong elite or apex institutions even in otherwise largely dysfunctional and/or corrupt environments. However, these successes don’t necessarily point to a potential for broader success at building capability for more transaction-intensive activities.

*Logistics.* A second type of capability is the ability of organizations to induce large numbers of agents to follow relatively simple scripts that rely on easily observable



## What type of organization capability is needed?

		Examples	
		Health	Finance
Policy making/elite services	   	Iodization of salt	Monetary policy
Logistics	 	Vaccinations	Payment systems
Implementation intensive service delivery	   	Curative care	Loans
Implementation intensive imposition of obligations	   	Regulation of private providers	Regulation of private providers
Wicked hard	    	Preventative health	Equity financing of start-ups

**Figure 5.2.** The five types of activities that have different capability needs in implementation

and judicable facts. In financial matters, an example is retail banking transactions, many of which can be carried out by a junior clerk (or for the most routine transactions, a machine).<sup>3</sup> To implement a “program” the agents of the organization need only to stick to a relatively fixed “script” (Leonard, 2002; Dobbin, forthcoming), in which the choices are few and judging the choice appropriate to the situation is relatively easy.

*Implementation-intensive service delivery.* Tasks that are discretionary (unlike logistics) and transaction intensive (unlike *policy/elite services*) we classify as “implementation intensive” as they require large organizations with agents engaged in complicated actions. The key distinction is whether these actions are devoted to “services” in which agents interact with people who (in principle) directly benefit from successful implementation.

*Implementation-intensive imposition of obligations.* The imposition of obligations can be implementation intensive, like policing, taxation, or regulation. Implementing such tasks entails overcoming the resistance of those upon whom obligations are being imposed; recipients may seek to use everything from passive resistance to physical threats to material incentives (bribes) to induce agents to be less than diligent in carrying out their duties.

*Wicked hard.* The most difficult tasks that combine *transaction intensive* (a large number of agents need to participate), *discretionary* (the decisions made by agents are based on

<sup>3</sup> The name “programs” has the advantage of following the usual development nomenclature (of policies versus practices) but also invoking the idea of a computer program.

difficult-to-verify knowledge) and *not based on a known technology* require a category all their own. As we are based in Boston we call these “wicked hard” tasks (where “wicked” is the local vernacular for “very,” not “evil”).

Going one level of specificity further, our taxonomy allows us to think analytically about the diverse range of tasks within a given sector. The taxonomy is at the level of tasks because these classifications do not correspond neatly to sectors; it is not that case that “education” is all “implementation-intensive service delivery” or “finance” is all “concentrated/elite.” Rather, within every sector and subsector there are examples of each type of task.

For instance, a girl turns up at elementary school eager to learn: What has to happen to provide her with high-quality instruction? The teacher has to know what to teach her and when, which means the curriculum has to have been established, preferably along with some norms for learning expectations grade by grade. This is a *policy formulation* problem as it is primarily technocratic and not transaction intensive. The girl has to be near a school with adequate facilities and learning materials. This is primarily a *logistical* problem as building schools and buying blackboards and desks can be reduced to a (reasonably) standardized process. There has to be a teacher there that know what to teach, knows the material, and knows how to teach it. This is *implementation-intensive service delivery*, as teachers exercise local discretion, hour by hour, class by class, and child by child. There are also elements of the *wicked hard*, as innovations are needed to not just keep learning levels constant but increase them over time. Similarly with procurement: the formal rules may be determined by a select committee (*policy formulation*), but ensuring that all relevant staff members in an organization know what these rules are might entail preparing a handbook and an online tutorial that can test knowledge (*logistics*). Knowing how to apply the rules in response to marginal, novel or ambiguous cases, however, will entail considerable discretion on the part of adjudicators (*implementation-intensive service*), while enforcing them in instances where there might be potentially lucrative kickbacks on offer (*implementation-intensive obligations*) is likely to entail adherence to strong professional norms and internalized codes of conduct.

### *Implications for Organizations of the Capability Taxonomy*

The purpose of the foregoing sections was to create an analytically grounded classification of types of organizational capability. This taxonomy is necessary because there is a tendency, in practice if not academically, to distinguish between policy formulation and policy implementation. Capability in policy formulation is the ability to frame objectives, analyze options and, based on

the relevant methods and evidence, choose the best policy. Capability in policy implementation is regarded in strictly logistical terms as creating organizational procedures that produce the process compliance of agents with the policy as formulated.

If everything the public sector did was “logistical” in our sense, and the organizations already had adequate capability, then this wouldn’t be a terrible approach. In Wilson’s (1989) classic on bureaucracy he points out that the US Social Security administration was roughly as cost-effective in the task of delivering old age pension checks to eligible recipients as any organization, private or public, could hope to be. This is because the task is entirely logistical: the policy formula for eligibility is well defined in hard facts (based on age and duration of contributions) and maps to a simple clear action (mail a check of a certain amount). Full stop. At logistical tasks like that, nothing beats a bureaucracy.

Carpenter’s (2001) history of the emergence of the US Postal Service (and others) as a modern Weberian bureaucracy recalls a period in which the bureaucracy was seen by forward looking reformers as the solution to the problems that riddled existing systems, which at the time were captured by local political interests and patronage networks. It struggled its way into existence by legitimizing stronger bureaucratic control over the post based on its superior efficiency. Indeed, even as post offices around the world are being corporatized and many functions shifting to private sector firms, those firms are competing to be more effective bureaucracies. But if one compares FedEx or DHL or UPS to the US Postal Service they are nearly identical in the way they are organized and operate—because they are competing to be a better bureaucracy at doing logistics.<sup>4</sup>

We have yet to meet anyone who can name a large firm of dentists. Everyone knows their dentist, but almost always in market economies their dentist works alone or in a partnership with one or two other dentists. Dentistry isn’t policy-making and dentistry isn’t logistics. Dentists have practices.<sup>5</sup> A “practice” is the organizational form for implementation-intensive service delivery when it is not in the public sector. Most law firms, physician practices, universities, household contractors, therapists, marriage counselors, music teachers, and sports coaches<sup>6</sup> are incredibly small relative to the national market. Even

<sup>4</sup> Indeed, the slogan of UPS for a time was “We Love Logistics”; interestingly, for our purposes at least, it is currently “United Problem Solvers.”

<sup>5</sup> Our rendering of “practices” should not be confused with Sunstein and Ullmann-Margalit’s (1999) intriguing notion of “second-order decisions,” which they define as the various strategies adopted in complex environments (by key actors such as judges, politicians, administrators) to *avoid* actually having to make discretionary decisions. Our discussion is more akin to, and in some senses builds on, Heifetz’s (1994) useful distinction between “technical” and “adaptive” decision-making.

<sup>6</sup> The exception that proves the rule in the “coaching” industry (e.g. music lessons, sports instruction) is the emergence of large organizations that provided courses that prepare students for a *standardized* exam.

though they are transaction intensive, the need for “local discretion” makes these tasks a mismatch with the logic of the logistical imperatives of large-scale, routinized, administrative control of agents to produce process compliance.<sup>7</sup>

The classic bureaucracy is appropriate for logistical tasks for which simple accountability is sufficient for adequate performance; “delegation” (what it is the agent should do) and “information” (measurement of the agent’s performance) are completely reducible to easily judicable facts. The post office is the quintessential example, as everything about what each agent should do to each parcel is easily contained in a few bytes (the address and the class of service). This creates compatible internal (management) and external (politics, compact, client power) formal and folk cultures of performance. What the postal clerk is expected to do by his managers (did he deliver the mail?) is measurable in exactly the same terms that clients can measure it (did my mail arrive?), the overall organization can be measured (what percentage of parcels were delivered on time?), and the political system can talk about it (is the post office doing its job at a reasonable cost?).<sup>8</sup> Note that this is a characteristic of task, not sector, and not whether it is in the public or private sphere. In the United States the internal mechanics and size and structure of organizations that deliver packages in the private sector (FedEx, UPS, DHL) look organizationally nearly identical to the post office—same trucks, similar uniforms, similar thin accountability tracked with thin information.

Everything about the way an organization tends to work depends on that task it confronts at its “operating core” (Mintzberg 1979). All large organizations have multiple elements and these elements have different capability requirements, but the “operating core” is the part of the organization that is the unique producer of value and *raison d'être* for the organization’s existence. Law partnerships, universities, and architectural firms all have units that handle accounting but accounting is not their “operating core”; it is a service function deployed in the interests of the technical core—legal services, teaching and research, designs of buildings respectively. When organizations can choose their structure the overall size, scope, and culture of the organization is driven by the characteristics

<sup>7</sup> In policing, for example, Goldstein (1990: 8) concludes that “studies identified the enormous gap between the practice and the image of policing. They identified problems in policing that were not simply the product of poor management, but rather reflections of the inherent complexity of the police job: informal arrangements . . . were found to be more common than was compliance with formally established procedures; individual police officers were found to be routinely exercising a great deal of discretion in deciding how to handle the tremendous variety of circumstances with which they were confronted.”

<sup>8</sup> The postal service itself, it should be noted, rightly seeks to convey a more noble account of its activities. Literally chiseled in stone on the National Postal Museum in Washington, DC is the following inscription, reminding visitors that a postal worker is actually a: Messenger of Sympathy and Love/Servant of Parted Friends/Consoler of the Lonely/Bond of the Scattered Family/Enlarger of the Common Life.

of the operating core. If the operating core is logistics, the organization reflects that. When the technical core is a “practice” the organizational structure reflects that, while incorporating service functions operating as logistics.

The major risk of not having an adequate taxonomy of organizational capability is the risk of mismatch between the approach to building an effective organization and the task at its technical core. As we articulate in the future chapters, the dominant tendency in public sector organizations has been to impose the Procrustean bed that public sector organizations are either “policymaking” organizations or “logistical.” Organizations that are responsible for implementation are treated as standard Weberian bureaucracies—which is fine if tasks that are logistical are in the technical core, but not at all fine (i.e. can fail badly) when more implementation-intensive activities are in the technical core.

Pritchett (2013, 2014) illustrates this mismatch in primacy education. As we saw above, primary education requires tasks of different capability types: *policymaking/elite* (standard, curriculum, assessment), *logistics* (building schooling, delivering inputs), and *implementation-intensive service delivery* (classroom teaching). It is clear that when delivered outside of public sector contexts that if instruction is the technical core then organizations are typically organized as “practices” because it is implementation intensive. However, for a variety of historical, political, and intellectual reasons primary education came to be dominated by “spider”<sup>9</sup> organizations which approached public education as a logistical problem of expanding enrollments. This mismatch between an organizational structure well adapted to logistics led to a situation in which the goal of expanding enrollment—through the construction of buildings, buying of inputs, hiring of teachers—has been met but many countries are admitting to a “learning crisis” as the quality of teaching and student learning is, not at all surprisingly, given the inversion of the operating core, very weak. In one state of India, enhanced budget and programs were able to improve all of the measures of facilities and logistics—and yet in less than a decade the system lost a million students to providers as parents chose to pay for private education rather than enroll children in the public system for free (Pritchett 2014).

## Accounting and Accounts in Accountability

Let’s return to accountability relationships and systems of accountability in light of the taxonomy we’ve just outlined. Packed into “accountability” are

<sup>9</sup> The terminology of “spider” and “starfish” as types of organizations comes from Brafman and Beckstrom (2006).

two fundamentally different notions embedded in different variants of the same word: *account* and *accounting*.

An “account” is the justificatory narrative I tell myself which reconciles my actions with my identity: am I fulfilling my duties? An account is the story of my actions I tell to those whose opinion of me is important to me (including most importantly, myself, but including family and kinsmen, friends, co-workers, co-religionists, people in my occupation and other people whose admiration I seek) that explains why my actions are (or, if the account is a confession, are not) in accord with a positive view of myself as an agent.<sup>10</sup>

Following the notion elaborated by Geertz (1973)<sup>11</sup> of a “thick description,” we create the distinction between “thick accountability” (the *account*) and “thin accountability” (the *accounting*).<sup>12</sup> Thick accountability is inevitably a folk process in which behavior is shaped by norms that are unwritten and informal, while thin accountability can be (re)produced within formal sector organizations.

Our argument is that successful organizations rely on a combination of thin and thick accountability, both internally and externally. Once agents have lost the sense that their account, either to their organization or to their fellow citizens or their fellow professionals, depends on their carrying out their formal duties, no amount of *accounting* can make a difference. A strong *account* and indeed *thick accountability* is required in public service delivery that is implementation intensive (and more so for the wicked hard).

As we saw in Chapter 4, when accounts and accounting diverge, organizations can often “fix” the accounting and thereby make the “administrative facts” of accounting a complete fiction. A public agent’s *account* actually rests squarely on many *folk* understandings. What is the *account* of the doctor in the Madhya Pradesh study, who doesn’t get off the phone when dealing with a patient presenting with chest pains? What is the *account* of a teacher who doesn’t smile at the students (much less laugh, joke, or talk to them)? What is the *account* of a policeman who takes bribes from motorists? Or the bureaucrat who issues licenses without the compliance? Fixing the *accounting* cannot fix the *account*, and the *account* is in the realm of the *folk*.

Our argument is that successful organizations are built on internal and external accounts for which accounting provides some support and plays some role. Think of any organization with a long track record of success (on the organization’s objectives): Oxford University, the Catholic Church, the

<sup>10</sup> Our views and description of an “account” is strongly influenced by MacIntyre (2007) and his views on Aristotelian notions of virtue.

<sup>11</sup> Geertz himself acknowledges the priority of Gilbert Ryle in the idea of “thick” description but he popularized the notion as a methodological stance.

<sup>12</sup> The term “thick accountability” is also used in Dubnick (2003), who describes the idea with many of the same meanings and implications we use here.

Red Cross, the US Marine Corps, Exxon. These organizations survive and thrive because key agents believe it is important that their account of what they do (indeed perhaps who they are) accords with the purposes of the organization. Indeed, the three of us can attest from experience that high-capability universities do not thrive because professors do accounting for their behavior, but professors at thriving universities do have an account of what they do because they are professors and this account is important to them.

Moreover, to some external audiences the organization has to justify itself for legitimacy and ultimately resources. This external accountability is not driven by accounting or detailed measures of cost effectiveness or proven impact or reducible to precise figures, but they have to continually prove to key constituencies that they work because there are competitors for their support base (students and faculty for universities, adherents for religions, donors and volunteers for philanthropic organizations, funding among other public uses for marines, capital markets and customers for Exxon) and if these external actors no longer believe the organization's account then they lose traction with their internal agents and external constituencies no matter what the accounting says.

Consider for a moment the thickness of information.<sup>13</sup> "Thin" information can be thought of as information that is easily amenable to being reduced to "information" in the Shannon (1948) sense of information as messages encoded in bits and bytes. "Is it Tuesday (right here, right now)?" is a "thin" question on which we all can readily agree and, if necessary, have third-party adjudicators agree to what the fact of the matter is. It is easy to create high-powered incentives on thin information: "I will pay you \$10 if it arrives on Tuesday and only \$5 if it arrives on Wednesday" is an enforceable contract because the fact of "Tuesday" is easily judicable and hence Tuesday is a contractible.

The world is, however, immensely thick. Only a tiny fraction of our everyday existence can be reduced to thin information. Was Tuesday a nice day? Was the bus driver rude to you on Tuesday? Was the Starbucks clerk friendly to you on Tuesday? Were you in a good mood on Tuesday? Was your lunch delicious on Tuesday? Were you attentive to your partner on Tuesday? Did you do your best at work on Tuesday? All of these are potentially important determinants of our wellbeing, but none of these are easily contractible. They

<sup>13</sup> The central issue in the "economics of information" is the costliness of the adjudication of information. The economics of information as an explanation of institutions and organizational behaviors starts with Williamson (1975), then builds through principal-agent theory to organizational compensation schemes (Lazear 1995), organizational strategies (Milgrom and Roberts 1992), allocation of authority (Aghion and Tirole 1997), and the theory of the boundaries of the firm itself as a problem of contracting (e.g. Hart and Holmstrom 2010). The economics of information approach has also been applied to delegation, contracting, and the scope of public sector organizations (e.g. Hart et al. 1997; Laffont and Tirole 1993).

are not judicable because the difficulty of establishing third-party intersubjective agreement on just what the facts on Tuesday really were about: nice, rude, friendly, delicious, inattentive, best effort, etc.

How does this “thick” versus “thin” distinction relate to the capability of the state for policy implementation?

When attempts at thin accountability—making agent rewards depend on judicable “facts” (like attendance, like were actual taxes owed)—are impossible because the overall institutional environment is weak, then even using incentives will not work.<sup>14</sup> Besley and McLaren (1993) used a model of tax collection and tax inspection to note that when punishment based on observed actions was sufficiently difficult (the probability of an effective audit with punishment was low) there was no advantage of paying a fixed wage high enough to deter corruption or encourage honest inspectors. In their model when actions cannot be contracted then a “capitulation wage”—paying low wages and admitting all tax inspectors who were not monitored would be corrupt, which results in a cynical and entirely dishonest set of tax inspectors—was the net revenue generating strategy.

Besley and Ghatak (2005) explore this issue referring to organizations with “mission” (what we call *internal folk culture of performance*) and show that if organizations can be matched to mission then this non-pecuniary form of motivation reduces the need for (if not desirability of) high-powered pecuniary incentives. The better organizations are able to recruit individuals motivated by mission (individuals whose personal thick accountability is strong) the less the organization needs to rely on thin accountability.

As mentioned earlier, logistical organizations such as FedEx can rely on thin accountability to function. In organizations that perform tasks that are predominantly of more difficult, non-logistical types (e.g. concentrated, implementation-intensive service delivery, implementation-intensive imposition of obligations, wicked hard) the internal folk culture required for performance is at odds with a formal culture of thin accountability (see Table 5.1). A high-performing university or hospital (either in the public or private sector) requires a culture of accountability for performance. But this does not translate into professors being tracked minute by minute by GPS. You cannot reduce the delegation of what a professor should do to be a high-quality professor to a sequence of bytes. The same is true of nurses. The same is true of policemen.

While there might be some minimal performance criteria that are thin (like attendance), what has been learned from decades of studies of schools, for

---

<sup>14</sup> One of the key insights of principal-agent theory is that the less precisely the desirability of the actions of the agent can be measured, the less high-powered the incentives should optimally be (e.g. Holmstrom and Milgrom 1991, and for an application to civil service Klitgaard 1997).



instance, is that the thin accountability parts of schooling do not affect education very much. While good teachers—as measured by their performance—matter a lot to student learning what being a “good teacher” means is not reducible to thin criteria like degrees or age or years of service (Chetty et al. 2011; Rivkin et al. 2005), or even, we would argue, student learning alone. Similarly, inputs alone, the kinds of things that education management information systems can measure and track, just do not have a very strong connection with the education a child receives—or the inequality in outcomes across schools (Pritchett and Viarengo 2009)—as “implementation-intensive service delivery” good schools require thick accountability as well as thin accountability, internally and externally.

Valuable local *folk* practices—idiosyncratic knowledge of variables crucial to the welfare of the poor (e.g. soil conditions, weather patterns, water flows)—get squeezed out, even lost completely, in large centralized development programs designed to address these issues (see Ostrom, 1990; Scott, 1998). The myriad informal “practices” that indigenous communities in particular have evolved over the millennia to address these concerns may be clearly ill-suited to the complexity and scale of modern economic life, but the transition from one set of mechanisms to the other cannot be made in a single bound. While not attempting the transition at all is a prescription for continued poverty, revolutionaries from Stalin to Mao to Nyerere to contemporary “shock therapists” have imagined that it was actually possible and desirable to ruthlessly “skip straight to Weber”—but with patently disastrous results. In the murky middle ground between the public services and risk management systems of “Djibouti” and “Denmark” lies the need for a much more delicate articulation of the two, an articulation that the technocrats and bureaucrats of large development (and other) agencies inherently and inevitably struggle to resolve.

These more graphic examples of large-scale bureaucratic disaster, however, have their counterpart in a host of smaller everyday instances of repeated failure by standardized delivery mechanisms to provide basic services to the poor. Some of these problems, of course, stem from the fact that in many instances the state itself (for whatever reason) was unable and/or unwilling to provide the services that citizens wanted. Our concerns, however, apply to systemic services failures that routinely occurred even in settings where intentions and resources were reasonably good.

In our taxonomy of the capability requirements of activities we want to stress that *moving* an organization from lacking capability to capability is itself *wicked hard*. That is, if one has a dysfunctional post office that is not delivering mail in an effective way the capability needed is pure logistics but *moving* an organization from lacking the capability to do logistics to having the capability to do logistics is wicked hard. Why? Because building capability, even using

our approach of PDIA is not a “known technology.” Changing organizations is changing the behavior of people and many aspects of human behavior are just too complex to pretend it can be reduced to a simple formula.

The same is true of moving from organizations with capability for logistics but not for the implementation-intensive components of what the organization needs to do. For instance, in building out a system of basic education capable of producing learning some elements—like building the school buildings or assigning teachers to schools, or ensuring attendance—can be reduced to logistics but other elements, like teachers displaying concern, cannot. Getting a large-scale organization from logistics to implementation-intensive capability is itself wicked hard and requires something like PDIA.

In Part II of the book we turn to the practical task of building the capability you need for success.