

PART I

Setting the Scene

I

The Puzzle of Reform

Congratulations! You've just been appointed an advisor to the head of the civil service. Your job is to recommend how to improve the effectiveness of dozens of government ministries and departments. These organizations, in turn, are responsible for delivering public services, developing policies, building infrastructure, implementing regulations, and nearly everything else the economy and society depend on. But your civil service is widely perceived—by civil servants, politicians, and ordinary people alike—as ineffective, even dysfunctional. Budgets are tight, and there are major political and legal constraints on your ability to hire or fire personnel. How would you advise the head of civil service to go about designing and implementing reforms to improve the performance of the civil service?

This question represents one of the most pressing practical problems facing governments around the world. It also poses an intellectual challenge of the first order. Ultimately, civil service reforms aim to change the day-to-day behavior and routines of the huge number of people who collectively make its decisions, enforce its rules, and deliver its services. But government is not a machine with simple levers that produce predictable responses when pulled. It is a complex system of interdependent organizations and teams inhabited by a diverse range of people, each with their own interests, preferences, experiences, and biases. The wrecks of failed reforms that litter the junkyard of public administration history can testify to the difficulty of transforming bureaucracies. What can this history of reform efforts teach us about the prospects for systemic reform to improve the inner workings of government?

This book aims to answer these questions by documenting and analyzing how six countries in Africa have repeatedly sought to reform their civil services over

the past three decades. These countries have been global hotbeds of experimentation in public administration reform, collectively undertaking over one hundred performance-oriented reforms during this period. The types of problems they have sought to solve and the solutions they've attempted to adopt and implement will be familiar to reformers around the world. Sometimes these reforms have been driven by public servants, and at other times they've been driven by donors or politicians—or often combinations of two or three of these. They have been influenced by global trends as well as homegrown ideas. So public managers and researchers from around the world have much to learn from the history and patterns of reform in these six countries.

I use a rich range of primary data and secondary literature to document and analyze the content, process, implementation, outcomes, and politics of each of these reform efforts. I then use this comprehensive portrait of reform efforts to identify repeated patterns in design, implementation, and outcomes that recur across many reforms, across all six countries, and across time. Of course, every reform effort and context is unique to some extent, and when we look at each reform in isolation, there are many potential explanations for why events unfolded as they did. But if we don't look historically and comparatively across efforts and across contexts, we risk missing the forest for the trees. Beyond these particularities and idiosyncrasies, are there repeated patterns we can identify and learn from?

I argue that there are two features that characterize the vast majority of large-scale reform efforts and that help to explain the persistently disappointing results of these reforms. First, most reforms focused mainly on changing formal structures, rules, and processes. This took different shapes across different reforms, but the common thread is an emphasis on creating formal processes that would compel or incentivize bureaucrats to behave differently. Second, reforms were typically conceived of and executed as discrete, one-off, often time-bound interventions. This means that reforms often ended up as projects with their own predefined work plans, acronyms, timetables, budget lines, and implementation teams. In other cases, they took the shape of new laws or structures on paper that, in practice, were implemented halfheartedly (if at all) or had purchase for a year or two before fading away. Together, these constitute reforms' two main *mechanisms of failure*. I demonstrate how and why this approach was largely ineffective at changing the complex system of organizations, rules, and perceptions that constitute government bureaucracy, and how and why it undermined reforms' potential to change the way that bureaucrats think about and carry out their everyday work.

At the same time, the history of reforms is much more than a story of failure—most reforms did achieve something, even if they fell short of their lofty goals. While reforms tended to make more changes on paper than in practice, some

reforms did have positive and meaningful impacts on the actual day-to-day work practices of civil servants. When they did so, it was usually through two linked *mechanisms of success*: creating opportunities for civil servants to discuss performance and how to improve it, and building energy and momentum to do so. These mechanisms of success worked not because they forced unwilling bureaucrats to improve their performance, but rather because they helped create an environment where bureaucrats who wanted their teams and organizations to work better were able to find ways to make this happen. Though these mechanisms of success often coexisted with mechanisms of failure within the same reform effort, they usually took a backseat role in the dominant reform approach.

If this is the diagnosis that explains the track record of past reform efforts, what is the prescription for how future reformers should approach their work differently? What can we learn from this history to help us better advise a head of civil service or other prospective reform leader on how they should go about improving performance across a whole bureaucratic system?

I argue that, instead of treating reform as a one-off change to formal processes, reformers should approach it as *catalyzing an ongoing process of continuous improvement in actual practices*. This alternative approach reframes the goal of reform as the improvement of day-to-day bureaucratic behaviors, both in terms of adherence to formal processes as well as the undertaking of the many tasks that are important for performance but can't be fully codified on paper. It reconceptualizes how to implement reform as a process of collective learning-by-doing that must be open-ended in order to get people to make these behavior changes. And it casts the task of reform leadership in a different light as a matter of enabling and supporting decentralized change at numerous points throughout the bureaucratic system rather than trying to legislate it or force it from above. This approach of *reform as process* aims to minimize the mechanisms of failure and maximize the mechanisms of success that have characterized past reform efforts, and to serve as a viable alternative for the many reformers and researchers who are seeking a different—and hopefully better—way to make changes in public bureaucracies.

Of course, there are many practical and political challenges to approaching reform as a process. The book discusses these obstacles and how to navigate them, based in large part on evidence from reform leaders who have experimented with these ideas and whose thoughts and efforts have helped inspire this book. Even so, it is definitely not easy, perhaps not even always achievable. But in most cases it represents a more promising approach—grounded in evidence, theory, and illustrative examples—to a challenge that has confounded generations of reformers and researchers alike and that represents one of the main barriers to improving performance in government organizations around Africa and worldwide.

The question I posed earlier—how you should advise a senior leader to go about designing and implementing reforms—isn't just a hypothetical scenario. In 2014, I sat in the office of Ghana's newly appointed Head of Civil Service, Nana Agyekum-Dwamena, who said: "I've got five years, and I want to change the way this place works. But I want to do it in an evidence-based way." What, he asked, did academics have to say about how to approach this? Agyekum-Dwamena brought nearly thirty years of experience in Ghana's Civil Service to tackling this challenge, but as a scholar, I felt uncomfortably limited in my ability to contribute either a broad and rigorous view of the evidence or a straightforward set of theoretical insights on designing and implementing system-level bureaucratic reforms. Given the complexity and context-dependence of such reforms, I wasn't even sure what an answer to this question would look like.

This book represents my effort to grapple with this challenge—to be able to offer stronger evidence and more useful conceptual frameworks to the numerous smart, dedicated, and experienced civil servants searching for insight to help them improve the effectiveness of their institutions. Doing so throws up an interesting set of puzzles and opportunities for researchers. This process has involved collecting and analyzing a great deal of new data, as well as learning from and synthesizing the immense amount of research already conducted on these issues by academics and practitioners from Africa and around the world.¹ I have tried to do this in a way that amplifies and complements this existing work on reform design and implementation and connects it to the rapidly growing adjacent literatures on bureaucratic performance, state capacity, organizational change, long-term development, and complex systems. While I think the book makes some novel theoretical and empirical contributions, it also reports on and builds on a great deal of insight from other researchers, and I have done my best to highlight these intellectual lineages and debts throughout the book.

I have also sought to draw as much as possible on the experience of practitioners themselves, not just as data points but also as deep theoretical insights. To the extent this book succeeds at its goals, it is largely because the generosity of numerous civil servants has enabled me to benefit from their ideas and bring their experiences into conversation with one another across countries and across time. Many of them are also frustrated with the shortcomings of past reform efforts and have long been actively thinking about and experimenting with alternatives. Their perspectives on these issues are naturally diverse and often contradict one another, and of course, the responsibility for the book's analysis and synthesis of this vast range of ideas and evidence rests with me. Nevertheless, I hope that they—and other readers who have tried to make change in government

bureaucracies or large organizations anywhere—see some of their experiences reflected in what I have written, together with a deep respect for their work.

Most of all, I hope that this book can help revitalize and recenter the study of large-scale bureaucratic reforms in social science research and policy thinking. Recent years have seen a general shift away from asking “big questions” toward focusing on more narrowly formulated research questions and interventions that are easier to implement and evaluate. But academics and practitioners alike need ways to bring evidence and theory to bear on complex challenges and questions that require systemic reforms, even if these are rarely amenable to clean, closed-ended, provably correct solutions. To paraphrase the political scientist Henry Farrell, studies of such questions are impossibly big and likely to be wrong (or, at least, not fully right) but nonetheless help us understand something important a little bit better.²

This book is written in that spirit. It asks a question of first-order importance for both research and practice: How should bureaucratic leaders design and implement systemic reforms? It attempts to answer this question by grounding itself in solid conceptual foundations, systematically collecting a vast amount of data, analyzing it carefully, and using these empirical findings to build a better understanding of why systemic reforms often fail and how they might be approached more successfully. But the answers it generates are not simple, closed-ended, provably correct solutions because that is not the nature of the research question—or of the task facing reformers. So this book is not intended to be the final word on studying and implementing bureaucratic reforms. Rather, its more modest goals are to help advance the state of empirical evidence and substantive insight on civil service reforms, to contribute some new theoretical ideas and methodological tools, and to encourage more researchers and reformers to grapple with the big, systemic challenges—and opportunities—facing government bureaucracies around the world.

THE CHALLENGES OF RESEARCHING SYSTEMIC REFORMS

That effective government bureaucracies are essential to prosperous economies, fair and open politics, and healthy societies is one of the few issues on which there is near-unanimity among academics and practitioners from across the ideological spectrum. This point has been demonstrated by hundreds if not thousands of academic studies—not to mention the lived experience of anyone who’s ever worked

in or depended on a government bureaucracy. If you've picked up this book and made it this far, I take it that you don't need to be convinced that effective bureaucracies that implement policies well and deliver services efficiently are generally a good thing and that taking steps to improve their performance is important.

When we think about the kind of evidence we'd want to help us do this, both academics and practitioners tend to ask questions like, "What is the effect of this intervention or reform on performance?" or, "What interventions or reforms should I adopt to improve performance?" There are many versions of this question, but they are all variations on the same underlying question of what lever needs to be pulled to improve performance. For example, reformers or researchers might investigate the effects of adopting a new performance management policy that aims to link measured performance to some type of rewards or sanctions. (Indeed, this book will show that such policies are among the most common reforms governments adopt to try to improve performance.) Or they might seek to understand whether certain structural variables outside the immediate control of the country's civil service—income levels, colonial or administrative legacies, political competition, and so on—help determine the effectiveness of the civil service.

These are questions of causal inference, and modern social science is increasingly dominated by methods that try to answer them in precise and rigorous ways, from quantitative randomized control trials or natural experiments to qualitative process tracing. There's a great deal of excellent research on bureaucracies and service delivery in this vein, and it has helped shed more light not just on the effects of incentive schemes on various dimensions of bureaucratic performance but also on selection, monitoring, motivation, funding, management, and politics.³ I've also contributed to some of this research alongside various coauthors, including through a long-term collaboration with Ghana's Civil Service that has included surveys, qualitative studies, and a randomized control trial—all of which also stemmed from the 2014 conversation with Agyekum-Dwamena that I referred to above.⁴

Despite the importance of these approaches, I want to argue that if we want to adopt an evidence-based approach to learning about reforming civil services at the *system* level, we need to be able to temporarily set aside our causal inference mindsets—or at least augment them with other forms of evidence and insight. There are two main reasons for this.

First, the systemic, multifaceted nature of large-scale reforms and bureaucratic performance makes it incredibly difficult to apply causal inference methods at the level of rigor that researchers and evaluators have come to expect. While causal inference approaches vary greatly in their methods, they share a set of

minimum requirements: interventions or policies that are narrow and precisely specified enough to be coded as binary or categorical treatment variable(s); exogenous variation in the application of the reform across units of analysis in a way that creates a clean counterfactual or “control” group; and good measures of bureaucratic performance across all (or at least most) dimensions that might be affected by this policy. But system-level civil service reforms nearly always bundle together numerous policies and processes that include both “hard” and “soft” elements and apply them to the whole of the system (i.e., to all units of analysis) simultaneously. Additionally, bureaucratic performance is notoriously difficult to measure in a consistent and comprehensive way, especially across ministries and agencies with very different functions and with multiple competing goals and tasks that have different levels of measurability and often require collaboration across various teams and stakeholders.⁵ And that’s without going into questions of whether the reform cases being studied are representative of the contexts to which we want to apply the evidence. It’s not impossible to meet these requirements for applying causal inference methods to large-scale reforms—and researchers should try to do so where possible—but it’s definitely easier to do so for microlevel interventions, small-scale pilots, and isolated behavioral nudges. Thus, the emphasis on obtaining causally identified answers to narrow, closed-ended questions has come to dominate research on changing bureaucracies.

Second, when we ask how to make bureaucracies more effective, every interesting answer is endogenous to the system we are trying to study. To be more precise, the prevailing emphasis on identifying exogenous variation in the explanatory variable of interest pushes researchers to only study variables over which current policymakers have no control—this is the definition of exogenous—and thus inadvertently avoid studying the very processes and situations that reformers might actually be able to change. An extreme example of this is studies that examine how macrohistorical or geographic factors have influenced present-day governance quality.⁶ It might be true that these deep and unchangeable variables explain much of the cross-country variation in various measures of government performance. But it’s often not a very useful insight for a present-day head of civil service, except, perhaps, in a very abstract sense. Even studies in which researchers collaborate with governments to run field experiments that test a specific policy in a real-world setting are not immune from distortion. While such trials or pilots can generate evidence that is valid and useful in that context, the same features of the setting that often make a rigorous field experiment possible—leaders’ interest in generating and using evidence, involvement of a skilled researcher, performance data availability, careful experimental control, an intervention that is self-contained enough to be delivered to some units but not others—also mark

out most such settings as atypical of bureaucratic contexts more widely. Of course, both macrohistorical research and randomized control trials (and many causal inference questions and methods in between) can provide important pieces of the puzzle of understanding bureaucratic performance and improvement. My point is not that we should do fewer of them but rather that limiting our research to factors or reforms for which we can obtain exogenous variation risks also limiting the types of reforms we study and the contexts in which we study them. If we want to generate empirical evidence and theoretical insight on complex, systemic bureaucratic reforms, we need to be able to ask and answer research questions about reforms that have been endogenously adopted and driven.

THIS BOOK'S APPROACH

Everything about this book has been profoundly shaped by these intellectual challenges. How can research on systemic reforms be as empirically grounded and as systematic as possible once we let go of the prescriptive strictures of causal inference—especially when the focus of our investigations is on these inherently endogenous reforms? What does generalizable knowledge look like when the adoption, implementation, and effects of reforms are subject to high-dimensional complementarities and contextual contingencies? How can we avoid the trap of throwing our hands up in the air and saying that it's all very complex, it all depends, and that no, we couldn't possibly use simple, straightforward language to offer a few actionable insights? And what would it look like to build a theory on implementing systemic reforms that is parsimonious but not oversimplified, profound but still actionable?

First, the guiding *question* of this book—How should bureaucratic leaders approach the task of systemic reform?—is a deliberately unusual one for an academic manuscript. A more orthodox approach would be to formulate a closed-ended research question (“Does *X* cause *Y*?”) that can be answered by deductive testing of a theory-driven, falsifiable hypothesis about the impact of one (exogenous) variable on another variable. Alternatively, when researchers do ask open-ended questions, they often ask variations of the question, “What factors cause *Y*?”⁷ My guiding question contains elements of both approaches while also insisting on seeking answers that are both actionable, practical, and generalizable enough to communicate to nonacademics—all in a setting where I've argued that conclusively establishing causal relationships between reform adoption and bureaucratic performance is nearly impossible. This research question is,

therefore, an especially challenging one to tackle, but it is (in my view) the most important question, so I maintain it as my guiding star and aim to find answers that are both grounded in evidence and useful for decisionmakers—incomplete and imperfect as these answers may be.

Second, the nature of an *answer* to such a question must consist not of a definitive one-size-fits-all solution but rather an insight into the theoretical *mechanisms* through which reforms succeed or fail that can be applied across a range of settings. Researchers and practitioners alike often desire unambiguous, causal answers: reform *X* causes outcome *Y*; adopt this reform, and it will improve performance. I've argued that for both practical and conceptual reasons, such evidence is near-impossible to generate for systemic civil service reforms. Even if it could be generated, different governments are facing different challenges at different moments with different existing structures and different reform histories; it should be obvious that there can't be a one-size-fits-all answer to this question that holds across all contexts.

Instead, useful and generalizable knowledge about systemic reforms has to consist of a set of midrange theoretical insights about different parts of the reform process, mental models that reformers can absorb, remix, and apply to the unique demands, constraints, and opportunities of their own particular contexts. In other words, learning from evidence about reform isn't like following a blueprint that tells you exactly what to do. Instead, it's like compiling a toolbox and learning what each tool does and doesn't do so that you can use these tools to solve your own challenges in your own context. In this book, then, I use empirical evidence to identify and explore what I refer to as *mechanisms of success* and *mechanisms of failure*, patterns that repeat themselves across contexts and that often coexist within a given reform effort. Learning about these mechanisms is analogous to learning how the tools in the toolbox work and is the kind of answer that this book aims to provide.

Third, the *scope* of the empirical analysis on which my arguments are based is broad rather than deep. Instead of trying to deeply understand the effects of a small number of cases of reform in great depth or conduct a deep history of reform in any one country, I study every identifiable effort at system-level, performance-oriented civil service reform in six different African countries over the last thirty years. This includes everything from high-profile reform efforts that have already been well-researched to little-known reforms that were announced but never got off the ground and quickly faded into obscurity. (I'll give more details about definitions and scope in a moment.) This approach gives me a relatively large sample of reforms to study and eases concerns about inadvertently picking unrepresentative cases based on existing perceptions of success or failure or on which cases have the

most information available about them. Even more importantly, studying many reform efforts across multiple countries makes it possible to start to distinguish broader trends and patterns that occur across many cases, whereas focusing on a single reform or single country often lends itself to identifying a wide and often overdetermined range of idiosyncratic failures that seemed to cause success or failure in that particular case. In my conversations, I have often heard the failure of a reform attributed to factors like a lack of political will, poor implementation, or a country's perceived work culture. And so reformers often think that a particular reform might have succeeded if only they had tried a bit harder or gotten a bit more lucky, or, alternatively, that their country is just uniquely unsuited to reform. But when we zoom out, we often see *repeated patterns* of similar efforts leading to similar results time and time again, often in quite different contexts. These repeated patterns can then be linked to the type of theoretical mechanisms of success and failure that provide generalizable insight to reformers. Studying a broad scope of reform episodes across several countries thus helps us move from idiosyncratic explanations to generalizable patterns.

The six countries I study are Ghana, Kenya, Nigeria, South Africa, Senegal, and Zambia. My temporal scope is roughly the three decades leading up to 2019, during which improving performance became a central and explicit goal for bureaucracies worldwide, but my starting year for each country varies between 1986 (Senegal) and 1999 (Nigeria), depending on the timing of key political transitions or reform waves. The countries are all democracies (at least, during the periods I study) and have undergone numerous reform efforts, but they span across western, eastern, and southern Africa and are diverse in terms of their size, wealth, and historical legacies. All inherited administrative structures and traditions from their colonial past but from different countries and with different characters, and these had each evolved and changed in different ways prior to the time period I study. While not randomly selected or fully representative of the continent or the world, as a sample, they strike a balance between comparability and contextual variation. It is important to emphasize that the aim of the book is not to make cross-country inferences about how exogenous characteristics affect reform experiences: the fundamental unit of analysis for the book is the reform effort, not the country, and reform cases differ as much within as across countries. By design, countries that have not been democracies for the majority of the past three decades are out of the scope of the book, as the dynamics of civil service reform are considerably different. This book, therefore, aims to speak mainly to reformers and scholars in countries with democratic, pluralistic political systems. While I do not gather systematic empirical data on reforms in any countries outside Africa, I do frequently draw on examples and evidence from all over the

world—including high-income countries—to motivate my analysis and consider the extent to which the patterns and mechanisms I describe are generalizable.

Within each country, I study system-level, performance-oriented reforms to the core civil service. By core civil service, I mean the central government ministries, departments, and agencies responsible for developing, implementing, and overseeing policy. I focus mostly on bureaucrats, administrators, and technical experts in offices in the capital or other large cities rather than on “frontline” employees like nurses, teachers, or local government officials dispersed around the country in public-facing service delivery roles.⁸ By reforms, I mean strategic and intentional structural or managerial changes to the internal administration of civil service organizations, whether *de jure* or *de facto*, aimed at improving bureaucratic performance. By performance, I mean the extent to which a government bureaucracy effectively delivers on its goals and objectives at the level of individuals, organizations, and the civil service as a whole.⁹ My focus is not on whether civil services are adopting optimal policies but on how well they are delivering on the policies they have adopted. And by systemic, I mean reforms that apply not just to a single agency or sector but to the whole civil service.¹⁰

My scope thus includes a wide range of reforms, from staff salaries and career structures to individual-level incentives and performance management, organizational performance and management, processes for monitoring and improving service delivery, and more. At the same time, it excludes reforms that take place within a single organization or sector, reforms that are exclusively customer-facing or that affect frontline workers only, reforms that are exclusively oriented around budgeting and fiscal issues, decentralization or other local government reforms, and anticorruption interventions. Of course, these kinds of reforms all potentially affect performance, as do many other factors besides reforms, such as politics, leadership, economic conditions, and so on. These are all important and worth studying—and have been researched extensively—but fall outside the scope of this book.

I collect data on each reform effort from a range of sources, both primary and secondary. I try to document the *content* of what changes the reform aimed to make; the *process* of how it was designed and implemented; the actual *implementation* of the changes it envisioned; the *outcomes* of the reform in terms of changing the everyday behavior of civil servants and improving overall bureaucratic performance (to the extent these are possible to gauge, given the challenges of measurement and causal attribution); and the *politics* surrounding the reform effort, both the “high politics” of parties and elections as well as the “low politics” of workplace relationships and vested interests. To do this, I draw on nearly one thousand systematically collected academic studies and government documents,

fifty-one interviews with elite civil servants who were personally involved in designing and implementing these reforms across all six countries, thirty-three interviews with rank-and-file civil servants who were affected by these reforms (in Ghana and Zambia), and in Ghana, archival records from the Office of the Head of Civil Service and Public Records and Archives Department. Table 1.1 summarizes these data sources. The amount and reliability of data I am able to gather, of course, varies dramatically across different reform efforts, and my own data collection and analysis are subject to many of the challenges of measurement and inference I highlighted above. I therefore calibrate the strength of my claims depending on the strength of the underlying evidence, particularly with respect to judgments about the impacts of reforms on bureaucratic behavior and outcomes. But by synthesizing and triangulating across these sources, it is possible to present a rich and fairly comprehensive picture of the design and implementation of dozens of reform efforts across thirty years and six countries.

Applying these criteria, I identified a total of 131 reform efforts across the six countries during this time period. When I began research for this book, I had originally planned to code the characteristics of these reforms to undertake quantitative or configurational analysis of how reform content and process were associated with implementation and outcomes, possibly moderated by political context. However, the process of compiling data and beginning to analyze it convinced me that such an analysis would be oversimplified and imply false precision at best, and be misleading and biased at worst. To some extent, this was due to

TABLE 1.1 Countries and data sources

	Ghana	Zambia	Kenya	Nigeria	Senegal	South Africa
Government, donor, and media documents	✓	✓	✓	✓	✓	✓
Existing academic literature	✓	✓	✓	✓	✓	✓
Interviews with elite reformers	✓	✓	✓	✓	✓	✓
Interviews with rank-and-file civil servants	✓	✓				
Government archives	✓					

Source: Author.

challenges of measurement—limited data availability, contradictory accounts, and the inherent ambiguity of many bureaucratic and political processes meant that it was often challenging to capture basic features like the end date of each reform with a single number, let alone its implementation process or outcomes. Even more fundamentally, reform episodes couldn't be analyzed as independent cases because the boundaries between reform cases were blurred: each reform built (in part) on what had come before and linked (in part) to other reforms that were ongoing simultaneously. Sometimes these influences reflected positive path dependence and efforts to maintain continuity and complementarity across reforms. Other times, the perceived failings of a preceding reform led its successors to try to do the opposite, or simultaneous reforms worked at cross-purposes because they were led by rival institutions. This means that even something as simple as counting reforms is difficult—the 131 figure I cited above is indicative, but different approaches to drawing case boundaries could lead to the number being far higher or lower. Trying to analyze these reform efforts as independent cases would be invalid, even without attributing causality to the observed relationships.

Rather than treating these methodological challenges as “bugs” to be worked around, I decided to treat them as inherent “features” of the phenomenon and processes of systemic civil service reform. So instead of relying on assigning codes to these episodes as discrete cases, I used my data to compile a narrative history of reform in each of the six countries, which presents a rich description of the reforms, their contexts, and their interconnections. These narrative histories serve to collate and harmonize messy data sources and diverse reforms into a format that is fairly consistent across reforms and countries while still permitting enough flexibility of exposition to capture complexities and nuances that formal coding would obscure. They then serve as the basis for producing descriptive generalizations about what governments are trying to do and how they are trying to do it, as well as abductive analysis of mechanisms of failure and success.

OUTLINE AND MAIN ARGUMENT

This book is, in some ways, two books in one. The first book is a narrative history of systemic, performance-oriented civil service reforms in each of the six countries. The second book is a thematically organized analysis of the track record of these reforms and what we learn from them, which draws on the narrative histories as data. However, the first book, containing the country-by-country narrative histories as well as further details of data and methods, is located at the end

in an extensive appendix. The thematic analysis thus comprises the main text of the book, and the contours of my arguments closely follow its structure.

Chapter 2 lays out a theoretical framework and empirical evidence for conceptualizing performance. Consistent with my abductive approach, this chapter focuses on building strong conceptual foundations as a starting point for empirical analysis rather than on developing hypotheses to be tested. The chapter begins by introducing the book's key theoretical building block: the idea that performance in organizations requires some actions that are *verifiable* (formal) and others that are *nonverifiable* (informal).¹¹ Formal actions are those that can be precisely specified *ex ante* and measured *ex post*; they are the kind of actions that can be easily described in a manual of standard operating procedures or measured on an annual performance appraisal. Informal actions, however, encompass all the other crucial actions that workers must undertake to carry out their jobs effectively but are too unpredictable, too complex, or too unmeasurable to be fully codified. I illustrate these theoretical distinctions with empirical evidence that these unformalizable aspects of performance are pervasive in public sector organizations, whether one looks at the level of organizational outputs, individual behaviors, or management processes. Thinking of public sector performance in this way helps us gain clarity about the core challenge of systemic bureaucratic reform: How can a reform leader (like a head of civil service) design and implement reforms to strengthen the performance of both formal and informal actions by thousands of bureaucrats across different organizations working on different tasks, in different teams, under different managers?

Part II of the book (chapters 3–6) is the empirical core, examining the track record of these six countries' reform efforts over the past three decades. It documents how governments have approached reform, how successful they have been, and what have been the main mechanisms of failure and success of these reforms.

Chapter 3 presents the key descriptive themes that characterize the history of civil service reforms across the six study countries. Chapter 3 begins by presenting a single reform case (Zambia's Public Service Capacity Building Program, 2000–2005) to give a tangible example of the kind of reform episodes this book studies—although these are very diverse and there is no single “typical” reform. The chapter then builds on this illustrative case by using the full range of data collected on reform histories to answer a basic set of descriptive questions:

1. How frequent have reform efforts been?
2. What has been the content of these reforms—the structures, processes, and practices they have tried to introduce?
3. How successful have reforms been?

In brief, my answers to these questions are:

1. Reform efforts were very frequent and often overlapping within each country. On average, in each country, a new reform effort was launched every 1.3 years, and there were 3.8 reforms simultaneously active.
2. Reform episodes mostly tried to achieve similar outcomes in terms of improved performance and did so mainly by repackaging and recombining a few different types of reforms (such as individual-level performance management, salary and career reform, and organizational capacity-focused interventions) into new bundles, meaning that countries often tried to implement the same types of reforms over and over again.
3. There were no examples of reforms that fully met all their goals, and most reforms fell far short of their (often overambitious) goals. But few were total failures. Most did achieve some of their aims—although more in terms of making changes on paper than in practice and not always through the mechanisms intended by the reform's designers.

Finally, the chapter examines what factors drove the design and timing of reform efforts, considering three sets of potential factors: the incentives and processes of key stakeholder agencies, in particular, international donor organizations and various agencies within governments; political cycles and time horizons; and the persistence, diffusion, and recurrence of ideas across and within countries. Much existing literature focuses on the roles of donors and politics, suggesting that reforms fail because they are externally imposed and not “owned” because governments are only pretending to reform in order to satisfy external donors and lenders or because African governments are forced to adopt and imitate models from the Global North rather than developing homegrown solutions.¹² I find some evidence consistent with each of these explanations, but also many cases in which reform implementation fell short despite a genuine desire within governments to improve performance and genuine belief among reformers that the proposed solutions were appropriate. Donors and political economy are thus both important parts of the story and often initiated or intruded on reform efforts, but they weren't the whole story, and the reformers I interviewed frequently pointed out ways that they themselves had major influence over reforms that, from the outside, appeared to be driven solely by external donors or political imperatives. Instead, reforms were usually shaped to a significant degree by reformers' own understandings or performance and reform, which were, in turn, shaped in path-dependent ways by their own idiosyncratic experiences, their institutional socialization, and by international trends and thinking. In

sum, mental models, ideas, and inspiration mattered. This offers some hope for researchers (like me) who hope that evidence and insight on such reforms might not just be of academic interest but also practical use.

Chapter 4 documents the first of the two main mechanisms of reform failure. It shows that most reforms tended to focus heavily on creating or changing *formal structures and processes* to try to force bureaucrats to perform better by using rules, carrots, and sticks. To illustrate this, the chapter focuses on one of the most common types of reforms introduced across countries: individual-level performance management policies, which aim to systematically link individual bureaucrats' performance to rewards and sanctions. There were thirty-four total efforts to introduce individual-level performance-linked incentives in these six countries over three decades. But despite careful design and widespread acceptance of the aims of these reforms, there were zero instances in which differentiated rewards and sanctions were actually delivered sustainably and systematically to civil servants. Only two delivered differentiated incentives at all—both of which ceased doing so within a few years—but neither delivered sanctions for poor performance. The neglect of the importance of unformalizable bureaucratic behaviors undermined both the implementation and impact of these policies. This pattern also emerged with respect to other common types of reforms, such as performance improvement funds, salary structure changes, and organizational performance reviews. These also tended to focus primarily on formal behaviors and formal rules and processes for changing them, and this pattern helped explain why they so consistently fell short of their goals.

Chapter 5 then examines the second main mechanism of reform failure: reforms were usually conceived of as discrete, one-off, often time-bound interventions. The chapter shows how this *projectization* of reform distorted the expectations, content, implementation, leadership, and politics of reform efforts in ways that undermined their ultimate goal of improving performance. Projectization thus explains a wide range of observed features and patterns of reform, from the overselling of reforms' potential benefits to its tendency toward top-down leadership styles, its focus on formalizable outputs and targets, and its frequent lack of political sustainability. While often politically and bureaucratically convenient, the projectization of reform was thus a second common mechanism through which well-intentioned reforms failed to live up to the expectations set for them.

Chapter 6 turns to investigating reforms' two main mechanisms of success: creating opportunities to discuss performance and how to improve it and creating energy and momentum for change. Opportunities for discussion gave civil servants who cared about performance the chance to find one another,

empowered them to identify and enact potential improvements, and spread the message that performance mattered. For example, annual staff appraisal systems never succeeded in systematically changing behavior through carrots and sticks but sometimes spurred workers to have conversations with their bosses for the first time about what they were responsible for and how they were doing. Energy and momentum helped to shift civil servants' expectations of one another from a negative equilibrium of low effort and low performance to a more positive pattern of extra effort and good performance being recognized and reciprocated. These mechanisms of success were driven largely by nonverifiable behavior changes: meaningful discussions can't be forced, workers can't be directly incentivized to care, and it's hard to drive innovation or problem-solving through predefined targets. While the means through which these mechanisms operated often involved formal structures and processes to some extent, they could also be easily crowded out by governments' efforts to use these formal systems for accountability and incentivization purposes. But when reforms were able to induce widespread and meaningful improvements in performance, one or both of these mechanisms usually played a leading role.

The four chapters of part II thus paint a rich portrait of the big-picture patterns of reforms and of the repeated mechanisms of failure and success that emerge when we analyze them collectively. While this book's main empirical analysis focuses on six countries in Africa, each of the chapters in part II also presents suggestive evidence that these patterns and mechanisms are also present in many other countries and other regions, including in high-income countries. Of course, these mechanisms manifest to different degrees, in different ways, and in different combinations across different reform efforts. But while each reform effort and context is unique in some ways, recognizing broader patterns can help us step back from these idiosyncrasies and realize that there is both a need and scope to rethink how we approach systemic reform more generally.

So if part II of the book focused on the "diagnosis" of the ways that reforms have failed and succeeded, part III (chapters 7–9) revolves around the "prescription." How can we conceptualize what an alternative approach to reform might look like and assess both its potential and the challenges it would face?

Chapter 7 draws together the empirical analysis of the previous chapters with the theoretical framework on organizational performance from chapter 2 to develop a theory of reform as *catalyzing an ongoing process of continuous improvement in actual practices*. This theory of reform as process reframes the goal of reform, focusing attention directly on changing the day-to-day work practices of rank-and-file civil servants and their managers rather than implicitly assuming that changes in behavior come from changes in formal rules. It reframes how

reforms should be implemented, conceiving of change as an ongoing process of many locally driven changes rather than a master plan rolled out through a one-off project or intervention. And it casts the role of senior leaders in a different light, seeing their task not as driving or forcing reforms from the top down but rather as catalyzing, enabling, and inspiring decentralized local change efforts by thousands of staff and teams spread throughout the whole civil service.

Chapter 8 then examines what reform as a process can look like in practice by conducting a focused case study of reforms led by Nana Agyekum-Dwamena during his time as Head of Ghana's Civil Service. His decades of experience with projectized reforms had convinced him that when he took charge, he wanted to focus not on introducing new policies or new structures but on improving the implementation of existing ones; on changing the management practices that guide ordinary bureaucrats' behavior and routines on a day-to-day basis; and on transforming the mindsets and culture of civil servants across the various ministries and departments that comprise the Civil Service. This chapter documents these reform efforts as well as their successes and limitations. I was privileged to be conducting research in collaboration with Ghana's Civil Service throughout this time period, alongside several coauthors, at Agyekum-Dwamena's request, so I witnessed many of these discussions and reforms as they unfolded. Not all of Ghana's problems were solved, and not all efforts at change were successful. But many were, and overall, the evidence paints a picture of a Civil Service that has made gradual yet meaningful improvements in a wide range of areas—all while diverging, in many respects, from the formula that guided previous reform efforts in Ghana and elsewhere. The chapter closes by discussing how Agyekum-Dwamena navigated the political and institutional challenges that arose in approaching reform as a process.

Chapter 8 also distills this alternative approach into a set of three simple rules of thumb that reformers everywhere can apply to their own contexts:

1. Focus first on what can be done within existing formal rules and processes; think of changing formal rules and processes as a last resort rather than a first step.
2. Approach change as a process of collective learning-by-doing rather than as rolling out a predesigned blueprint. The priority should not be to make the perfect plan but instead to start changing actual practices—even small or apparently minor ones—as early as possible.
3. Decentralize the leadership of reform as much as possible. The role of a leader is to encourage and support dispersed improvement by other actors across the system rather than to drive it by themselves.

Of course, exactly what actions these rules of thumb imply will be different in different times and places and for actors in different roles. They are general principles rather than rigid guidelines: Changes to formal rules are sometimes important, some types of changes ought to be rolled out as discrete projects, and top-down leadership is occasionally necessary. But both theory and the evidence of how and why reforms have persistently fallen short of expectations suggest that reform leaders—whether operating at the level of whole systems, single organizations, or even just their own teams—would generally be better served by using these rules of thumb to guide their thoughts and actions.

Chapter 9 concludes the book by exploring the nuances and scope of these findings and recommendations. To do so, it discusses how the dynamics of reform might differ along four dimensions: the *purpose* of the reform, in terms of whether it is aiming to change mostly verifiable or mostly nonverifiable practices; the *context* of the reform, in particular the degree to which compliance with formal rules is enforced and expected; who the *people* driving the reform are, both in terms of their seniority and position and whether they are internal or external to the bureaucracy; and the *politics* surrounding the reform, especially the time horizon of the reformer(s) and the degree to which external actors impose pressures or constraints on the bureaucracy. The chapter suggests that these factors determine the extent to which this book's analysis and advice is likely to hold for other countries, other change efforts, and other types of organizations. While there are reasons to think that this book's diagnoses and prescriptions are widely applicable, the determination of how relevant they are to any particular situation ultimately rests with you, the reader, and your knowledge of your own context and goals.

You've likely noticed by now that the book does not have dedicated chapters on two issues that are often discussed in the context of civil service reform in Africa and elsewhere: the role of international donor institutions and the politics of reform. This is because I made a deliberate decision to focus not on the external forces that constrain or influence reform leaders within country governments but on reformers' agency and room for maneuver within these constraints. To be sure, donors are important stakeholders and funders in many (but far from all) of the reforms I study, so I do discuss their roles and influence extensively—but always as members of the supporting cast, not as the leading actors. With respect to politics, I treat it the same way that public servants typically experience it: neither as an afterthought to technical discussions nor as entirely deterministic of reform opportunities but rather as inextricably bound up with the technical and organizational aspects of reform. Each of the core chapters, therefore, integrates analysis of how political considerations have shaped the past track record

of reforms and must also be navigated by any alternative approach while still foregrounding the agency of reformers in doing so.¹³

Finally, it's important to emphasize that this book has many allies, cousins, and forebears in its effort to rethink how we conceptualize behavior, performance, and change in organizations in general and public bureaucracies in particular. In the realm of academic research, scholars from diverse fields and perspectives have written about bottom-up approaches to reform, relational contracts theories of management practices, backward mapping in policy implementation, continuous improvement, process-oriented methods for studying organizational change, problem-solving and muddling as change strategies, the importance of organizational cultures and bureaucratic autonomy, transformational leadership and system leadership, the building of pockets of bureaucratic effectiveness, polycentric governance, agile government, and problem-driven iterative adaptation and other complexity science-inspired ways of thinking—to name just a few of the areas of existing literature that this book connects with, draws inspiration from, and seeks to build on.¹⁴ I'll discuss these theories as they arise over the course of the book, especially in chapter 7. And of course, reformers around the world have been grappling with these ideas for far longer than academics have been writing about them.

So in analyzing how past reform efforts have fallen short and what a better approach might look like, I don't mean to imply that no one else has ever recognized any of these challenges before or tried to experiment with solutions. Still less do I want to suggest that overcoming these challenges is easy. On the contrary, systemic bureaucratic reform is complex and difficult, both conceptually and practically, and most past reforms were themselves the product of smart, dedicated people trying their best to piece together something that they thought would work. I don't pretend that this book has all the answers, but I hope that reading it helps you feel at least a little better prepared than I was the next time a head of civil service—or a minister, mayor, CEO, manager, or other leader—calls you into their office and says, “I want to change the way this place works, how would you recommend I go about it?”

2

Theory and Evidence on Organizational Performance

What do civil service bureaucracies do? And what does it mean to do it well?

The last chapter argued for the importance of studying large-scale, systemic reforms by asking whether and how they actually affect the everyday behaviors of rank-and-file individuals and their managers. This chapter builds answers to these two questions from the bottom up, focusing on understanding performance and behavior at the levels of individuals, teams, and organizations. In so doing, it weaves together a conceptual framework that serves as a foundation for the book's subsequent investigation of how system-level reforms do or do not translate into improved performance at the micro- and meso-levels.

The first section of the chapter focuses on defining bureaucratic performance and building a simple but powerful theoretical framework for thinking about it. I approach this not by asking what we want civil services to look like in terms of macrolevel characteristics or structures but by asking what performance looks like in terms of everyday individual behaviors and organizational processes. Some of these behaviors can be formalized with tools like key performance indicators and standard operating procedures, but other important behaviors are too difficult to specify or measure to be effectively formalized. In the language of organizational economics, the former are verifiable (formal) behaviors, and the latter are non-verifiable (informal). The core challenge of organizational performance—and of performance-oriented civil service reforms—is thus how to get employees to undertake both types of behaviors.

After setting out this conceptual framework, I then present three empirical “stylized facts”—midrange empirical patterns that broadly hold true across a range of contexts—about performance in bureaucracies at the organizational

and individual levels. First, most of the tasks that government bureaucracies are responsible for are not fully verifiable in that they either can't be perfectly specified in advance or perfectly measured after the fact. Second, most organizational management processes are also not fully verifiable, meaning that they require individuals to undertake informal as well as formal actions. Third, there are large variations in organizational performance even among organizations in the same context operating under the same set of formal rules and external constraints—due, in part, to the challenges of getting people in the organization to undertake both the formal and informal parts of these processes. Taken together, this evidence paints a picture of public service organizations in which the nonverifiable, unformalizable aspects of performance are pervasive.

The role of this chapter is to provide a simple, evidence-grounded, micro- and meso-level foundation from which to undertake the abductive analysis of system-level reforms that is the empirical core of the book. Rather than setting out hypotheses to be tested against subsequent empirical evidence, it provides a starting point for analysis and a framework for interpreting it. So while this chapter doesn't posit any answers to this book's motivating question—how a head of civil service should approach performance-oriented reform—it does give us a structured way to think about what bureaucratic change efforts must achieve to improve performance and why that is so challenging. With that in mind, let's begin by defining what we mean by performance.

DEFINING PERFORMANCE

Throughout this book, I use the term *performance* to refer to the extent to which an individual or organization delivers on its goals and objectives. In economic terms, this means maximizing internal efficiency subject to externally given goals, policies, and constraints. In administrative terms, it means implementing policies and processes as they are intended to be implemented. Used this way, performance can refer to a property of individual behavior or to an emergent collective property of the cumulation of linked individual behaviors within teams or organizations. It can also apply to the undertaking of specific actions or tasks, the execution of organizational processes, or the production of final outputs or outcomes by an organization (or even a set of organizations).

This definition is simple, intuitive, and flexible, but that inevitably comes at the price of abstracting away from a great deal of complexity. Agreeing on a precise definition of public sector performance is hard because government organizations

always have multiple goals that sometimes conflict with one another (e.g., service delivery, transparency, cost-effectiveness, impartiality, democratic accountability), and different observers disagree about how to weigh these in measuring performance. Most of these goals are heavily influenced by factors external to the bureaucracy—political, social, economic, environmental—and so the more comprehensively we attempt to conceptualize performance in terms of the outcomes citizens care about, the less control the bureaucracy itself has over their achievement. Measuring public sector performance is also hard because many (perhaps most) public sector outputs are nonpriced (i.e., they are not market goods), and their quality and/or quantity often cannot be characterized precisely and objectively. These conceptual and definitional challenges are the subject of dozens, if not hundreds, of books and articles on public sector performance.¹

While these nuances and definitional debates are important, I opt for a simple and limited definition of performance as the implementation of defined policies and processes for three reasons. First, the reforms I study in this book differed in their own definitions of performance, often without grappling with these complexities or specifying how they balance these different goals. So my definition must be flexible enough to capture what these varied reforms were aiming at. Second, I do not attempt to construct my own formal performance measures for civil services or reform success, relying instead on the diverse measures and perceptions of civil services themselves, interviewees, external evaluators, and academics. Third, the idea of the civil service as primarily charged with implementing politically defined goals and objectives corresponds closely to the legal role of the civil service as the executive arm of the government and, thus, to the main objective of a reformer like the head of civil service. So while I do engage with some of these complexities in later chapters and the appendix with respect to evaluating the success or failure of reforms and in discussing strategies for reform, I deal with these conceptual and practical difficulties as they arise rather than trying to impose a rigid conceptual schema on them in the abstract.

The task and institutional context of civil services also affects how we conceptualize bureaucratic performance. The majority of civil servants are not workers standing alone at fixed assembly lines performing repetitive tasks to create standardized widgets under the direction of a single boss. They work on many tasks at the same time, often collaboratively in teams or cross-organization networks. The tasks themselves are often not repetitive or well-defined, requiring bureaucrats to use their discretion or initiative to figure out what needs to be done. The quality or value of what has been produced is often hard to measure, and attributing good or bad outcomes to individual bureaucrats or organizations is difficult. And far from answering to a single boss, most bureaucrats must

simultaneously balance multiple managers and institutional imperatives, each of which is embedded in numerous layers of hierarchy. So the performance of individual bureaucrats is a complex thing to define and understand. Is there a simple way we can think about whether and how well bureaucrats do what they're supposed to do on a day-to-day basis that incorporates these nuances but still gives us a clear vision of what good performance looks like?

VERIFIABILITY

All bureaucratic operations ultimately consist of individuals (agents) taking actions at the behest of their manager or organization (principal) in pursuit of organizational goals. Some of these actions are written down in regulations or standard operating procedures, some are delegated by managers, and others rely on individuals themselves taking some initiative. Performance—in the sense I defined above—consists of getting individuals to reliably take the actions they're supposed to take in order to achieve organizational goals.²

The challenge with getting public servants to do what they're supposed to is that it's not always clear in advance what the right action to take in a given situation will be, nor is it always clear after the fact what an individual actually did. If everything bureaucrats needed to do could be specified in the standard operating procedures and everything bureaucrats actually did was perfectly observable by their managers, then getting them to perform would be easy: Just fire anyone who doesn't do what they are supposed to. But anyone who's worked in an organization knows that it's not that simple. Regulations and organizational handbooks might be a good starting point for defining what workers are supposed to do, but there are always lots of unforeseen situations or little details that can't be written down—not to mention situations when the written rules specify a course of action that conflicts with the organization's goals. And whether public servants are working in remote forests or in open-plan offices in ministry headquarters, most of the time, managers can't or don't watch their every action. And even if they were, they might struggle to measure exactly what it was that the official did, particularly in jobs where most work is cognitive and team-based rather than physical and individualized. So writing and following rules is only one part of what is required to achieve good performance.

Let's make this intuition a bit more precise, using the concept of *verifiability* from organizational economics.³ An action is verifiable if a third party (e.g., a court, tribunal, auditor) can objectively determine if the agent took the correct

action in a given situation on behalf of the principal. Actions are only verifiable if two things are both true: (1) the correct course of action to take in the situation can be specified *ex ante*, and (2) what the agent actually did can be measured *ex post*. Of course, in reality, these two properties can hold to a greater or lesser extent, so verifiability is a spectrum rather than a dichotomy. But for convenience, I will generally refer to actions as either verifiable or nonverifiable. Or, in less technical language, as formal or informal, formalizable or unformalizable.

This matters because if an action is verifiable, then organizations can compel it to be performed with formal rules or contracts associated with rewards or sanctions. Think of your own employment contract: It probably includes a specified list of duties that you can be fired for not performing—even if you work in a highly unionized or regulated sector, as is often true of public sector workers. But this list probably consists of a few bullet points or paragraphs that leave lots of ambiguity about what this looks like in practice, and there’s probably also some vague language around “and other such tasks as may be required” but with no clear way to define what might reasonably be asked of you. Your contract or job description might also say that the quality of your work should be “of a high standard” or something similar, but it’s impossible to precisely specify or measure that in all circumstances in an objective way. If you were sanctioned for one of these fuzzier parts of your job description, you would probably be able to appeal to an administrative tribunal or sue in an employment court, and your employer would have a hard time proving their case. So it is much easier for organizations to compel compliance with the highly verifiable parts of your job than the less verifiable parts.

But these hard-to-formalize actions are often crucial for performance at both the individual and organizational levels. This is particularly true in environments like the civil service, where the quality of outputs is hard to measure, individuals have to collaborate across stakeholder networks on complex and often unpredictable tasks, and individuals often have to innovate or use their discretion—their own judgment of a situation or their common sense about what is the right thing to do—to navigate day-to-day challenges. Doing only the formal, verifiable parts of a job is akin to following the letter of the law, whereas achieving the spirit of the law—the course of action that will lead to achieving organizational objectives—often requires supplementing these with informal actions that can’t be fully specified beforehand or perfectly measured after the fact. These informal aspects of work are so important that refusing to perform them by “working to rule”—when workers perform only the specific actions listed in their contracts or job descriptions but do none of the extra or unspecified tasks that actually make everything work—can even be used as a form of industrial protest because it is so disruptive to overall organizational functioning.

Let's again draw on the language of organizational economics to make this intuition more precise and distinguish between three stylized levels of performance. Doing exactly what a rule or contract says—nothing less, nothing more—is called *perfunctory performance*. Complying not only with formal rules but also supplementing them with informal practices that give meaning and fulfillment to them is called *consummate performance*. When actors neither comply with formal procedures nor undertake supporting informal practices, we can characterize this as a state of *nonperformance*. Getting to consummate performance is thus the core goal—and core challenge—of improving organizational performance. Table 2.1 summarizes these definitions.

The verifiable/nonverifiable distinction is similar in many respects to some other terms that are widely used in public administration and development. For example, James Q. Wilson famously categorized public agencies by how “visible” their outputs and outcomes are, and he wrote about how managing agencies with highly visible tasks was different from those with less visible ones.⁴ Similarly, Lant Pritchett and Michael Woolcock and Matt Andrews and colleagues have categorized public services according to, among other factors, whether they require agents to exercise discretion or are nondiscretionary.⁵ And terms like *hard* and *soft*, *tangible* and *intangible*, and so on are widely used by scholars and practitioners alike to capture similar distinctions. These concepts aren't all identical, and I think there are some analytical advantages to thinking in terms of verifiability. But these terms do aim to capture similar distinctions and are more alike than different.

The next section fleshes out these ideas with examples and evidence. Before proceeding to this, however, an important clarification: the term *informal* is sometimes used among public administrators as a euphemism for “corrupt,” “illegal,” or “improper,” but this is not at all the sense in which I use it in this book. I use *informal* simply as a synonym for nonverifiable: not fully specifiable by law but also not necessarily contravening any laws. Understood in this way, the vast majority

TABLE 2.1 Three stylized levels of performance

	Nonperformance	Perfunctory performance	Consummate performance
Formal practices defined and executed?	No	Yes	Yes
Supporting informal practices undertaken?	No	No	Yes

Source: Author.

of informal practices are completely legal, and many informal actions and processes are necessary for carrying out even the most basic organizational functions. Almost everything an individual civil servant or an organization as a whole does has both formal and informal components, albeit to varying degrees. These formal and informal elements are both important, both take effort and coordination, and undertaking both together is a crucial goal for organizational management.

MOST ORGANIZATIONAL TASKS ARE AT LEAST PARTIALLY NONVERIFIABLE

The idea that much of what bureaucracies and bureaucrats do is hard to specify and measure has a long intellectual tradition, from the observation that many public servants do work that can't be easily monitored to theories of differential observability of agency outputs and outcomes, the consequences of delivering nonpriced goods and services, and the openness to interpretation of organizational goals.⁶ While much of this literature has assumed that these characteristics vary at the level of organizations—some ministries do things that are easily measurable, others less so—another strand focuses on variation in measurability across different tasks or outputs within an organization. In economics, theories of multitasking explore how management tools like incentives can cause bureaucrats to focus their effort on more measurable tasks rather than less measurable ones,⁷ and public administration scholars have examined whether and how strict targets for one bureaucratic activity can distort public servants' efforts on that task or divert it from other ones.⁸ So there's a scholarly consensus that difficulty in clearly specifying and measuring bureaucratic action is a salient factor of public management—but does this refer to a small part of what government does or to the majority of it?

One effort to quantify this was undertaken by Imran Rasul, Daniel Rogger, and myself as part of a research collaboration with Ghana's Office of the Head of the Civil Service (OHCS) that stemmed from Agyekum-Dwamena's request for assistance in building an evidence base to help inform OHCS's reform efforts.⁹ In 2015, members of our research team worked with staff from OHCS and Ghana's Management Services Department to collect, digitize, and hand-code characteristics of every task or output listed on the annual workplans and reports of each ministry and department in Ghana's core civil service. These tasks included work on physical infrastructure projects (e.g., "construction of secondary data centre at Kumasi," "identify bungalows and initiate procurement process"),

public-facing activities and awareness-raising (e.g., “talk shows in four rural district markets in the region on the GIPC act held,” “sensitize printers and suppliers on the procurement law and packaging”), internal administrative tasks (e.g., “preparation of 2014 annual report,” “conduct second phase of housing audit”), as well as tasks related to policy development, training, financial management, permitting and regulation, and other categories.¹⁰ Altogether, we examined 3,620 tasks from thirty government organizations.

For each task, we coded: (1) *ex ante* clarity, or the extent to which “the task can be defined in such a way as to create little uncertainty about what is required to complete the task;” and (2) *ex post* clarity, or the extent to which “a report of the actual action undertaken leaves little uncertainty about whether the task was effectively completed.” These correspond closely to the characteristics of pre-specifiability and measurability that together comprise the concept of verifiability. These two variables were coded on a scale from 1 to 5, with 5 meaning “no ambiguity” and 1 meaning “undefined or so vague it is impossible to assess what completion would mean.”¹¹ Coders had relatively little information on which to base their coding—usually just a phrase or sentence in a reporting table—although this is also the exact same information used by organizational management and by OHCS to monitor performance, so it presents a fairly realistic way to quantify the verifiability of tasks.

The results of our coding are reproduced in figure 2.1. Just 22.7 percent of tasks were judged by the coders to be perfectly clear *ex ante* (score of 5), and 19.1 percent were judged to be perfectly clear *ex post*. Only 11.2 percent of tasks were coded as completely unambiguous in each category—that is, perfectly verifiable. At the same time, very few tasks were completely unverifiable. The vast majority of tasks were partially clear on both measures. Based on this measure, then, almost 90 percent of what civil service bureaucracies do is at least partially nonverifiable—and organizational managers thus have to find ways to guide, monitor, and encourage both completion of formalizable and unformalizable tasks or task components.

ORGANIZATIONAL PROCESSES AND INDIVIDUAL PERFORMANCE HAVE NONVERIFIABLE ELEMENTS

The formal versus informal distinction can also help us understand the organizational management processes that direct, coordinate, and monitor individuals and teams as they work to achieve these tasks and outputs. In 2013, I conducted interviews with sixty civil servants across forty ministries, departments, and agencies in Ghana’s public service about how a common set of fifteen different organizational

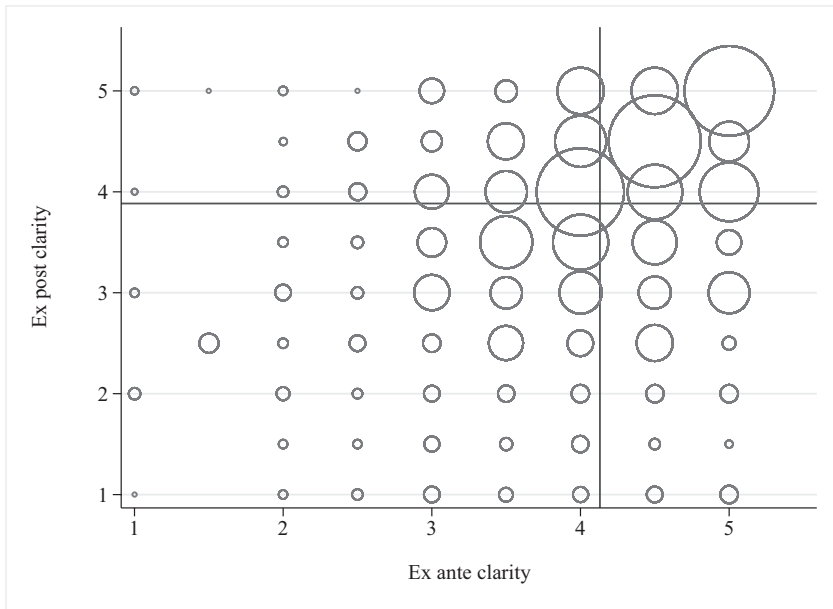


FIGURE 2.1 Task Clarity, Ex Ante and Ex Post

Notes: Circle size is proportional to the number of tasks that fall within each bin of width 0.5. Lines indicate mean values for each measure of task clarity. Reproduced with permission from Imran Rasul, Daniel Rogger, and Martin J. Williams, “Management, Organizational Performance, and Task Clarity: Evidence from Ghana’s Civil Service,” *Journal of Public Administration Research and Theory* 31, no. 2 (2021): 259–77.

processes worked in their organization.¹² Our conversations focused on how these processes were actually executed in practice—not just what they were supposed to look like on paper. For some of these processes, the formal aspects were actually defined at the system level rather than the organizational level, so in principle, the process should have looked the same in every organization.

What I observed in practice was a huge variation in de facto management processes across organizations. This variation corresponded naturally to the three categories of performance I defined above: nonperformance, perfunctory performance, and consummate performance. To illustrate this, let’s first examine a process that was meant to be the same in every organization: annual staff appraisals.¹³ In each organization, this was supposed to involve completion of the same appraisal template by each officer, as well as the same annual cycle of defining targets and responsibilities in conjunction with the officer’s supervisor at the start of each year, assessing performance at the end of the year, and feeding the performance assessment into decisions on training allocation. These assessments

were also intended to feed into promotion decisions; while these were made at a central level rather than by the organizations themselves, in practice, organizations had some influence on promotions because they could choose whether and how strongly to recommend officers to be considered for promotion by the central authorities.

Table 2.2 summarizes the range of annual staff appraisal practices that were actually used in organizations. One group of organizations fell into the category of nonperformance: the formal parts of the staff appraisal process were not complied with, nor were there systematic informal efforts to make the appraisal

TABLE 2.2 Variation in formal and informal staff appraisal practices, Ghana 2013

Standardized <i>de jure</i> process in all ministries			
All officers should sit with their manager at the end of year and complete an annual appraisal (based on a standardized template) to assess performance against pre-defined targets.			
Variation in <i>de facto</i> practices across ministries			
	Nonperformance	Perfunctory performance	Consummate performance
Formal aspects of practice	Appraisals are not conducted annually. Instead, multiple years are filled out at the same time when individual are due for promotion.	Appraisals are always conducted annually.	Appraisals are always conducted annually.
Informal aspects of practice	The appraisal is a self-assessment; actual discussions about officers' performance rarely happen.	Appraisal discussions are not meaningful, and process is viewed as a formality.	Appraisals are used as an opportunity to give meaningful feedback, and are supplemented by discussions in weekly divisional meetings.

Source: Adapted from Martin J. Williams, "From Institutions to Organizations: Management and Informality in Ghana's Bureaucracies," working paper, London School of Economics and Political Science, September 9, 2015.

process function as intended. In these organizations, the appraisal process was often not even carried out annually; rather, individuals tended to complete several years' worth of appraisals retrospectively as they neared the date for their promotion interviews, simply entering a different year on each form. In these organizations, the appraisal itself was regarded as essentially a self-assessment that the supervisor merely signed off on, except in cases of extreme misconduct, with no link to any rewards, consequences, or remedial action. As one interviewee explained, "The appraisals don't improve the system. They're a formality." Another reflected on the disconnect between assessed performance and the allocation of training opportunities, saying, "One director even says he'll send you 'if your face looks nice.'"

Another group of organizations carried out the process perfunctorily. Interviewees reported that appraisal forms were indeed carried out each year, and supervisors and supervisees did sit together to complete a list of targets at the start of the year and assess performance at the end of the year, but the target-setting process was often disconnected from the actual work officers did during the year, and the assessment tended to be equally uninformative. In a similar vein, research on Ghana's performance appraisal system around this period by Justice Nyigmah Bawole and colleagues describes it as "rhetoric rather than an important practice."¹⁴ and Frank Ohemeng and colleagues refer to the process as "much ado about nothing."¹⁵

A third group of organizations undertook the staff appraisal process with what can be described as consummate performance. These organizations not only carried out the formal appraisal process as designed but also instituted complementary informal measures to make the formal process more meaningful. Supervisors held the standard annual and midyear performance review meetings with each of their staff but also held informal discussions on performance in weekly divisional meetings so that individuals were not only kept aware of their own performance but also how they compared to others in the division. Due to the centrally imposed constraints on pay and promotions, organizations found other ways to recognize and reward good performance: awards nights for top performers as voted by their colleagues, occasional written letters of recognition for exemplary work, opportunities for training, and so on. While accelerated formal promotions were rare, given the rigidity of the system, it was common for these organizations to reward star performers with "informal promotions": appointment to committees, roles as a focal person, and other opportunities that further career development and may also entail some financial benefit. As one interviewee remarked, "This is one of the ministries where you are recognized based on your work, not your rank."

The same pattern was evident with respect to other types of management processes, including those that pertain to the organization as a whole rather than individual managers and officers. For example, each organization was mandated to hold regular management meetings to discuss performance, but the ways this formal mandate was implemented varied wildly. Some ministries didn't even comply with the mandate, holding performance review meetings on an irregular and ad hoc basis only, perhaps four to five times per year. Even when these meetings were held, they were typically not very useful, with one interviewee reporting that such meetings were "more for human management and the condition of the office and that kind of thing than results management. There's very little discussion of what are we supposed to do and where are we."

Other organizations complied perfunctorily, holding regular management meetings (most commonly once a month) at which some matters related to performance were reviewed. But these discussions rarely were communicated to non-management staff or had implications beyond the management meeting; they were merely a formality. These organizations could truthfully say on their annual report that they had held regular performance meetings, but the nonverifiable reality was that such meetings likely contributed little to performance improvement.

Another group of organizations carried out these performance reviews consummately, holding regular management meetings for operational and administrative issues (usually more frequently than monthly) as well as periodic broader reviews. Management staff would typically hold follow-up meetings with their teams to brief them after the management meeting so that staff were aware of what was happening elsewhere in the organization and how it affected their work, and cross-team working groups were formed to tackle issues that arose. For these organizations, the formal process was not only carried out but also prompted and underpinned a wide range of less formal supporting practices that were important for achieving the intended goal of the formal process. These supporting practices were difficult to formally mandate but were nonetheless carried out systematically in the group of organizations that executed this process consummately.

While thinking of management processes and their performance as comprising both formal and informal aspects is broadly useful, some nuances and clarifications are in order. First, not every organization I interviewed fell neatly into one of these categories, and there was some variation across managers within each organization. So these three levels of performance—nonperformance, perfunctory performance, and consummate performance—are not rigid, discrete categories but rather simple heuristics for characterizing the inevitably more complex world of organizational processes and practices. Second, there is a fourth potential category of performance that could exist: organizations that don't have or

TABLE 2.3 Variation in formal and informal organizational performance review practices, Ghana 2013

Standardized *de jure* process in all ministries

All organizations should hold regular (usually monthly) meetings of management-level staff to review performance and coordinate across divisions, as well as quarterly and/or annual reviews with participation from a broader range of staff.

Variation in *de facto* practices across ministries

	Nonperformance	Perfunctory performance	Consummate performance
Formal aspects of practice	Management meetings are held occasionally but sometimes don't happen. Quarterly reviews exist some of its programs, but only the donor-led ones are regular.	Performance review meetings with management staff happen every month, as do annual and/or quarterly all-staff meetings.	Performance review meetings happen monthly or even weekly with management, as well as periodically with different groups of staff, so everyone gets to know about the agency's performance.
Informal aspects of practice	Quarterly and annual reports are seen as a formality undertaken to meet external requirements, not an opportunity to review performance. Management meetings "more for human management and the condition of the office and that kind of thing than results management. There's very little discussion of what are we supposed to do and where are we."	Directors are supposed to brief their officers after management meetings, but this does not always happen. More senior officers would learn about the organization's performance through working with their directors. Agreed actions would be minuted to the officer responsible. "That's if the minutes can be done in good time, sometimes they can take long."	Directors meet regularly with their divisions to update them on discussions from management meetings and to review their own performance, sometimes holding informal briefings immediately after management meetings. Even routine reporting requirements, like quarterly reports, are seen as opportunities to document and improve performance.

Source: Adapted from Williams, "From Institutions to Organizations."

don't carry out the formal aspects of these processes but do systematically employ informal practices to achieve the goals of the process. This is a theoretical possibility, but in practice, I didn't observe this as a deliberate managerial approach in any of the forty ministries, departments, and agencies I studied. To the extent performance-oriented informal practices did exist without being linked to or supported by any formal practices, it was nearly always a matter of individual managers trying to do their best to cope with organizational failings rather than a conscious strategy for good management. Third, with these two management processes, as well as the other types of formal management processes imposed by the reforms I study later in the book, my focus on the execution of formal processes is not meant to imply an assumption that each of these formal processes is optimal or even necessarily positive for performance in the organization as a whole. Rather, I focus on execution or implementation under the much weaker assumption that civil services aim to implement the processes that they define for themselves in meaningful ways and that doing so, in general, will lead to performance improvement (overall, if not necessarily for every single process). In other words, it's likely that some processes defined by any given civil service are likely to be suboptimal or even harmful for performance and, hence, need to be continuously evaluated and improved. In a similar vein, the design of formal processes cannot necessarily be separated from their implementation—a theme we will return to later in the book.

The evidence described above and this book's conceptual starting point is that good organizational performance generally stems from both executing the organization's defined formal processes, at least for the most part, and from simultaneously carrying out complementary informal practices that give meaning to these formal processes. This view of management and performance as comprising both formal and informal components that must complement each other also accords with evidence from other types of organizations around the world.¹⁶ So moving toward consummate performance of organizational processes, as opposed to perfunctory or nonperformance of them, can serve as a simple, parsimonious, and broadly applicable goal for performance improvement efforts.

PERFORMANCE VARIES ACROSS ORGANIZATIONS WITHIN THE SAME CONTEXT

These process-level variations aggregate up into substantial differences across organizations in overall management quality and performance. Academics and practitioners alike sometimes assume that the picture of performance in the

public sector is uniformly bad. But a large body of evidence points to a different conclusion: There is actually substantial variation in performance across organizations, teams, and individuals within any given government. This is despite these organizations sharing the same national-level variables that are commonly thought to affect bureaucratic effectiveness—geography, political settlements, administrative structures and legacies, colonial history, societal education and human capital, and so on¹⁷—and in many cases, operating under the same set of formal civil service laws and regulations.

The interviews I conducted in 2013 across forty Ghanaian ministries, departments, and agencies described in the previous section provide one way to illustrate this point. In addition to recording qualitative information about the management processes being used, I also quantitatively benchmarked the quality and consistency of practice within each organization. Adapting the influential and widely used World Management Survey methodology developed by Nicholas Bloom and John Van Reenen,¹⁸ in each organization, I scored each of the fifteen processes I studied on a scale from 1 to 5, where 1 meant that the organization had essentially no structured or consistent approach to the process, and 5 meant that the organization consistently and thoughtfully implemented both formal and informal aspects of that process. The interviewing and benchmarking procedures incorporated a wide range of methods to avoid bias and ensure comparability across organizations,¹⁹ and the benchmarking criteria were designed to be neutral with respect to the style of management and instead focus on whether for each practice the organization had a process or routine that was consciously designed and was followed in reality.

Figure 2.2 shows the portrait of management processes that emerges from aggregating these scores within each organization, with the black diamonds representing the organization's average score across practices and the grey circles representing the organization's score on each process. (In this figure, all scores have been normalized so that the mean is equal to 0 and the standard deviation is equal to 1.)²⁰

The resulting picture was one in which there was substantial variation both across organizations in overall management quality and within each organization across different processes. Some organizations were better managed than others, and some organizations that were managed quite poorly (or well) overall had specific processes that they executed well (or poorly). So the qualitative differences in how organizations implemented each management process aggregated up into broader differences in overall management quality despite these organizations all operating in the same institutional context.

Organizations in the same government also vary in their level of performance. In the research that Imran Rasul, Daniel Rogger, and I conducted in

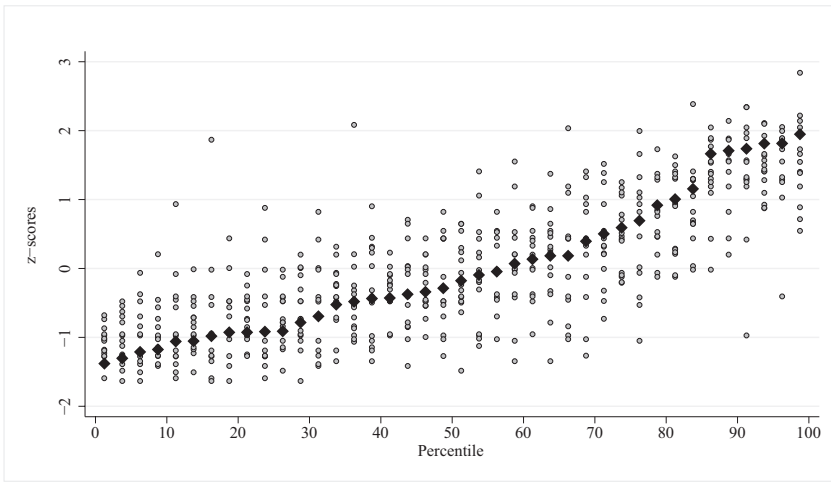


FIGURE 2.2 Variation in Management Quality Across and Within Organizations, Ghana 2013

Notes: Black diamonds represent the overall management quality score for the organization, grey circles represent the score for each of the fifteen management processes individually. All figures are normalized z-scores. Reproduced from Williams, “From Institutions to Organizations.”

collaboration with Ghana’s OHCS in 2015, we not only coded the ex ante and ex post clarity of each of the 3,620 tasks (as I described earlier in the chapter) but also the degree to which the organization completed each task. We coded completion using the information provided in the organization’s own reports, with checks on data quality and accuracy,²¹ also using a 1–5 scale where 5 meant “no action was taken towards achieving the target” and a score of 5 meant “the target for the task has been reached or surpassed.”²²

Figure 2.3 shows the average task completion rates for each of the thirty organizations we studied (black diamonds) and each division within them (grey circles). The picture was again one in which there was substantial variation in performance both across organizations and within them.

As part of this research, we also partnered with Ghana’s OHCS to conduct a large-scale survey of 2,971 civil servants across forty-five ministries, departments, and agencies. We asked about the de facto management processes these civil servants experienced on a day-to-day basis, using a 1–5 coding methodology similar to what I had used in my earlier interviews and Rasul and Rogger had previously used with Nigeria’s Federal Civil Service.²³ Using this quantitative metric based on thousands of interviews, we again found the same pattern: substantial variation in management process quality across organizations, across divisions within

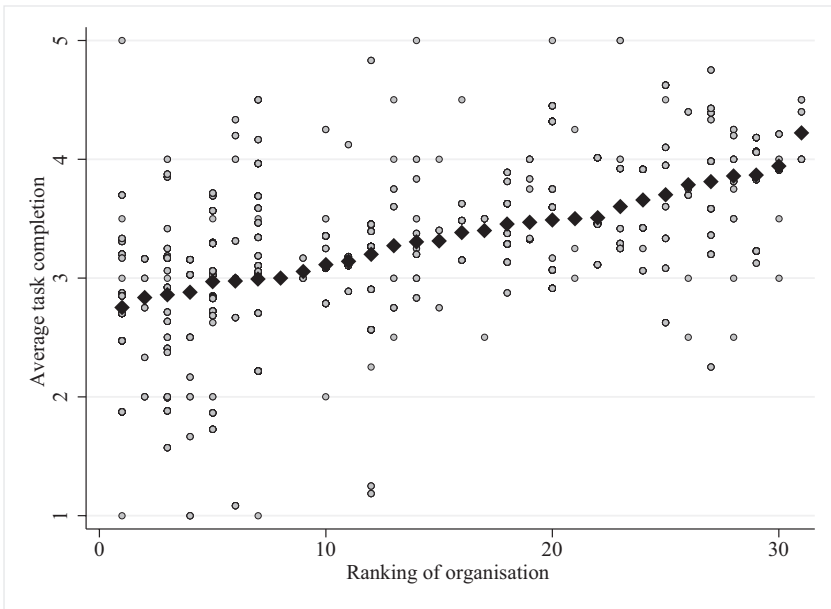


FIGURE 2.3 Variation in Task Completion Across and Within Organizations, Ghana 2015

Notes: Analysis undertaken for Rasul et al., “Management, Organizational Performance, and Task Clarity.” Black diamonds represent organization average task completion rates, grey circles represent division average completion rates.

organizations, and across different process types within organizations. This variation in management quality was strongly associated with variation in task completion, even after controlling for a range of other features of organizations, tasks, and personnel. So better or worse organizational performance could be traced back (in part) to better or worse organizational management quality, which could, in turn, be traced back to variation in the execution of both formal and informal aspects of specific management processes.

These studies are far from unique in finding that there is variation in performance among organizations undertaking similar functions in similar contexts. In Ghana, for example, Francis Owusu used survey-based methods to identify high- and low-performing agencies within government and argued that organizational recruitment criteria and remuneration levels were associated with variations in performance.²⁴ Also in Ghana, Erin McDonnell conducted in-depth qualitative research on four high- and four low-performing teams within the same organization and showed how managers of high-performing teams simultaneously

tried to create distinctive performance-oriented cultures while also carving out the autonomy to build and maintain these cultures.²⁵ More broadly, there is a substantial literature on “islands” of effectiveness and “pockets” of excellence, building on the foundational work of Judith Tandler and Merilee Grindle, which primarily uses qualitative case studies to demonstrate that high-performing organizations exist in the public sectors of states around the world with supposedly weak institutions.²⁶ Surveying this literature, McDonnell points out that even the patterns of which types of agencies perform better vary across countries—an observation that pushes back against simplistic, function-oriented explanations for this variation and hints at the importance of internal management and leadership in producing organizational effectiveness.²⁷

Quantitative evidence of variation in effectiveness also exists in other governments around the world. For example, Daniel Gingerich, Katherine Bersch and colleagues draw on different sources of quantitative data to show that there is substantial cross-organization variation in various measures of capacity (among other variables) in the governments of several Latin American countries.²⁸ In Russia, Michael Best and colleagues used data on public procurement from 2011–2016 to show that 39 percent of the variation in prices paid for identical items is due to the differential effectiveness of individuals and organizations.²⁹ In their work in Nigeria’s Federal Civil Service, Rasul and Rogger report that infrastructure project completion rates vary from 4 percent to 89 percent across different organizations. In the United States, Patricia Ingraham and colleagues document a wide range of variations in management and performance across federal government agencies.³⁰ Even among private sector firms, a large empirical literature documents “persistent performance differences” among firms producing identical products in the same market, as Robert Gibbons and Rebecca Henderson summarize and theorize.³¹

Taken together, these studies demonstrate that while forces external to organizations (history, politics, geography, etc.) might encourage or inhibit bureaucratic performance to some degree, they do not fully explain it. This finding is a hopeful one for reform efforts in that it suggests that internal management processes also play an important role in determining organizational performance—and thus that reforms that find a way to succeed in improving management might also be able to improve overall performance.



This chapter has sought to provide a conceptual and empirical starting point for this book’s study of systemic, performance-oriented civil service reforms.

To do so, it zoomed in to the levels of individuals, teams, and organizations and introduced the language of verifiable and nonverifiable—formal and informal—actions and processes to characterize what civil servants do and what it means to do it well. It also provided empirical evidence to illustrate and support this view of work in public organizations.

What this chapter hasn't done is attempt to provide any answers to the question of how governments should design and implement reforms. Rather, the point of this chapter is to better frame the goal of (and key challenge for) performance-oriented reforms to provide a strong foundation for the subsequent parts of the book. The four chapters in part II will thus aim to document and explain the track record of such reforms across the six governments covered in this book. Part III will then turn toward trying to provide an answer, generating a theory of organizational change and system reform that is informed by the evidence from part II and that builds on the foundations laid in this chapter. With this conceptual framework in mind, then, let's proceed to part II.

