2011

Public Accountability: Performance Measurement, The Extended State, And The Search For Trust

Melvin J. Dubnick
University of New Hampshire - Main Campus, Mel.dubnick@UNH.edu

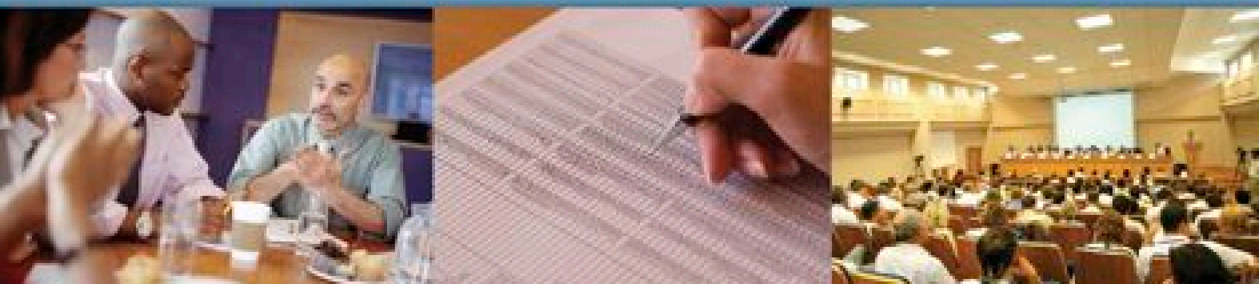
H. George Frederickson
University of New Hampshire - Main Campus

Public Accountability

Performance Measurement, the Extended State, and the Search for Trust

Melvin J. Dubnick

H. George Frederickson



National Academy of Public Administration®

Kettering Foundation

The Academy

The National Academy of Public Administration is a nonprofit, independent organization of top public management and organizational leaders who tackle the nation's most critical and complex public management challenges. With a network of more than 680 distinguished Fellows and an experienced professional staff, the Academy is uniquely qualified and trusted across government to provide objective advice and practical solutions based on systematic research and expert analysis. Established in 1967 and chartered by Congress in 1984, the Academy continues to make a positive impact by helping federal, state, and local governments respond effectively to current circumstances and changing conditions. Learn more about the Academy and its work at **www.napawash.org**.

Kettering Foundation

The Kettering Foundation is an operating foundation rooted in the American tradition of cooperative research. Kettering's primary research question today is, what does it take to make democracy work as it should? The research is conducted from the perspective of citizens and focuses on what people collectively can do to address problems affecting their lives, their communities, and their nation. The foundation seeks to identify and address the challenges to making democracy work as it should through interrelated program areas that focus on citizens, communities, and institutions. Established in 1927 by inventor Charles F. Kettering, the foundation is a 501(c)3 organization that does not make grants but engages in joint research with others. It is an operating foundation headquartered in Dayton, Ohio, with offices in Washington, D.C., and New York City. For more information, call 800-221-3657 or visit the foundation's Website at **www.kettering.org**.

PUBLIC ACCOUNTABILITY:

Performance Measurement, the Extended State, and the Search for Trust

Melvin J. Dubnick
H. George Frederickson

ISBN Number 978-0-923993-36-8

An electronic version of this paper is available online at **www.napawash.org**.

This page intentionally blank

INTRODUCTION

Established in 1921, the venerable General Accounting Office of the United States of America changed its name in 2004 to the Government Accountability Office (GAO). Why? First, as the investigative arm of Congress, the GAO was always more than an accounting agency. Second, given the present unsavory reputation of some accounting firms and some accounting practices, it was wise and timely of the GAO to dump that word with all its baggage. Third, it was clever of the GAO to change two words in its title and still keep the GAO symbols—symbols of nonpartisan fairness, balance, and toughness. Fourth, and most important, by replacing *accounting* with *accountability* in its title, the GAO associated itself with one of the most powerful symbols in modern politics and administration.[1]

Public accountability, in ordinary and traditional usage, refers to grand acts of account giving like the War Crimes Trials in Nuremberg, Germany, the International Military Tribunal for the East in Japan following World War Two, or the truth and reconciliation process in South Africa after the collapse of apartheid. In everyday usage, *accountability* refers to or describes the process and results of one person settling up or getting right with another. While still used in this way, *accountability* now has far broader meanings.

It seems that accountability is the air we breathe, the oxygen of politics, and the hope of administration. In short, *accountability* is the word that is eating government. According to proponents of accountability-centered reforms, enhanced accountability will (among other things) result in greater transparency and openness in a world threatened by the powerful forces of hierarchy and bureaucratization, greater access to impartial arenas where abuses of authority can be challenged and judged, increased pressure and oversight that will promote more ethical and appropriate behavior on the part of public officials, and further improvements in the quality of government service. Most important for the purposes of this essay, accountability will bring formal and precise measures of performance to government so the public can know how well their government is meeting public expectations.

1 For an explanation of the GAO name change, see the comment of then-Comptroller General David M. Walker (2004): "I am not suggesting that agencies need to change their names—but most of them do need to come to grips with the fact that some of their most basic policies, processes, and procedures are years out of date. We at GAO have a proud history, but we are not defined solely by our past. We will still be known as GAO, but our new name will make clear that our first priority is to improve the performance of the federal government and ensure its accountability to Congress and the American people."

Modern applications of accountability carry with them promises of enhanced democracy, greater justice, more ethical behavior, and improved government performance. Among these, it is the last claim—that modern applications of accountability will improve government performance—that we find most interesting, because that claim is at the core of what governments are supposed to do. We ask: Is there a basis for the assumed relationship between accountability and public organizational performance? Will greater accountability mean improved public organizational performance? Will improved public organizational performance guarantee greater accountability? If so, how, and to whom?

We find it puzzling and troubling that, although a direct and clear positive relationship between accountability and improved public organizational performance has not been widely explored, proponents of public-sector reforms based on performance measurement assume it. Both accountability and public organizational performance, particularly performance measurement, are part of the wider body of so-called "new public management" (NPM) administrative reforms, and in those reforms the words *performance* and *accountability* tend to be used interchangeably and to mean approximately the same thing. In the circles of government and public affairs there is, then, an influential and vocal accountability and performance-measurement reform hegemony (Radin 2006). While such reformist rhetoric is well intentioned as a means for justifying and energizing the changes that advocates prefer, proselytizing to "the gospel of accountability" is not conducive to thoughtful reflection or analysis.

The purpose of this monograph is to unpack the concepts and perspectives upon which the logic of accountability and performance rest. This examination of the foundations of accountability and performance reaches for a clarification of meanings and with that clarification attempts a reasoned evaluation of the strengths and weaknesses of the assumed association between accountability and performance in both theory and application. To aid in this evaluation we employ the example of the application of accountability and public organizational performance to the so-called Extended S tate, the modern practice of outsourcing to third-party contractors and grantees the organization, management, and implementation of governmental projects and programs. We choose this example out of the conviction that the contemporary Extended State is a particularly vexing challenge to the application of accountability concepts and performance-measurement protocols.

We start by presenting a framework for understanding the nature of contemporary accountability and the problems associated with it as applied in the modern extended form of the Administrative State. The basic argument we make is straightforward: the current problems associated with accountability in the Extended State are rooted in a widely held but unsubstantiated belief in the capacity of accountability mechanisms alone to bring about the three things we require of our government today—efficient control, democratic legitimacy, and effective performance. We call this the "promises of accountability" framework. It reflects the view that (more often than not) policies are passed and programs designed based on untested assumptions (i.e., promises) about what accountability mechanisms can accomplish. We contend that the problematic governance of today's increasingly powerful Extended State is tied to the idea that the wide range of accountability mechanisms we apply in efforts to achieve those three core objectives of governance are in and of themselves sufficient to achieve them.

The basic argument we make is straightforward: the current problems associated with accountability in the Extended State are rooted in a widely held but unsubstantiated belief in the capacity of accountability mechanisms alone to bring about the three things we require of our government today—efficient control, democratic legitimacy, and effective performance.

It is within the general framework presented in the next section of the monograph that we also try to expand our understanding of the complex relationship between accountability and performance. We follow this with an overview of the three major objectives of accountable governance (i.e., control, legitimacy, and performance), and after discussing each in turn we turn to the case of performance measurement in the Extended State. Finally, we make the case for an approach through which the promise of accountability (as well as the other promises) might be achieved.

The Promises of Accountability

Accountability is both a word and a bundle of concepts. As a word, *accountability* is notoriously ambiguous (Brooks 1995), chameleon-like (Sinclair 1995), and its synonymic nature—that is, its capacity to stand in for a range of other strong terms, such as *responsibility*, *liability*, *responsiveness*—render it subject to rhetorical uses and abuses. In recent years it has also taken on an icon-like symbolic power, used too often to manipulate political forces (Dubnick 2002). It is notable, however, that the word is etymologically of recent vintage, emerging in 12th-century English from Norman French origins.[2] Yet, conceptually the idea underlying the word has been around for centuries, allowing many historians to speak of accountability found in the governance of ancient regimes.[3]

Despite this ambiguity and conceptual looseness, we need to establish (at least at the outset) a working definition for *accountability*. For present purposes we will assume a simple, traditional view of accountability as a launching point for the promises of accountability schema presented below. Accountability can be approached as having two major definable characteristics: (1) it is a social relationship between at least two parties (2) in which at least one party to the relationship perceives a demand or expectation for account giving between the two. This definition highlights the fact that accountability is social in nature—it must involve two or more individuals to come into play in any relationship.[4] That relationship, however, need not be explicit or involve the active or ongoing engagement of both parties; rather, the demand for account giving needs to be assumed (i.e., expected and accepted) by at least one of the parties in the perceived relationship.[5]

For example, consider the case of two individuals (let's call them Mary and Joe) who are complete strangers in all other respects except for the fact that they both pass the same four-way stop intersection at approximately the same time each morning on their commute to work. Despite being total strangers, as licensed drivers they operate their vehicles within the social setting created by traffic laws. As they approach the intersection at the same time, each expects and assumes the other feels accountable for their behavior at the intersection and will come to a full and complete stop. At the same time, each possesses a sense of being accountable for their

2 Dubnick (1998) argues that the term is "anglican"—so much so that it had no equivalent even in the Romance languages—not even French, from which it is derived—until recently. Instead, most languages treated the English (for translation purposes) as a variation of *responsibility*.

3 For example, see Roberts (1982) and Elster (1999) on accountability in ancient Greece; Ezzamel (1997) examines accounting and accountability in ancient Egypt.

4 In that sense, the attribution that someone is accountable is similar to saying that she is influential or powerful: implied in such characterizations is the question "to whom?" The frequent confusion between accountability and responsibility may rest on this point. Responsibility can imply the "to whom?" question, but it primarily focuses attention on "for what?"; in contrast, accountability can imply "for what?" but draws one initially to the "to whom?" query. For a related discussion, see Uhr (1999).

5 This is a subtle but important point because it puts aside any notion that someone is inherently or existentially accountable. At the individual level, it restricts our attention to at minimum an unreciprocated "felt accountability," which involves at least a perceived relationship with an "other." On the role of "otherness" in accountability, see Shearer (2002) and (for a more philosophical perspective) Butler (2001).

actions at the intersection. The ever so brief (and one hopes safe) relationship at that intersection is a classic instance of simple and effective accountability. Had either momentarily abandoned or lost the sense of accountability—e.g., had Mary blatantly ignored the stop sign and crashed into Joe's car—she would be held to account for her actions because the event occurred within the frame of an established social relationship. Joe, relying on the mechanisms provided by the laws of the jurisdiction, would have the ability to hold Mary to account, and Mary (given the fact that she was a licensed driver) would submit to being held to account. Of course, Mary might offer Joe some explanation or excuse for the accident ("a fallen tree branch hid the stop sign from view"), but that does not mean she is not subject to being held to account.

Most often, daily instances of accountable relationships are most clearly in evidence after an event, such as a car accident or other mishap, that involves perceived errors of choice or judgment on the part of one party to a social relationship. Under this view, the accountability is typically associated with actions taken "after the fact" or *post factum* to an accountable matter or event. This common perspective is frequently applied to more complex situations. It includes post-factum reconsideration of big policy choices like the decision to incarcerate Japanese Americans during World War Two or the decision to use atomic weapons on Japanese cities in that war. Traditional accountability also includes post-factum attempts to fix responsibility for perceived human errors, such as those thought to have caused the *Challenger* and *Columbia* space program disasters. It also encompasses efforts to fix responsibility for ineffective governmental responses to unanticipated events such as the Katrina and Rita hurricanes in 2005 and other natural disasters, as well as man-made disasters like the 2010 Deepwater Horizon oil spill in the Gulf of Mexico.

Most often, daily instances of accountable relationships are most clearly in evidence after an event, such as a car accident or other mishap, that involves perceived errors of choice or judgment on the part of one party to a social relationship. Under this view, the accountability is typically associated with actions taken "after the fact" or post factum *to an accountable matter or event. This common perspective is frequently applied to more complex situations.*

What we often forget is that those post-factum responses are premised on the existence of *pre-factum*, or before-the-fact, expectations and assumptions about the behavior of individuals, groups, and even nations. Relationships based on accountability, whether perceived as the ability to hold someone to account or the capacity to be held to account, require an awareness and appreciation of both pre-factum and post-factum dimensions. As we will argue later in this monograph, break or qualify the link between the two—that is, disconnect the accountability relationship from its social, legal, or even moral setting—and what you have is something that is accountability in name only. We believe that such a disconnect has occurred in the general effort to enhance accountability, and that the source of the problem can be traced to the emergence of what we term the "promises" of accountability.

To fully appreciate the power of the promises of accountability that we discuss below, we need to highlight another feature of the pre- and post-factum connection. To deal with this, let's return to our intersection where both Mary and Joe have dutifully come to full and complete stops at exactly the same time, but this time they are joined (at precisely the same moment) by a third vehicle driven be Eloise and a fourth driven by Carl. Each contemplates the next move since the general rule (the vehicle in the intersection first has right of way) does not apply, nor does the common secondary rule (the vehicle furthest to the right should proceed). Finally Joe steps on the accelerator pedal just as Mary has decided to do the same. Despite the existence of pre-factum accountability in the form of traffic rules and norms, the resulting accident raises issues about who is at fault under these circumstances. After a number of such incidents over a period of time, local residents and other regular commuters urge community policymakers to do something to address the problem. The fact that Mary and Joe are working out the issue of post-factum accountability through the legal system is not enough, and there is a call for some modification of pre-factum conditions to deal with the situation, hopefully to prevent a recurrence. Since replacing the stop signs with a traffic signal would prove too expensive for the community and the existing rules are not always relevant, the traffic engineers suggest posting a sign that reads "Proceed with caution and courtesy" under each stop sign.

What this little fictional narrative highlights is something that is central to the emphasis we increasingly place on accountability-based reforms. When we call for changes in policies and programs in order to enhance accountability, we are primarily focused on modifying pre-factum accountability. Whether it involves posting signs of caution, requiring certain forms be filed, or the passage of a "three strikes" law that mandates harsher sentences for repeat offenders, what we are attempting to do is prevent or preclude post-factum accountability through changes in pre-factum accountability settings.

When we consider the push for greater accountability in the public sector, our attention is usually drawn to efforts aimed at public administration.[6] Historically, there have been three major problem areas where calls for greater accountability have had a major impact. The first is the problem of government efficiency and productivity. For more than a century, reformers have pursued a range of changes that would enhance the ability of public agencies to control and rein in "waste, fraud, and abuse." This problem normally affects the organizational setting within which government conducts its business, and for present purposes we call this the "problem of control."

The second problem, unique to work of governments, reflects the fact that public administration in the United States must operate within a system of governance defined by its commitment to democratic norms and values that give public programs and policies their legitimacy. In fact, we term this the "problem of legitimacy."

Finally, beyond the push for greater legitimacy and efficient productivity, government administrators are under pressure to ensure that the outcomes of their programs and policies are effective—and that they perform up to the expectations of the wide range of stakeholders that are invested in the work of government. Here we discuss the "problem of performance."

As we elaborate below, the problems and challenges generated by each of these government settings (see Table 1) have created a demand for both more and different forms of accountability, some post-factum (e.g., stricter enforcement of existing rules and regulations) but most of the pre-factum sort. These pre-factum forms involve taking steps designed to "cause" individuals or agencies to assume an account-giving attitude or posture "before the fact," thereby motivating them to act accountably either in the present or the future.

6 In many ways the American judicial system, based on the rule of law, is a large-scale accountability system. Because of the vastness of the judicial system and because we are focusing on organizational, performance-based, and democratic process accountability, we do not include it here. The logic of due process is designed to guarantee just treatment at the bar. The American judicial system—the police, the courts, jails and prisons, probation and parole—are elaborate and complex systems of bringing citizens to account for their actions. In this dense complexity the accountability promises of fairness or equity are vexing. Justice is one thing, fairness is another. Fairness issues tend to be policy specific, as in fair housing, equal schooling, and so forth. In this policy specificity, fairness issues also tend to manifest themselves at the stage of policy implementation and are laden with accountability questions. The splendid recent field research by Steven Maynard-Moody and Michael Musheno details the vexing accountability choices faced by police officers, teachers, and counselors as they make choices between clients in the context of scarce time and resources (Maynard-Moody and Musheno 2003). Fairness and justice at the street level are artifacts of bureaucratic accountability, as line-level agents practice a kind of accountability triage, as they sort through claims of need and worthiness on one hand and resources on the other.

The problems and challenges generated by each of these government settings have created a demand for both more and different forms of accountability, some post-factum (e.g., stricter enforcement of existing rules and regulations) but most of the pre-factum sort. These pre-factum forms involve taking steps designed to "cause" individuals or agencies to assume an account-giving attitude or posture "before the fact," thereby motivating them to act accountably either in the present or the future.

In a very significant sense, many of these pre-factum arrangements reflect a general expectation that things can and do go wrong—that "stuff" happens, whether in the form of human errors, natural disasters, or some other unanticipated events. Thus, the roles and significance of pre-factum mechanisms in today's governance structures cannot be overestimated and certainly not ignored. Some of them are quite visible, often seeming like ritualistic demonstrations of deference of the accountable official to those to whom she or he is accountable (e.g., the State of the Union/State/City addresses given annually by elected executives can be seen in this light). At the other end of the spectrum are mundane and almost invisible mechanisms, such as required record keeping and audits, which almost every public entity must maintain and submit to on a regular basis. These are commonplace, pervasive, and frequently invisible until, usually as a result of some misfeasance or malfeasance or "act of nature," things run amiss. At that point the most subtle forms of pre-factum mechanisms—the record keeping, the annual audits, the retained e-mails—become the basis for investigations, fault finding, blaming, shaming—as well as explaining, excuse making, disclaimers, justification, and other means by which the gaps between administrative actions and policy or program expectations were bridged.

Setting	Seeking	Focus on	Problem of
Organizational	Order and compliance (e.g., to achieve operational efficiency)	Inputs	Control
Policy/Program	Goal attainment, programmatic ends (e.g., effective delivery of services)	Outputs	Performance
Governance	Good governance, democracy, responsiveness (e.g., greater participation, transparency)	Processes	Legitimacy

TABLE 1: Problems of Accountability

But as important as pre-factum mechanisms have become in the effort to reform government, they have also became a source of problems and dysfunctions even greater than those they were designed to tackle. As a result of the relative success of these pre-factum mechanisms in dealing with the problems of control, legitimacy, and performance, advocates of reform often have been lured into assuming that it was the pre-factum mechanisms per se that were the keys to accountable governance. Rather than treat the pre-factum accountability mechanism as one among several factors and contingencies that improved control, legitimacy, or performance, they begin to approach it as a sufficient (*post hoc, ergo propter hoc*) condition for successful reforms.

The result can be seen in one of two patterns developed in the treatment of accountability. On the one hand, pre- and post-factum mechanisms became completely decoupled, and the pre factum come to be regarded as managerial and policy tools sufficient unto themselves to bring about the appropriate or desired degrees of control, legitimacy, and performance. Thus, if you wish to have more ethical behavior and less corruption, establish a code of ethics or make the relevant parties sign a formal acknowledgement that they understand their responsibilities and will adhere to ethical and legal standards of conduct.

On the other hand, the relationship between pre- and post-factum mechanisms gets turned on its head, with instances of post-factum accountability being used in support of (enhancing the influence of) pre-factum mechanisms. Consider a widely known practice of tax collection agencies. By selectively pursuing and investing in the prosecution of one or two high visibility cases of tax evasion out of thousands of potential cases, the agency intends to enhance the effectiveness of rules and regulations (pre factum) rather than seek justice (post factum).

This situation is at the center of what we term the "promises of accountability problem."[7] What has taken place as a result of this decoupling from and subordination of post-factum accountability is that the pre-factum mechanisms associated with accountability have emerged as tools that promise to deliver accountable behavior—and thus to directly solve—the problems of control, legitimacy, and performance. Want greater efficiency? Use reporting and similar mechanisms to control the production process. Seek to reduce the democratic deficiencies of an agency? Require greater transparency. Wish to improve the outcomes of agency efforts? Impose

7 For an elaboration of the promises perspective, see Dubnick and Frederickson (2011).

This situation is at the center of what we term the "promises of accountability problem." What has taken place as a result of this decoupling from and subordination of post-factum accountability is that the pre-factum mechanisms associated with accountability have emerged as tools that promise to deliver accountable behavior—and thus to directly solve—the problems of control, legitimacy, and performance. Want greater efficiency? Use reporting and similar mechanisms to control the production process. Seek to reduce the democratic deficiencies of an agency? Require greater transparency. Wish to improve the outcomes of agency efforts? Impose a system of performance measurement. These are the promises of accountability—or, to be more precise, the promises of pre-factum accountability.

a system of performance measurement. These are the promises of accountability—or, to be more precise, the promises of pre-factum accountability.

Solving the problems of efficient control, democratic legitimacy, and effective performance are without doubt worthy aspirations, and efforts to do so ought to be encouraged. And there is little doubt as well that accountability has and can play an important role in doing so. But what we have learned over the past decades of intense accountability-based reform is that the relationship between pre-factum account-giving mechanisms and each of those objectives is neither simple nor easily implemented.

In the following sections we provide a brief overview of the three problems and how they have been addressed in the past. We also consider how the promises of pre-factum accountability emerged in each area. This is followed by a more intensive examination of how promise of performance has become so prominent, and then we turn our attention to how all this relates to the emergence of the Extended State.

The Problem of Control

In the traditional practice of public administration it is understood that specialization, merit appointment, coordination, hierarchy, standardized procedures, record keeping, accounting, policy directives, and the other elements of bureaucratic control result in both individual and group accountability to agency and program goals. This is the bureaucratic model of accountability, the form of accountability with which we are all familiar. Indeed everyone who is part of a public agency or private company is part of bureaucracy, kept accountable by the policies and procedures of organization. Despite the empirical challenges to this logic on the part of those describing organizational goal displacement, agency capture, and other bureau pathologies, it is certainly the case that the dominant bureaucratic form of accountability in governance was—and continues to be—the preferred means for asserting and maintaining control over the human factor in the machinery of government.

It is notable that what we typically term the "Weberian" bureaucratic model of accountability was not the design of sociologist Max Weber, but rather his articulation of an "ideal type" he saw emerging during the modern era. Each part of the model—including its explicit and implied accountability mechanisms—developed in response to some real or perceived threat to the effectiveness and efficiency of the increasingly important modern organization. For example, the stress on meritorious appointments and promotions was meant to guard against threat to efficiency posed by patronage related to corruption. Similarly, the depersonalization of roles and hierarchical arrangements of positions was developed in response to the threats posed by undisciplined and rogue behavior by the organization's members. In short, these features and the accountability arrangements they generated (e.g., job descriptions, written records) were pre-factum responses designed to preclude the need for post-factum actions (e.g., sanctions, firings).

The dark side of these pre-factum arrangements has not gone unnoticed, especially as they related to efforts to enclose workers within what Weber called the "iron cage" of bureaucratic rationality. Franz Kafka, a German novelist, wrote so effectively of the dehumanizing implications of a bureaucratized life that we often use the term *kafkaesque* to describe that condition. More recently, public administration scholars, such as Ralph Hummel (2008), Guy B. Adams and Danny Balfour (2004), and others, have stressed that dark view. On the larger stage of modern philosophy, Hannah Arendt (1976) wrote critically of the banality of evil and thoughtlessness generated by bureaucratic orders, and Michel Foucault (1991) spoke of the cold and calculating "governmentality" of contemporary governance.

Despite these criticisms, pre-factum bureaucratic accountability arrangements have proven so effective that their connection to the post-factum forms was soon lost. Instead, the order, predictability, and reliability of government derived from bureaucratic accountability were associated with a much simpler logic based on a general sense that these account-giving

Pre-factum bureaucratic accountability arrangements have proven so effective that their connection to the post-factum forms was soon lost. Instead, the order, predictability, and reliability of government derived from bureaucratic accountability were associated with a much simpler logic based on a general sense that these account-giving mechanisms by themselves hold the promise of control.

mechanisms by themselves hold the promise of control. It was a logic not unlike that of the classic Pavlovian stimulus-response: impose before-the-fact forms of accountability and the result will be compliant behavior and control over the human factor in the organization's operations. And it is through such control that the objectives of efficiency and rationality will be attained.

In one of its most popular and positive forms, this promise-based pre-factum approach is found in widely accepted motivational theories, such as Douglas McGregor's Theory X/Theory Y (McGregor 1960). It is a matter of selecting the right motivational tools once you've established whether you are dealing with lazy *X*s or self-actualizing *Y*s. Activate the right tool (typically one related to account giving) and responsiveness to the needs of the organization will follow.

Of course, while the promise of control through pre-factum accountability is a popular one, the reality is that few organizations have dropped the traditional mechanisms (that is, those linked to post-factum functions) entirely. We all know of the instances when the promises of organizational and managerial control through pre-factum accountability fail or prove insufficient. In those instances the subject changes from the routines and protocols of before-the-fact accountability to the processes of traditional post-factum accountability. An agency that seeks to foster innovation or greater productivity through a pulling back of burdensome oversight and red tape (e.g., the expansion of telework arrangements for public employees) is not likely to do so without providing at least subtle reminders that there are obligations and legal parameters tied to increased discretion over one's activities. An organization where experiments with pre-factum schemes have resulted in lower productivity or other negative results may drop them altogether and return to accountability approaches that explicitly reconnect behavior to demands for post-factum account giving.

We can see another dimension of the promise of control at work in the use of accountability to promote ethical choices and appropriate behavior among agency workers. Accountability as a word and a concept increasingly implies the promise of ethical behavior of public officials, or at least a reduction in corrupt behavior, ordinarily through a variety of procedural mechanisms, such as reporting and disclosure requirements for elected and appointed public officials.

Traditionally, professional and jurisdictional codes of ethics and oaths of office were intended to encourage ethical behavior by implying that consequences will follow (post factum) violations of those standards. Similarly, it could be argued that the almost universal adoption of merit-based civil service systems by American governments at all levels during the Progressive Era was an accountability reform linked to the desire for ethical (that is, less corrupt) behavior in the appointment of competent personnel to operate government agencies. Accountability in the Progressive Era was understood to be based on a broad consensus that merit systems would do more than bring competence and expertise to government—it would also reduce corruption on the part of public officials and thereby make government more honest and therefore more accountable to the public (Frederickson, Johnson, and Wood 2004).

In the last several decades, things have changed as the impetus to use the logic of a merit-based civil service to reduce corruption has weakened. Merit-based selection and promotion of civil servants has, in many jurisdictions, and particularly the federal government, been watered down as new layers of politically appointed officials have been inserted above (and in lieu of) the merit-based civil service. What was once associated with often blatant examples of cronyism and corruption has now been rationalized as a necessary means for enhancing executive authority and responsibility. This too has been based on a form of the promise of control through accountability, defining accountability as the claim of responsiveness to political principles and the agenda of particularly elected executives (president, governors, and mayors). Under the promise of control, such a claim of responsiveness is sufficient; demand a pledge of responsiveness and control will follow. The erosion of the civil service based on this promise-based pre-factum logic—in addition to the widespread practice of contracting out—have almost certainly weakened the role of post-factum accountability through bureaucratic control (Aberbach and Rockman 2000).

Of the three promises of accountability, the promise of organization and managerial control has the strongest historical claim. Despite complaints about "the bureaucracy," the evidence is that public organizations do what their enabling legislation calls for them to do, given the limits

The promise of organization and managerial control has the strongest historical claim. Despite complaints about "the bureaucracy," the evidence is that public organizations do what their enabling legislation calls for them to do, given the limits of time and money. Furthermore, public agencies are seldom "out of control."

of time and money (Wilson 1989). Furthermore, public agencies are seldom "out of control," generally responding to the preferences of elected officials (West 1995). In short, public agencies are, generally speaking, accountable to the law and to their political leaders. This is particularly the case when accountability is understood to be systems of rules and policies, organizational arrangements, and managerial leadership that cause or result in accountability before the fact. The problem, in many cases, is not the absence of pre-factum bureaucratic accountability to public purposes; the problem is the assumption that any or all forms of pre-factum accountability are sufficient by themselves to result in appropriate and optimum levels of control.

Consider the story behind "red tape." By definition (Bozeman 1993), *red tape* represents those burdensome "rules, regulations and procedures" that remain in effect despite the fact that they "have no efficacy for the rules' functional object" (1993, 283). Whatever their original rationale or purpose, they stand as useless—and often dysfunctional—legacies to past organizational requirements or needs generated for the sake of control. A past case of fraud or embezzlement might result in a new procedure or rule (a pre-factum mechanism) designed to preclude the reoccurrence of the episode. While such a reoccurrence proves increasingly unlikely due to changes in circumstances, the pre-factum mechanism remains in place, having become institutionalized and perhaps now supported by a belief that its maintenance remains critical to the management-control system that had developed in and around that rule over time.

This was understood clearly by 16 former senior government executives holding the rank of assistant secretary, serving as a panel of the National Academy of Public Administration, who issued a report making this point:

> Over many years, government has become entwined in elaborate management control systems and the accretion of progressively more detailed administrative procedures. This development has not produced superior management. Instead, it has produced managerial overburden. . . . Procedures overwhelm substance. Organizations become

> discredited, along with their employees. . . . The critical elements of leadership in management appear to wither in the face of a preoccupation with process. The tools are endlessly "perfected"; the manager who is expected to use these tools believes himself to be ignored. . . . Management systems are not management. . . . The attitude of those who design and administer the rules . . . must be reoriented from a "control mentality" to one of "how can I help get the mission of this agency accomplished" (NAPA 1983).

Upon coming to the same point, the astute James Colvord, a long-time senior civil servant, said that the government needs to have "a bias toward action, small staffs, and a high level of delegation which is based on trust" (Wilson 1989, 369). Although not always in the name of accountability, repeated cycles of organizational and managerial reform or, particularly in the federal government, have lessened the likelihood of effective accountability. Describing this paradox, James Q. Wilson suggests that "to do better we have to deregulate the government" in the same way we deregulate the private sector (Wilson 1989, 369). In order to make government agencies more accountable, we need to clear the underbrush of red tape created by earlier iterations of accountability means and mechanisms that had lost their functionality.

What would a deregulated bureaucracy have looked like (DiIulio 1994)? Under one scenario, left to its own devices a deregulated administrative state might have developed into some nightmarish vision of power-abusing public officials, wielding their unchecked authority to fulfill or enhance their own self-interests. Under a more likely scenario, dedicated public servants, ever mindful of the need for democratic legitimacy and demands for performance, would use their managerial and organizational skills and authority to dispose of unjustifiable red tape while reshaping the pre-factum accountability environment.

A third scenario, of course, was to abandon the direct provision of public goods and services through various forms of outsourcing—a path that will (and has) subverted the Administrative State and put in its place the Extended State. Short of the complete withdrawal of government from involvement through a radical form of privatization, this approach must still rely on accountability mechanisms to ensure that those operating on behalf of the State are doing so in a way that meets the three objectives of control, legitimacy, and performance.

The Problem of Democratic Legitimacy

The relationship between accountability and democracy is widely accepted, to the point that we commonly equate democracy with accountable governance—and vice versa.[8] In that sense, accountability does promise democratic government, particularly as it establishes the answerability and responsiveness of public officials and a set of effective procedural guarantees.

Nothing is more fundamental to representative democratic government than is political accountability to citizens. Constitutionally established principles, such as regular and fair elections, the rule of law, majority rule, and guaranteed minority rights, are designed to achieve democratic accountability. In theory, elected officials are held accountable each election day, and electoral mechanisms are more often than not perceived as the perfect set of accountability mechanisms within our pre/post-factum framework. Assuming an incumbent is running, the election itself is an opportunity to hold that individual (or party) to account. Where there is no incumbent, the election can be perceived as reflecting a number of messages, from judging the campaign promises made to a mandate—or even a warning—from the electorate. The very existence of an election acts on the pre-factum side of the approach, making the officeholder or potential candidate behave in ways that can be regarded as responsive or accountable to the wishes of the electorate that she or he will face. The problem with this set of mechanisms, of course, is that electoral accountability processes are intermittent and distinctly rigged in favor of incumbents,[9] which considerably weakens their capacity to really enhance democratic accountability.

Other electoral forms, particularly initiative, referendum, and recall, are likewise understood to be means by which the people can directly circumvent representative government and either directly legislate or remove officials from office. Finally, the U.S. Constitution as well as the 50 state constitutions, and particularly the Bill of Rights, include firmly established norms to which public officials—both elected and administrative—are to be held accountable (Rosenbloom 1983; Rohr 1986).

8 For a recent example, see Fukuyama (2011).
9 Recent studies include Stonecash (2008) and Stone et al. (2010).

The evidence indicates that American citizens have a generalized dissatisfaction with government, including feelings of exclusion and powerlessness (Mathews 1999; Harwood and Creighton 2008). It was Richard Harwood who provided evidence that Americans are not apathetic but are mad "at a political system that pushed them out of their rightful place of governing the nation" (Mathews 1999, 11). Coming in for the strongest criticism in the Harwood study was government controlled by a political class, by lobbyists, by powerful interest groups, by too much money, by a biased and negative media, and by costly and retrograde electoral campaigns. In sum, many citizens do not believe that government is accountable to them. Citizens want both good government *and* good politics.

From an accountability perspective, the modern American concept of politics is too narrow. The politics of elections and law and budget making are necessary, but they are not sufficient to create or sustain an accountable democratic government. A wider politics is typically required, including the politics of public deliberation; the politics of public engagement in neighborhoods, schools, civic clubs, and organizations; and the politics of reframing the political agenda are widely practiced but less well understood than electoral politics.

From an accountability perspective, the modern American concept of politics is too narrow. The politics of elections and law and budget making are necessary, but they are not sufficient to create or sustain an accountable democratic government. A wider politics is typically required, including the politics of public deliberation; the politics of public engagement in neighborhoods, schools, civic clubs, and organizations; and the politics of reframing the political agenda are widely practiced but less well understood than electoral politics. There is both a growing awareness of the emergence of this wider understanding of politics, as well as a growing literature that describes it. This understanding of politics comes in several forms. Perhaps best known are the communitarians "back to basics" movement that emphasizes families, schooling, and communities and calls for a louder moral voice. That voice calls for citizens taking responsibility as an alternative to claiming rights (Etzioni 1993; Schorr 1997). Another variant in the wider understanding of politics is the social capital perspective associated with Robert D. Putnam and the work of the Saguaro Seminars. Putnam describes the decline of community-based intermediate institutions as well as a decline in political participation. He and many others call for a reemergent agenda for social capitalists—citizens trained and organized to be politically engaged (Putnam 2000). A third variant in the wider understanding of politics is the citizenship and civic-engagement advocates (Skocpol 2003; Skocpol and Fiorina 1999; Schudson 1998; Mathews 1999). Perhaps most notable in this variant is the work of National Issues Forums, a national network of locally based groups and organizations engaged in deliberation on the most vexing policy issues of the day.

There are far more similarities than there are differences between these approaches to a wider understanding of politics. All are local or grass roots. All are deliberative. All assume the legitimacy of direct forms of citizen engagement, which is to say that American government is not the sole province of elected legislators and executives. All assume that the key to an engaged public is to be found in a well-informed electorate and that deliberation is the key to informing the public (Adsera, Boix, and Payne 2003). David Mathews makes this incisive point:

> On reflection, it seems to us that the accountability movement is powerful because it serves its own kind of politics, which is as entrenched as the movement. If that is so, nothing will change until there is some reconsideration of the dominant mode of politics. While the accountability movement would likely claim to be serving democracy, we would like to know what kind of democracy; specifically, we would like to know what concept of citizenship is implied in accountability practices. Our system tends to sideline citizens as occasional voters and consumers of political fare but not recognizing them as people capable of producing public goods by their collective efforts. So, if the prevailing concept of who citizens are (or should be) is the central issue, we reason, any corrective has to begin there.
>
> More and more, we suspect that there are significant differences in the way citizens understand accountability and the way officeholders and institutional officials (governmental and nongovernmental) understand it. We don't think one understanding is necessarily better than another but that the differences may be responsible for much of the frustration of both institutional leaders and "the people." *A very tentative hypothesis is that institutions think of accountability primarily in informational terms while citizens think of it more in relational terms* (Mathews 2011, ix).

This wider understanding of politics includes the direct voices of the public in all their diversity. These voices call for direct accountability to the people on the part of those whom they have elected, and this accountability is far more than merely being elected or reelected.

Rather than the accountability of elected officials to the people, a second form of political accountability involves those to whom the daily work of government has been delegated by the policies and decisions of elected principals—that is, public administration. Earlier in this monograph we described accountability patterns found in the organization and management of government, and we claimed that these patterns evidenced a generalized pre-factum accountability committed to the objective of providing efficient and orderly governance. Here we stress the second objective—a commitment to democracy, to the law, to elected officials, and to the public. Depending on the level of government and on the policy field, there are wide variations in the effectiveness of government programs, variations based on the intrinsic tractability of the policy challenge ranging from the operation of Social Security to the operation of schools or prisons. Accountable public administration turns on policy tractability, the availability of resources, and goal clarity. But it also turns on the legitimacy of its processes in the eyes of those who put their trust in government agents and agences.

The large and growing outsourcing of the day-to-day implementation of public policy is a fundamental challenge to the accountability of public agencies in their efforts to achieve and sustain democratic legitimacy. It is unlikely that so-called third party government or the hollow state (what we discuss below as the Extended State) can be rendered as accountable either to elected officials, to the public, or to the law as in traditional public-sector agencies. The delegation of authority through contracts or grants may include provisions that can simulate superficial forms of legitimate behavior, but democratic legitimacy is not easily passed on to third parties or grantees, even if they are themselves subject to legitimacy pressures (e.g., in the United States, states that act as agents of federal requirements in their implementation of federally funded programs).

The problem posed by the Extended State is a magnified version of the issues facing any form of authority delegation in the public sector. Viewing the problem of delegation of authority in the more general context of the European regulatory state, Giandomenico Majone concludes, "it is up to the political principals [elected officials] to structure relationships with their agents [regulatory agencies] so that the outcomes of the agents' efforts comply with democratic accountability, given the choice to delegate in the first place" (Majone 1999, 13). The extension of state functions to third parties and grantees was regarded, in the not too distant past, as a fundamental violation of basic constitutional principles in the United States, but over time practice has overtaken principle and it is now unchallenged if not formally justifiable (see Lowi 1969). Applying pre-factum accountability mechanisms, such as those suggested by Majone to address complaints of "democratic deficits," can help, but only if linked to post-factum consequences. Too often, however, there is little or no such linkage. Merely requiring transparency[10] and other superficial mechanisms of account giving will not suffice, a point Majone makes when he points out that "in the final analysis" their legitimacy "depends on their capacity to engender and maintain the belief that they are the most appropriate ones for the functions entrusted to them" (Majone 1999, 22).

Problems of Performance Measurement as Accountability

The third problem accountability can address—effective performance—is achievable through assessments of performance and the application of performance standards. This is the common mantra among most modern administrative reformers (Radin 2006). Accountability through performance measurement, performance management, and performance budgeting is the fashion of the day, a hegemony seldom subject to question. The assumed linkage between accountability and performance is so powerful, in fact, that the two are used as indicators of each other: to be accountable is to live up to expected performance and to be performing up to (or in excess of) standards is a clear sign of being accountable.

The role of performance measures was initially applied to problems of managerial control. The idea of using performance assessments and measurement as the basis for improving the management and delivery of public services is traceable at least to the Progressive Era, and especially to the work of municipal research bureaus (Williams 2003) and related organizations. The emphasis at that time was on gathering a wide range of data that would be used as social indicators of existing conditions and applied in early forms of "needs assessment" that would guide choices of emerging social service agencies and charities. For those concerned with the corruption and inefficiencies of government, those indicators became the basis for advocating civil service and general administrative reforms, leading to more specific measures of agency and program performance. Measures of outcomes were a logical next step since the same numbers

10 On transparency as a governance mechanism, see Hood and Heald (2006); Fung, Graham, and Weil (2007); also Piotrowski (2007).

could be employed to assess the success of specific programs (e.g., the performance of a sewer system is measured by changes in the mortality rates over time).

What Williams (2003) uncovers in his detailed exploration is that those early efforts were driven by mixed motives, reflected in the three types of measurement that emerged at the time: needs assessment, outcome measurement, and productivity measurement. Needs assessment was somewhat relevant to performance in providing information on current conditions that could be used as a basis for later assessments. Productivity measures served the needs of those who wanted to uncover corruption and inefficiencies that can be subjected to the wide range of progressive reforms. Outcome measures of performance were of secondary importance in those early years, but played a role nonetheless in bringing about reforms. The roots of performance measurement, in short, are found in the general commitment of Progressive Era reformers to understand and deal with the major social and political problems of their day.

As the energy of the progressive reform movement dissipated, the further development of performance measures and monitoring systems shifted to the private sector where scientific management approaches and its progeny were flourishing. The technical advances made in performance measurement were generally aimed at enhancing productivity rather than outcomes—and any concern with the performance of individuals was typically related back to questions of how to improve overall cooperation in the drive for greater productivity.

Performance measurement reemerged in the public sector during the 1960s as part of systems thinking and evaluation research,[11] but here again there is little explicit connection to issues of accountability, especially at the individual level. Management-by-objectives (MBO)—promulgated as an idea by Peter Drucker (1954)[12] and eventually endorsed (albeit briefly) by the Nixon administration[13]—provided the connection, however, with its basic assumption that employees who played a role in setting the goals and objectives of the organization would be both motivated and willing to be held to account for their achievement. By 1980, in both the United States and the United Kingdom, performance measurement had come into its own as a major tool

11 See the classic (and quite devastating) critique of the systems approach in Hoos (1972); also Hoos (1973).
12 Drucker (1976) points out that MBO has deeper roots than his own articulation in the 1950s.
13 The Nixon administration adopted MBO in 1973, but by 1976 it was essentially regarded as a failure (see Sherwood and Page 1976; Newland 1976; Rose 1977); however, in 1995 Poister and Streib reported that it was alive and well in a number of surveyed municipalities.

of governance in the Administrative State, and it played an even more significant role in other nations (e.g., New Zealand, Canada) as the NPM movement spread.

Assessments and critiques of performance measurement and its associated management technologies are easy to find, and it would be absurd to offer a comprehensive overview of the various comments and observations. But from our perspective, there have been two significant and related developments. First and foremost, performance measurement increasingly became a management tool reflecting a belief[14] in what we term accountability's "promise of performance" (also see Ammons 2002)—an unsubstantiated and articulated assumption that the pre-factum imposition of performance standards and goals is sufficient to produce demanded or appropriate outcomes through accountable behavior on the part of an organization and its members. This ran counter to all forms of performance-management approaches that preceded the promise, even MBO which required participation and commitment in the setting of objectives by those who would be held to account.

Second, and perhaps most important, accountability's promise of performance has become politicized to the point of putting major decisions about the design and implementation of performance-based management into the hands of politically driven policymakers who use the call for greater accountability (i.e., greater performance) as a means of mobilizing electoral support. There is perhaps no better example than the No Child Left Behind Act (NCLB), originally passed in 2002 and up for renewal at the time of this writing. Although this high-stakes form of education policy might, in some earlier iteration, have been based on a well thought through logic tied to a substantial theory of how to enhance student performance, such a theory—other than the promise of performance—has been lacking since passage of the act.

NCLB also stands in as well as an example of the accountability challenges posed by the Extended State. While the U.S. government lacks explicit constitutional authority to set policies for K-12 education in states and localities, there is little doubt that both Republican and Democratic administrations in Washington have education as a high priority. In addition, NCLB stands out as an example of a policy born of enthusiastic bipartisan support in Congress as well. And yet it is

14 Kelly (2002) notes that there is little actual evidence or facts to support that belief, so we most take it "on faith."

a policy that requires that federal administrators make use of nonfederal parties to carry out its program objectives. How does the U.S. Department of Education go about ensuring accountable managerial control, democratic legitimacy, and effective performance in the implementation of its programs? And how does it do so without relying (as it does now) on simplistic assumptions about the relationship between accountability and performance? These are issues to which we will return later in this monograph.

Although NCLB represents the most visible and, at present, the most controversial of efforts to fulfill accountability's promise of performance, perhaps the most significant reform in this regard is found in the Government Performance and Results Act of 1993 (GPRA) and its progeny: the Bush administration Program Assessment Rating Tool (PART) and the GPRA Modernization Act (GPRAMA).

The text of the 1993 GPRA legislation identifies its purposes as improving the following: the public's confidence in government, program effectiveness, customer service, accountability, service delivery, congressional decision making, and the internal management of the federal government. It seeks to address the perceived inefficiency and ineffectiveness of federal agencies, which it attributes to poorly defined agency missions, goals, and objectives and to inadequate performance information. Not only do these conditions breed poor performance, advocates for GPRA passage claimed, they also lead to the public's low confidence in government and hamper congressional decision making.

As a remedy for inefficiency and ineffectiveness, GPRA required that federal agencies: (1) establish strategic plans that provide broad descriptions of agency goals and objectives covering a period of three to five years; and (2) develop annual performance plans on the basis of measures, preferably quantified, by which agencies determine the extent to which goals and objectives are being met and report annually on agency performance based on these measures.[15]

Put in perspective of the three problems discussed above, GPRA can be regarded as a congressional mandate for federal agencies to shift their focus from organizational and managerial

15 For perspectives on the 1993 act as it was being implemented, see Kravchuck & Schack (1996) and Radin (1998).

control to performance. GAO, in its 1996 "Executive Guide" for implementing GPRA, stated the implications of the act clearly, noting:

> Every major federal agency must now ask itself some basic questions: What is our mission? What are our goals and how will we achieve them? How can we measure our performance? How will we use that information to make improvements? GPRA forces a shift in the focus of federal agencies—away from such traditional concerns as staffing and activity levels and toward a single overriding issue: results. GPRA requires agencies to set goals, measure performance, and report on their accomplishments. This will not be easy. (U.S. General Accounting Office 1996, 1)

Although GPRA was still relatively new and not yet fully implemented when the Bush administration took office, in the early stages of the Bush presidency a rather different approach to performance measurement was developed through implementation guidance in the form of a PART developed by the Office of Management and Budget (see Gilmour 2007).[16] At the direction of the White House, OMB assembled a team to reform or augment GPRA, and in July 2002, OMB introduced PART, a large questionnaire designed to be answered by agencies. PART included 25 to 30 questions grouped into the following four categories: program purpose and design; strategic planning; program management; and program results/accountability. The first question in each category was as follows:

- Is the program purpose clear?
- Does the program have a limited number of specific long-range performance measures that focus on outcomes and meaningfully reflect the purpose of the program?
- Does the agency regularly collect timely and credible performance information, including information from key program partners, and use it to manage the program and improve performance?
- Has the program demonstrated adequate progress in achieving its long-term performance goals?

On the basis of agency answers to these and other questions, OMB assigns either a yes/no response or a response of yes/large extent/small extent/no response. Each answer is to be accompanied by an explanation.

16 Also see Gilmour & Lewis (2006) and Breul (2007) for more on the Bush administration's effort to assert managerial control.

The relationship between GPRA and PART proved to be problematic (GAO 2005). While GPRA reflected accountability's promise of performance, PART was clearly focused on the use of pre-factum mechanisms to deal with the problem of managerial (i.e., White House) control. The GPRA statute provides a framework under which agencies prepare strategic plans, performance plans (now, performance budgets), and performance reports that set goals and annual targets for those goals and report on the extent to which they are achieved. PART was a systemic method of assessing performance of program activities from the perspective of the White House, focusing on their contribution to an agency's achievement of the administration's strategic and program performance goals as expressed in the president's budget.

On the basis of agency responses to the PART questionnaire, OMB assigned a score to each question and an aggregate score for each program assessed through PART. The aggregate score was based on the application of different weights to each of PART's four sections, with the greatest weight placed on the program's performance results. After assembling the results of agency responses to the PART questionnaire, OMB assigns a rating, based primarily on the score, to each program. Programs can be rated "effective," "moderately effective," "adequate," "ineffective," or "results not demonstrated." The PART responses and ratings are published in the subsequent year's submission of the president's budget. For example, PART assessments conducted in 2004 were published the following January with the release of the president's fiscal year 2006 budget. In the original version of PART the summed scores assigned to each program or agency were described as "the percent accountable."

PART was not continued by the Barack Obama administration, and in its place the White House undertook an Accountable Government Initiative (White House 2010) that focused almost exclusively on issues related to problems of managerial control.

> As part of its Accountable Government Initiative, the Administration has moved to cut programs that don't work and other wasteful spending, streamline what does work, modernize how government operates to save money and improve performance, and make government more open and responsive to the needs of the American people.

That initiative was part of a more general White House call for "High-Performance Government," and it followed the guidance OMB gave to agencies in June 2009 for developing the FY2011 budget. In that guidance, the Obama's OMB had effectively abandoned the PART approach and its "performance rating" approach by deferring to individual agency priorities. In a memo, then OMB director Peter R. Orszag "asked agency heads to identify a limited number of high-priority goals and begin developing strategies to address them." Commenting on this move, John Kamensky called it "unusual," but "praised Orszag for learning from past administrations that agencies, not the White House or OMB, should set priorities" (Newell 2009).

This was followed by a potentially even more significant development. During its lame duck session in late 2010, Congress passed the GPRA Modernization Act (GPRAMA) of 2011 that amended the original GPRA to address some of the tensions that had developed as a result of conflicts between the congressional push for performance and the White House need for mechanisms that gave it greater managerial control. In the GAO's assessment of the new legislation (Dodaro 2011), it highlighted provisions that would result in

- a more coordinated and cross-cutting approach to achieving meaningful results,
- efforts to address weaknesses in major management functions,
- ensuring performance information is both useful and used in decision making,
- sustained leadership commitment and accountability for achieving results, and
- engaging Congress in identifying management and performance issues to address.

Despite all the qualifications and adjustments that have occurred over the past 18 years, what the history of GPRA, PART, and GPRAMA represents is the elevation of accountability's promise of performance to the driving force behind the American approach to governance. Performance measurement, in short, has become central to the institutional infrastructure of government operations. In the pantheon of various efforts at measurement or assessment-based reforms, from cost-benefit analysis to PPBS and MBO to ZBB, performance measurement has been among the most durable, primarily (we believe) because it is linked to accountability's promise of performance. The question is whether the institutionalization of the promise of performance is a positive development, or an obstacle to effective governance. With the accountability-based promise of performance playing such an important role in the

Despite all the qualifications and adjustments that have occurred over the past 18 years, what the history of GPRA, PART, and GPRAMA represents is the elevation of accountability's promise of performance to the driving force behind the American approach to governance. Performance measurement, in short, has become central to the institutional infrastructure of government operations.

administration of federal programs, the question necessarily arises as to whether such an approach is indeed effective in improving the performance of government agencies and agents. What is most inviting about accountability's promise of performance is that it makes logical and conventional sense to those who operate in a world reflecting the characteristics of an ideal-type Weberian bureaucracy operating within the confines of the modern Administrative State. The problem is that the reality of today's governance is far removed from those conditions.

Some Issues Surrounding Accountability's Promise of Performance

Accountability, it is argued, often comes with an accountability paradox. Beryl Radin (2006) claims that there is an inherent tension between accountability and performance, and efforts to improve the performance of public agencies through accountability mechanisms tend to have the opposite effect. In the accountability paradox, agencies are held to account for how well they implement formal accountability processes and paperwork procedures rather than for how well they actually perform their primary purposes. The accountability paradox argument also points to high transaction costs: expenses of time and resources devoted to account giving are resources that could have been used to improve performance (Frederickson and Frederickson 2006).

A more important challenge to the presumed connection between greater accountability and enhanced performance comes from Jennifer Lerner and Philip E. Tetlock. Their review of research findings regarding how individuals react to account-giving situations is sobering and worth quoting:

> This review underscored the falsity of the conventional wisdom—often born of frustration at irrational, insensitive, or lazy decision makers—that accountability is a cognitive or social panacea: 'All we have to do is hold the rascals accountable.' Two decades of research now reveals that (a) only highly specialized subtypes of accountability lead to increased cognitive effort; (b) most cognitive effort is not inherently beneficial; it sometimes makes matters worse; (c) there is ambiguity and room for reasonable disagreement over what should be considered worse or better judgment when we place cognition in an institutional context. In short, accountability is a logically complex construct that interacts with characteristics of decision makers and properties of the task environment to produce an array of effects—only some of which are beneficial (Lerner and Tetlock 1999).

This is high academic jargon for Lerner and Tetlock's basic claim—greater accountability seldom results in improved organizational performance. Systems of accountability are essential to the effective oversight of public administration. So are systems of performance measurement and evaluation. But they are not the same thing. The assumed relationship between accountability and performance may be conventional wisdom, but it is an untested assertion. We already know from the Hawthorne studies that attention and trust induce performance improvement. Can the same be said for accountability systems based on performance measurement (Dubnick 2005)?

There is little question that the potential of thoughtful and well-crafted performance-measurement-based accountability principles can contribute to the quality of governance and, thereby, to improvement of prospects that the public sector can achieve its purposes. Instead, we attempt here to answer this question: How can performance measurement be better understood as accountability?

There is a considerable body of useful accountability research and theory. Performance measurement has become shorthand for precision and objectivity in the definitions of goals and purposes. Precise quantitative measures, so called metrics, are the means by which levels of goal achievement can be determined. In this evaluation, the emphasis is on the application of modern accountability principles in the public rather than the corporate world because, as Harry Boyte puts it, "public work" is a world of changing values, ambiguous purposes, and elevated expectations (Boyte and Kari 1996).

Table 2 is a simplified template of the theory and practice of performance measurement as a primary means of accountability. To describe the contemporary use of performance-based accountability claims, we employ a "theory versus practice" heuristic to aid in the discussion of the effectiveness of performance measurement as accountability. The rhetoric and theoretical premises of performance measurement as accountability are set out in the center column of Table 2. The actual practices of performance measurement as accountability are set out in the right-hand column. Put another way, the center column lists the theoretical claims of performance measurement as accountability, the right-hand column lists the evidence.[17]

17 Table 2 synthesizes and summarizes an extensive literature on performance measurement. That literature can be roughly divided into: (a) performance measurement promotion and advocacy using some case-study example and (b) empirically based critiques of performance measurement (Ingraham 2007; Frederickson and Frederickson 2006; Radin 2006; Moynihan 2006; Forsythe 2001; Kamensky and Morales 2005; Kearny and Berman 1999; Berman 1998; Hatry 2006).

	Accountability	In Theory	In Practice
I	Measures of performance and results:	answers facts objective-neutral policy guides long run	questions interpretations slanted weapons, argumentst short run
II	What to measure:	the important results outcomes	the measurable surrogates of results processes
III	Performance domain:	comparable	particular
IV	Assertions of causality can be:	demonstrated determined	are unclear indeterminate
V	Performance measures show that programs often:	perform poorly	perform poorly, but things would be worse without them
VI	Policy is influenced by:	measures of performance and results	rhetoric, changing values, winning coalitions
VII	The propensity to innovate is:	a. enhanced by measuring performance and results b. manageable	enhanced by risk taking which deemphasizes performance measurement subject to gaming responses
VIII	Performance measurement results in:	a. strategic differentiation b. the diffusion of best practices	a. isomorphism b. reflected status

TABLE 2. Performance Measurement As Accountability in Theory and Practice

In theory, measures of performance are thought to be the factual answers to questions about organizational effectiveness. In practice, performance measures are themselves questions, or at least they raise as many questions as they answer. Performance measures are exactly that, quantitative representations of some reality. Because performance measures are presented quantitatively, they may have the appearance of fact and may convey impressions of objectivity and neutrality, but in practice such measures are quantitative interpretations of reality in the same way words are narrative interpretations of reality. In practice, very little is judged to be neutral and objective; all performance measures may be used as arguments and weapons in policy debates. If one supports a program and performance measures appear to indicate that the program is doing well, then those performance

In practice, performance measures are themselves questions, or at least they raise as many questions as they answer. Performance measures are exactly that, quantitative representations of some reality. Because performance measures are presented quantitatively, they may have the appearance of fact and may convey impressions of objectivity and neutrality, but in practice such measures are quantitative interpretations of reality in the same way words are narrative interpretations of reality. In practice, very little is judged to be neutral and objective; all performance measures may be used as arguments and weapons in policy debates.

measures are seen to be objective, factual, and neutral determinations of program results. The same performance measures will be judged to be slanted interpretations of questionable data when an individual or a group opposing a policy or program views them.

Measures of performance, if based upon reliable data and kept up to date, may tell the organization in a general way how it is doing. As guides for incremental program adaption, reliable performance data are helpful. But such measures seldom tell an organization what to do.

It is in the nature of organizations to be advocates, to represent particular values and beliefs. Should such organizations not be expected to design strategic plans and attendant performance measures so as to demonstrate success? Is it reasonable to imagine, particularly in the short run, that organizational executives will indicate that the investment of scarce organizational resources is being wasted?

For these and other reasons described below, performance measures are best understood as information that may help sharpen questions rather than the answers to questions. Such measures, particularly in fields with illusive bottom lines, are best thought of as clues, interpretations, impressions, and useful input rather than facts. And of course all such interpretations carry with them certain biases, assumptions, and values.

It is useful to recall the wisdom of Derek Bok, in his splendid treatment of performance measurement in government:

> Attempts to measure the effectiveness of American government are necessarily crude. In certain cases, to be sure, one can measure the progress of particular programs with clearly defined goals and even compare the government's record with that of private firms performing similar tasks. But scattered examples of this kind hardly add up to a

> complete picture of the government's performance. And even if the examples could be multiplied, there is no way of aggregating the results into a reliable overall measure of effectiveness (Bok 2001, 20-21).

If we are to be accountable based on measuring performance, how shall we know what to measure? Virtually all the evidence on the public side is to the effect that (1) the measurable drives out the important; (2) it is extremely difficult to measure actual results so analysts turn to surrogates of results; (3) analysts rely on measures of process rather than actual outcomes. Each of these choices has its own rationality. Choosing to measure the measurable rather than the important supports a theoretical point made 50 years ago by Herbert Simon. It is rational, he claimed, for administrators to use tested techniques, affordable techniques, and techniques that are near at hand. Data gathering and analysis is expensive and time consuming, and performance measures go rapidly out of date.

In the language of performance measurement the most common surrogate or proxy is a so-called output matrix, a measure of organizational production. Performance measurement advocates strongly prefer measures of outcomes.

Surrogates are very useful in performance measurement. For example, we may not agree on what constitutes good education because it is next to impossible, particularly in the short term, to measure education. But we can require students to take standardized tests, we can evaluate teachers, and we can measure overall school performance. All are output, proxy, or surrogate measures. So it is with civil investing; we agree that community and citizenship are very important, but measuring them is tough. So analysts measure bowling, participation in intermediate institutions, and public opinion, and when thoughtfully aggregated these measures give us clues, sharpen our questions, and guide incremental adaption of policies and programs (Putnam 2000).

But it is process that is most important of all. Consider justice. Justice is difficult to define. Instead procedural due process is established, and the result of that process is called justice. Consider democracy. Elaborate electoral processes are put in place and the result is called democracy. Over time, reasoned evaluations of justice and democratic processes inform attempts to improve them. It is exactly the same in civil society, community development, and citizenship. The National

Civic League's *Civic Index: Measuring Your Community's Civic Health* (1999) is a process measure of civil society. Many of the findings based on the Social Capital Benchmark Survey of the Saguaro Seminar are process measures. The received wisdom in public policy and administration is that it is much easier to find consensus on a process or a course of action than it is to get people to agree on objectives, particularly objectives that are more specific than bland generalizations like "better education."

In sum, in the public sector it is both rational and helpful to measure what we can, to measure proxies of results, and to measure processes in lieu of outcomes.

Comparative analysis as in the use of benchmarks is a common feature of performance management. There are two difficulties associated with using comparative analysis to determine how well organizations are performing. To begin with, estimating a program's contribution to performance is extremely difficult. The effect of law enforcement on crime is one prominent example; the independent impact of schools on student achievement is another. The second problem is that conditions vary substantially from one city or community to another, one church to another, one university to another. One cannot help but wonder whether variations in the amount of progress made in various fields of endeavor are due to the performance of institutions or to differences in the external conditions under which individual programs operate. For example, if the United States has made less progress than other countries in eradicating poverty, reducing pollution, or achieving higher student test scores, it is at least conceivable that our record reflects such factors as larger size, greater population diversity, heavier defense burdens during the cold war, or the unfortunate legacy of race.

Relatively noncontroversial, unambiguous federal programs, such as the work of the Social Security Administration and the Postal Service, lend themselves to performance measurement, and they tend to do well. But many government programs have little basis for reasonable comparison. The work of diplomacy, national defense, medical research, and justice are not easy to measure against business practices. In the absence of a basis for comparison the question changes from how well is national defense doing? to how well are we doing national defense? Or the question changes from how well are schools doing? to how well are we doing schooling? In the theory of performance measurement this may be disappointing. In practice, however, it has its own rationality.

All of this suggests the frailty of comparison as a useful performance- measurement tool. Despite its weaker logic, measuring the performance of a program all by itself and on its own terms is likely to yield information at least as useful as comparative analysis. Context matters and comparisons should be carefully and cautiously made.

Despite advances in social science methodology it is still very difficult to demonstrate precise causality. This difficulty is particularly notable in public fields like education, law enforcement, or national defense, in which goals are ambiguous or in conflict. Nevertheless, experienced observers and participants know the importance of resources to schooling, to law enforcement, to medical research, and so forth, although it is often difficult or impossible to precisely measure causality. Reason suggests that the same is true in civil society, although descriptions of causality are narrative rather than quantitative and it is difficult if not impossible to *prove* causality. Therefore, in a world of unclear or indeterminate causality, we rely on softer social science—history, participant observation, surveys, interviews, and case studies. When softer social science is carefully done, it conveys a form of evidence that may not be causality but is useful evidence nevertheless.

It is an axiom of public policy and management that the public sector ends up with the nasty problems (Hargrove and Glidewell 1990). As a consequence, many programs designed to ameliorate these problems perform poorly. In the measurement of the results of specific programs it is often impossible to know how much would have been achieved had the program not existed. Inner-city schools are cases in point. So are jails and prisons and the problems of recidivism and rehabilitation. Poor overall performance may have less to do with the effectiveness of public and nonprofit programs and more to do with the intractable nature of many social problems. It is because of the intractable nature of many social problems that well-meaning people and organizations do their best to make things better. To use performance measures to relentlessly remind us that things are bad and don't appear to be getting much better is neither evidence that social programs are failing nor justification for program discontinuation. The application of strict accountability logic to organizations dealing with the toughest social and public problems may be overly accountable to boards and legislatures and not sufficiently accountable to those in need.

One of the great fallacies in public life is that performance measures and similar data are influential in policymaking. In rank order of salience, policy is moved by changing conditions or circumstances, changing values, rhetoric, and winning coalitions.

One of the great fallacies in public life is that performance measures and similar data are influential in policymaking. In rank order of salience, policy is moved by changing conditions or circumstances, changing values, rhetoric, and winning coalitions. Consider the great epochs in American public life—the Civil War and the abolition of slavery, the reform-progressive movement, the Depression and the New Deal, the civil rights movement—each was at some level influenced by information, accurate and inaccurate, but each was really moved by critical circumstances, by changing social values, and by political rhetoric. Only then did winning coalitions change and thereby change policy.

If those who lead public agencies, believe that changing public policy is their purpose, yet do not invest over time in trusted institutions, they run against the wind. But if they do all they can through trusted institutions to achieve agency purposes *for their own sake,* the prospects for changed policy are great. And if the programs of public agencies are only a bit effective, when circumstances change—a war, an epidemic, a recession—strengthened and resilient institutions will be better able to respond to the challenge (Wildavsky 1988).

All this is not to argue that policy does not matter, because it does. "Broken windows" policy, community policing policy, "three strikes and you're out" policy, and many others are obviously important—and controversial. But these policies are moved more by circumstances (or perceived circumstances), changing values, and political rhetoric than they are moved by accountability-driven performance-measurement protocols.

These days it is fashionable to put together the logic of performance measurement and the organizational search for innovation. Management consultants and corporate CEOs are partial to the logic that innovation can be managed, which assumes the compatibility of performance measurement and innovation. There are three major components to the management of innovation logic applied in the private sector: innovation stimulus, the organization's capacity to innovate, and the innovations themselves.[18] The general argument is that innovations in either the production

18 Based on Prajogo and Ahmed (2006).

process or in the product itself will be positively associated with a greater capacity to innovate, which in turn can be stimulated by any number of manageable (and managed) factors. The "theory"—carried over to the public sector—is that by strategically integrating the measurement of product and production process innovations into the management (i.e., innovation stimulus) scheme, both the capacity and motivation to generate further innovation will be enhanced.

On the surface, that logic makes sense given the growing literature on innovation which points to positive relationship between an organization's innovation capacity and real changes (often real advances) in products and production processes. But the devil is in the details of the argument. In both historical (see Johnson 2010) and empirical studies (Christensen 2000; Collins and Porras 1994; Light 1998; Nayak and Ketteringham 1994), the nurturing capacity for innovations is found to be greatest in loosely managed organizations that provide open, collaborative "spaces" that are flexible, loosely coupled, decentralized. Such environments are not conducive to being tightly managed or constantly subjected to high-stakes performance pressures (see Prajogo & Ahmed 2006). Moreover, the kind of innovative and entrepreneurial behavior generated by such overly managed, excessively measured environments may not be of the most desirable sort, and in fact might prove highly dysfunctional and counterproductive.

For an example, we return to the case of the high-stakes performance-measurement regime that characterizes No Child Left Behind. The pressure imposed on schools and teachers under NCLB is intended to do more than make educators work harder; it is also intended to get them to work smarter, and to improve their teaching through the use of new and innovative approaches and techniques that make more effective use of their classroom time. Critiques of the testing regime used in NCLB's push for "accountability" abound, and even were we to put aside the evidence of outright cheating and lying (Gilliom 2010; also Nichols & Berliner 2005), there remain the responses sparked by blame and risk avoidance (Hood 2002), shame and guilt avoidance (Posner & Rasmusen 1999) and even (as suggested in Popham 2005) pain avoidance. These various emotive reactions to pressures for innovative teaching often motivate individual and collective attempts to develop innovative responses with the intent of mitigating the real or potential threat of high-stakes assessments.

Diane Ravitch (2010), in her devastating critique of NCLB, offers an overview of the "gaming" that the policy has generated (2010, 154-161). She calls "plain old-fashioned cheating" the "most reprehensible form of gaming" (2010, 155), but then highlights a variety of "not outright illegal" approaches fostered by the high-stakes accountability system. Principals have strategies to reduce the number of low-performing test-takers, either through control of school admission policies, encouraging them to stay home on test day, or by reclassifying them as "special needs" students who are not eligible to take the exam. They can also enhance the performance of test-takers by certifying them as students in need of "accommodation," thus allowing them access to dictionaries or other forms of support during the testing period. Resources devoted to test preparation increases significantly under the high-stakes regime, and increasingly time and curriculum content are devoted to instruction that clearly fostered a teach-to-the-test approach. Assistance also comes from the school district or state education offices where test content is determined, as is the determination of what constitutes a "passable" score. In New York State, the level of passable performance on various parts of the statewide "Regents Exam" (a score of 65) was "converted" through scoring that made low percentage scores the equivalent of 65. In Cleveland, district policies allowed them to "scrub" from the records the test scores of students who were not continuously enrolled during the school year—a practice that boosted the overall performance of 4th and 10th graders significantly.

Looked at objectively (and somewhat ironically), the lesson of the NCLB case is that the use of performance measures in a high-stakes accountability regime can produce fresh and innovative thinking, but not necessarily of the sort that one would have expected. In theory, innovation; in practice, gaming.

The assumed positive relationship between performance measurement and innovation implies a more fundamental belief that such measures can, if appropriately designed and applied, be used to strategically change the organization or program so that it is better able—particularly through differentiation—to contend with an otherwise complex environment and uncertain/risky future. A measure of current performance, according to this logic, can highlight the strengths and weaknesses of an organization, providing at least a clear picture of the two basic ingredients in a SWOT (strengths, weaknesses, opportunities, threats) assessment (see Bryson 2004). Measures of performance can also be used comparatively and competitively, providing an agency with a means for exploring opportunities and threats. Performance measurement, in short, can be used to enhance the capacity of an organization to deal with the risks and uncertainties of its future.

When Kaplan and Norton initially developed their balanced scorecard approach (BSC) (1992), they did so with the idea of offering organizations a framework for supplementing and offsetting the typical overreliance on financial measures. In that sense, the BSC focused on the nonfinancial measures as indicators of past and present performance, but they had little else in mind. As they noted nearly a decade later, what they discovered was that many organizations adopting the BSC framework were using it as part of their strategic-planning process.

> Originally, we thought the Balanced Scorecard was about performance measurement…. Once organizations developed their basic system for measuring strategy, however, we quickly learned that *measurement* has consequences far beyond reporting on the past. Measurement creates focus for the future. The measures chosen by managers communicate important messages to all organizational units and employees. To take full advantage of this power, companies soon integrated their new measures into a *management system*. Thus the Balanced Scorecard concept evolved from a performance measurement system to become the organizing framework, the operating system, for a new strategic management system (Kaplan 2001, 99, 102; emphasis in original).

The theory underlying BSC is that performance measures, when put within a coherent and well-articulated frame, can create a more effective organization that is capable of dealing with—and perhaps shaping—its future circumstances. The promise of performance is extended to the promise of planning, with the implication that a BSC-like performance measurement system can become the foundation for developing a more differentiated and capable organization.

In practice, however, the results are more contingent, with factors such as the interplay of organization culture and leadership being extremely significant in shaping the future of any organization. Efforts to use performance measures as a basis for strategic planning and management has its limitations and distorting effects—a point acknowledged in several follow-up case studies by Kaplan and Norton (e.g. Kaplan 2001, 368). The BSC approach reflects what Henry Mintzberg calls the "performance–control" model of traditional management. When applied in the public sector, the "ultimate effect is to reinforce the old machine model."

> In other words, the performance model decentralizes in order to centralize; it loosens up in order to tighten up. And tightening up comes at the expense of flexibility, creativity, and individual initiative. Thus, the brave new world of [performance-based] public

> management all too often comes down to nothing more than the same old machine management—new labels on old bottles. It works fine where machine management worked—sometimes even slightly better—but not anywhere else (Mintzberg 1996, 81).

Finally, although the Government Performance and Results Act of 1993 was described as a major step forward in agency and program accountability, the act makes no mention of the fact that the actual work of federal agencies is now mostly done by third parties. As we discuss in the next section, in the ensuing 18 years it has become increasingly clear that of the many factors influencing agency and program accountability, it is federal policy implementation by third parties that is most important (Radin 2006; Frederickson and Frederickson 2006; Handler 1996; Cooper 2003). The failure of lawmakers to anticipate the salience of third parties explains in part why performance measurement has only marginally influenced the accountability of federal agencies and programs.

When the eight points in Table 2 are brought together they form the basis of these claims—accountability is most effective when:

- there is caution regarding causality;
- measures of performance are understood to be numeric interpretations of reality rather than facts;
- measuring performances can provide information useful to the sharpening of questions;
- measures of the measurable, measures of processes in lieu of outcomes, and measures of the surrogates of results are understood to be useful and legitimate;
- measures of performance should further the distinctions and uniqueness of institutions and communities rather than encourage isomorphism;
- it is understood that doing what can be done is far better than doing nothing and that poor performance measures are evidence of investing in the nasty problems;
- performance measurements rarely improve the prospects for institutional or community-level innovations; and,
- public investments are most effective when they are long-term; short-term performance measurement is seldom the friend of long-term investing.

Accountability in the Extended State

For the past quarter century, students of American political development (APD), such as Stephen Skowronek, Karen Orren, and David Carpenter (among others) (see Orren and Skowronek 2004), have been exploring the emergence of the modern administrative state in the United States. Looking back at the watershed reforms of the Progressive Era, they have provided us with a historical narrative that helps explain some of the more idiosyncratic characteristics of today's American governance system. The nominal American state that existed prior to 1877 (a pivotal year in Skowronek's analysis, when Reconstruction ended in the South) was eventually replaced over the next four decades by one that injected national administrative capacities into the existing constitutional frame. The process of doing so, however, was highly politicized, and the result was a rather unique configuration of the modern bureaucratic state (Skowronek 1982).

At the same time as APD scholars were helping us understand the distinctive nature of the American Administrative State, students of public administration were attempting to understand the fundamental changes that were taking place from the 1970s onwards that challenged the foundations of that state. Privatization, deregulation, tax revolts, outsourcing, managerialism—all pointed to a fundamental shift in the nature of the American Administrative State. Several scholars (Brint Milward and Keith Provan [2000], B. Guy Peters [1994], R.A.W. Rhodes [1994]) termed this the rise of the "hollow state" in which many of the tasks and functions of government were being contracted out to third parties who would act in the name of the state (for general overview, see Milward & Provan 2000). Frederickson (1999) spoke of the fragmentation and "disarticulation" of the state and the reduced capacity of traditional bureaucratic forms to deal with the increasing demands made of the state.

Into this potentially critical situation there emerged the Extended State, as institutional forms and various networks and other collaborative arrangements began to fill the growing vacuum left by the increasingly incapacitated bureaucracies of the traditional Administrative State. Practical solutions quickly outpaced theories as public administrators found themselves "repositioned" as managers of governance in addition to their duties as managers of agencies (Frederickson 1999, 2004). This is the context within which the promise of performance must now operate.

The scope of federal government work that is outsourced, done by contractors, and grantees is astonishing. To illustrate, more than a decade ago it was estimated that there are more than seven contract employees for every one directly employed federal civil servant (Light 1999). Some make much higher estimates. In this way the state is extended or made very much larger without having made the directly employed civil services larger.

The scope of federal government work that is outsourced, done by contractors, and grantees is astonishing.[19] To illustrate, more than a decade ago it was estimated that there are more than seven contract employees for every one directly employed federal civil servant (Light 1999). Some make much higher estimates (Verkuil 2007).[20] In this way the state is extended or made very much larger without having made the directly employed civil services larger. Many agencies of government are indeed "hollowed out" in the sense that most of the substantive work of the agency is all contracted out, leaving the agency to engage primarily in contract management and oversight. Ensuring accountability to federal government purposes in the context of service delivery via contractors is, at best, problematic (U.S. Special Inspector General for Iraq Reconstruction 2009).

A recent study of federal policy implementation by contracts and grants from the perspective of accountability and performance measurement is helpful in giving a sense of the complicated situation created by the Extended State. The Frederickson and Frederickson (2006) findings in *Measuring the Performance of the Hollow State* generally confirm the point made more than 80 years ago by Mary Parker Follett, who wrote that "we see that the place in which orders are given, the circumstances under which they are given, may make all the difference in the world as to the response which we get. Hand them down a long way from President or Works Manager and the effect is weakened. One might say that the strength of a favourable response to an order is in inverse ratio to the distance the order travels" (Follett 1926/2003, 27). Follett's hypothesis had to do with giving orders in a standard organizational hierarchy. If one substitutes goals for orders, and if one assumes that Follett's description of a "favourable response" is essentially the same thing as accountability to program goals, then the Fredericksons' findings support her

19 In 1996, President Clinton declared that "the era of big government is over." Upon looking back it could be said that "the era of big government is over, to be replaced by the era of really big government."

20 Amazingly, it is nearly impossible to come up with more than a guess on this figure or, for that matter, the number of real federal employees. For example, a position that might be counted as an FTE government employee might be (and in some agencies, frequently is) filled by a temporary contract employee (usually through a contracting employment company, who employs staff to run the contract relationship), while at times those positions might be counted as actual employees (FTEs) but go unfilled either by government worker or contractor for extended periods of time.

hypothesis. On the basis of those findings, and adapting Follett's hypothesis to a modern third-party government, they suggest that the strength of agency accountability to program goals is in reverse ratio to the distances actual implementation travels and the layers of third parties through which implementation travels.

In the Extended State, much of the modern federal government could be organizationally understood as vertical networks of third parties. If, as Follett claims, the length of these articulations matter, so too do the unique characteristics of the third parties in those networks matter. For example, agency performance in the context of federal-state relations and state implementation of federal policy is very different from agency performance in the context of for-profit contractors.

The applicability of accountability mechanisms is clearly influenced by the characteristics of third-party agents. This is as true for private sector outsourcing as it is for the public sector. Academics who study "transaction costs economics" note that there are three basic costs to outsourcing: production costs, bargaining costs, and opportunity costs (Vining & Globerman 1999). The first of those is straightforward and addresses the question: how much will be saved or gained (in terms of human and material resources, and so on) by turning over the work of the organization to outsiders? The second and third costs, however, are more complicated and more directly involve accountability. Bargaining costs involve not only the striking of a bargain or agreement, but also developing and sustaining monitoring (pre-factum) and enforcement (post-factum) mechanisms related to the agreement. Opportunity costs cover those behaviors and events that occur outside the bargain—events that change or exceed the circumstances considered in the agreement. Fraudulent behavior by an employee, a natural disaster, the collapse of financial markets, or lack of access to funding source—some unanticipated, some part of the assumed risk that such "bad stuff" won't happen. These too generate a need for some accountability mechanisms. In fact, one might consider that portion of overall transaction costs linked to pre-factum and post-factum mechanisms as a distinct category—accountability costs.

Of course, there are third parties, and then there are third parties. The extent to which accountability mechanisms seek to deal with the problem of control, it matters whether the third parties are states, commercial contractors, research grantees, nonprofits, or commercial firms engaged in governmentally guided performance of public sector functions. Direct control is difficult given the nonhierarchical nature of the relationship, although there are instances when the contractor is subordinated into the bureaucratic process of an agency. For example, a number of employees at one federal passport processing center are contractors who work side-by-side with government employees doing the same or similar tasks, and for all intents and purposes they act (and often feel like) regular employees during normal, day-to-day operations.

But in the more typical nonhierarchical relationship, the procedure of choice for establishing control defaults to some form of performance measurement, and here the situation becomes complicated. Following the logic of performance measurement that is based on the bureaucratic model of the state, accountability to agency goals will improve as a result of the implementation of the GPRA and, during the Bush administration, PART performance-measurement requirements. According to the interview respondents in the Fredericksons' study, the five federal agencies included improved their performance measurement capabilities. Therefore, following the logic of performance measurement, these agencies were more accountable for program outputs and even outcomes. Although there is little evidence that performance measurement actually causes improved performance, agency officials tend to believe it does.

Frederickson and Frederickson (2006) describe the federal third-party agents as "articulated vertical networks of third parties," articulation taken to mean jointed, as in an articulated bus or the articulation of cars in a train. The quality of articulation has to do with the extent to which separate organizations or institutions in a network are coupled, fit together, linked, or combined, as well as with the nature and quality of those connections. Articulation can take many forms, including both block and categorical grants from the federal government to the states and territories or to other subgovernments; contracts let by the federal government or by subgovernments to nonprofit or for-profit third parties; the processes of regulating third parties; and the provision of funding to third parties with clients who have claims for services from these third parties. The point is that the wide variation of arrangements between the federal government and third parties and the remarkable variation in the purposes of these arrangements cannot be understood as just grants or contracts with third parties or subgovernments. The subject of federal grants and contracts is far more complex. Although contracting out and privatization have received the bulk of public attention, contracts are only one form of the devolution of federal program implementation by third parties. Formalized contracts between the federal government and third parties, however, are all various forms of articulations, almost all of them vertical (Salamon 1989).

In third-party grant and contract regimes, the form of articulation between agencies and their third parties ranges from remote control to managed networks. Remote control, according to Salamon (1989), is a loosely coupled grant or contract articulation in which third parties exercise wide discretion and latitude in both the management and the substance of policy implementation. Contracts for health research between the National Institutes of Health and hospitals and universities are a typical example. Managed networks, at the other extreme, are tightly coupled articulations between agencies and grant or contract third parties. Tight network management can be hierarchical or exercised through agency oversight programs, auditing procedures, carefully drafted contracts, tight contract management, negotiated shared understanding between agencies and third parties, and so forth.

The nature and quality of principal-agent network articulations that are on remote control rely primarily on accountability through pre-factum contracts granting considerable latitude or discretion to third parties on post-factum performance-measurement-based assumptions. By comparison, principal-agent network articulations that are managed illustrate the accountability promises of detailed contracts with little latitude and hierarchical control or bureaucratic contract and grant management processes.

Another factor to take into consideration is the fact that many governmental programs have multiple purposes. The level of goal agreement or congruence in an agency's statutory or regulatory foundations and the level of policy agreement among agency principals influences how they respond to the imperatives of accountability. In the context of contradictory or ambiguous goals, the fixing of accountability is rendered more difficult and less effective (Heinrich 1999). Nowhere is this more evident than in the application of accountability expectations to the Food and Drug Administration (FDA). The FDA is expected to enable pharmaceutical and food products to reach market as quickly as possible, while at the same time assuring the public that those pharmaceutical and food products are safe. Which of these goals FDA chooses to maximize is a function of many factors, with political forces in the congressional and executive branches of government and the influence of the pharmaceutical industry being the most important.

In the 1990s, the FDA, along with many other regulatory agencies, was part of a distinct shift in approaches toward federal regulation of businesses (Heimann 1998). This shift was away from traditional regulation toward both deregulation and a logic of partnership whereby regulators and regulated industries share regulatory roles. For the FDA this meant that pharmaceutical and food companies paid fees for regulatory services and expected, in return, timely reviews. In addition, regulated industries, such as the seafood industry, using agreed-upon standards and reporting protocols, were authorized to engage in self-regulation. But in 2004 and 2005, two drugs approved under these partnership arrangements, Vioxx and Bextra, were found to have undisclosed but known dangerous side effects and were taken from the market. The peanut butter-salmonella crisis again highlighted the FDA goal conflict issue. The tension between the FDA's food and drug safety responsibilities and its food and drug promotion responsibilities shift as political forces shift from one emphasis to the other and back again (Heimann 1998). Accountability at the FDA is played out in the context of these contradictory and shifting goal emphases. Evidence of the influence of competing FDA goals is seen in the slowness of their GPRA performance-measurement development and in the ongoing aggregation and disaggregation of FDA performance measures. The decision to shift between two contradictory emphases is a policy rather than a performance challenge and one that performance measurement is ill-suited to solve. This policy challenge leaves the FDA in a position of attempting to manage its way around and through the problem of conflicting goals, and leaves the matter of agency accountability unresolved.

Medicare and Medicaid also have competing goals. Unlike FDA's regulation/promotion goals, Medicare and Medicaid have financial service goals achieved through articulated networks of third parties—the states and their third-party contractors in the case of Medicaid and large-scale health finance and insurance contractors in the case of Medicare. Clients are at the bottom of these vertical networks—clients who are entitled to Medicare and Medicaid services from doctors, pharmacies, hospitals, and nursing homes. Medicare and Medicaid are expected to finance health-care services to those who are eligible and at the same time to control health-care costs. Although goal conflict is evident in the operations of the FDA, for Medicare, and Medicaid, it appears that open consideration of conflicting goals by agencies is unacceptable as a matter of policy. Accountability mechanisms, such as GPRA and PART performance measurement, did not attempt to reconcile these competing objectives.[21] Performance measures and the data needed to make them operative were never described in terms of possible goal conflict. Agency and program goals are described as if they are in agreement or are compatible. Based on GPRA and PART reports, it is as if agency goal conflict and competition is the problem that dares not speak its name. In interviews with agency officials, however, agency goal conflict is a regular theme. There is some evidence, particularly at the FDA and the Center for Medicaid and Medicare Services (CMS), that performance measures are attempts to pave over goal conflicts, as if to claim that really good agency performance will somehow demonstrate that agencies can be equally accountable to conflicting or competing goals (Frederickson and Frederickson 2006.) This would not be the first time that better management is thought to be the answer to vexing questions of goal conflict.

Performance measures and the data needed to make them operative were never described in terms of possible goal conflict. Agency and program goals are described as if they are in agreement or are compatible. Based on GPRA and PART reports, it is as if agency goal conflict and competition is the problem that dares not speak its name. In interviews with agency officials, however, agency goal conflict is a regular theme.

In the third-party government of today's Extended State, policy implementation is, by definition, formally extended beyond an agency's hierarchy. There are variations in forms of agency centralization and decentralization, most particularly programmatic fragmentation and the geographic spread of policy implementation agents. Levels of centralization and decentralization of third-party policy implementation complicate agency accountability. Programs that use grants or contractors to provide services are decentralized by definition.

There appears to be an association between third-party government and the decentralization and fragmentation of the federal government. Patterns of policy implementation that are centrally funded and locally managed and delivered are now deeply established in contemporary federalism. In addition to geographic decentralization, federal program fragmentation is evident in the health

21 Although the newly passed (2011) GPRAMA does attempt to address that situation through authorizing budgetary and management initiatives, time will tell if this adjustment in administrative policy overcomes the problem.

fields; each state and territory operates dozens of federally funded programs. Much of the actual operation of these programs at the street level is contracted out by the states and territories to private nonprofit or for-profit third parties. In every state and locality, federally funded health contractors operate as financial intermediaries or carriers, as health researchers or as providers of direct health and hospital services. It could be argued that third-party government, decentralized federalism, and program fragmentation are nearly the same thing.

How does the correlation between third-party government and decentralization influence accountability, and vice versa? The Fredericksons' study of GPRA implementation indicates, first, that federal agencies with state, territorial, or tribal third parties will build performance-measurement accountability mechanisms into more general patterns of negotiated cooperative federalism or federal-state mutual adjustment. In this context, third parties participate in goal-setting and performance-measurement decisions and become partners in GPRA-based performance-measurement implementation. The most volatile examples are Medicaid and Medicare, which both exhibit constantly changing and churning granting or contracting relationships between CMS and the states—a kind of marathon dance of accountability.

Most of the GPRA-related elements of agency-state relations tend to be among top executives at each level and among GPRA implementation specialists. The rank and file in state agencies that implement federal programs know little of GPRA or PART, and they assume that accountability through performance measurement is someone else's responsibility. Medical research third parties are insulated from formalized accountability mechanisms—both agency goal setting and performance measurement. NIH has entirely walled off their grant recipients from any participation in accountability-based performance measurement and has developed centralized and effective GPRA compliance systems in the context of a decentralized and fragmented medical research model. In other words, performance-based accountability is not often left to third parties.

This suggests that under loosely coupled arrangements, GPRA is treated by the agencies as Washington-level bureaucratic requirements rather than as important elements of their day-to-day contract and grant management of third-party policy implementation. What may, in Washington, be thought of as essential features of accountability and good management tends not to be reflected down loosely articulated chains of third-party implementers. Real accountability, therefore, following the definition of accountability at the beginning of this essay, may have less to do with formal systems of performance measurement, such as GPRA and PART, and more to do with grant and contract management and oversight at the operational level and with executive and legislative branch politics.

In the name of accountability, some nonprofit or for-profit contractors will be obliged to build data and other performance-measurement requirements into contracts. In ordinary contract circumstances third parties will be required to keep certain records and aggregate certain data. Contractors may not know the connection between what they are obligated to provide as a part of contracts and how that data might be used in measuring agency performance. Such contract regimes are top down principal-agent models that assume that it is possible to formalize reliable and consistent data requirements and expectations through compliance language (Radin 2006). Despite rules and regulations, the gathering, organization, and analysis of reliable and consistent performance-measurement data has proven to be an illusion.

When direct government programs operated by civil servants, on one hand, and grant- and contract-based third parties, on the other, were compared, in general terms direct government received significantly higher overall PART scores and, more specifically, higher scores for the management section of the PART. Research and development programs, which are mostly third-party based, are a marked exception to this generalization. Nevertheless, the quality of management and the nature of accountability in hollowed-out third-party-operated federal programs is very different from the management and accountability of directly operated federal programs. The accumulation of GPRA- and PART-based performance measurement results shines a strong light on these differences. One might conclude, on the basis of these data, that third-party-operated federal programs are less accountable and less well managed; or one might conclude that articulated vertical networks of third parties will be inherently or, by their nature, less accountable to federal government objectives and less well managed, at least less well managed in terms of traditional definitions of management (Frederickson and Frederickson 2006, 182-183).

Outsourcing the work of government, and particularly large-scale outsourcing, is a significant challenge to making effective use of accountability mechanisms as a means for dealing with the problems of governance. Putting aside the patterns of post-factum accountability associated with important errors, breakdowns, or scandals and turning instead to the routines of before-the-fact accountability, we learn that:

- in a general way, accountability, and particularly pre-factum accountability, is kept by systems and processes of organizational and managerial control; and
- large-scale outsourcing of the provision of governmental goods and services generally reduces the accountability of governmental agencies acting as contractees or grantees.

We further learn that, in the context of Extended State outsourcing, accountability is more effective when:

- the high variation in forms and types of public purposes and the unique patterns of contractee-contractor relations associated with them are recognized and developed. Specialization matters in the public sector, no less so when the administration of specialized public projects and services are outsourced. One size never fits all;
- the possible attenuation of agency program goals associated with layers of third-party implementers are understood and mitigated;
- the nature of relations between public agency contractees and their contractors are understood (tightly coupled, loosely coupled, managed, remote control, and so forth) and the appropriate form of control management are put in place;
- third parties are not expected to solve the problems of ambiguous or competing public agency goals; and
- third parties expected to gather and provide performance-measurement data are included in the processes by which the reasons for those data are determined and their possible uses are understood.

Accountability and the Search for a High-Trust Culture

We turn now from our evaluation of the contemporary promises of public sector accountability to our understanding and description of a preferred accountability—an idealized, aspirational, and hopeful accountability. From our evaluation of public sector accountability, it may seem that to this point we have emphasized the shortcomings and disappointments of the applications, mechanisms, and processes of accountability and downplayed many of the values or virtues of accountability. It is axiomatic that reforms, including the public sector accountability reform, tend to overpromise. It is also axiomatic that evaluations of reform will point to the shortcomings and disappointments in reform. Such is the nature of reform, and such is the nature of the evaluation of reform.

It is axiomatic that reforms, including the public sector accountability reform, tend to overpromise. It is also axiomatic that evaluations of reform will point to the shortcomings and disappointments in reform. Such is the nature of reform, and such is the nature of the evaluation of reform.

While we describe the disappointments and shortcomings in the application of public sector accountability reforms, we nevertheless acknowledge the significant contributions the accountability reform movement has made to governance. At the foundation of these contributions is a particularly important shift in what public sector accountability is understood to be. Traditionally, accountability was understood to look back, to be held to account for past errors or mistakes; accountability was understood to be associated with an event, an incident, or an action for which one or more persons would be expected to be accountable. Contemporary administrative reforms and the creation of modern bureaucratic infrastructures (e.g., red tape) can be regarded as efforts to promote efficiency and rationality, but also a means for shifting accountability from post factum to pre factum. Indeed, one can see a similar shift in managerial theories of motivation, with a shift from the emphasis on rewards and sanctions to those that stress commitment, morale, and self-actualization. But just as overreliance on post-factum mechanisms can prove dysfunctional, so can an overemphasis on pre-factum mechanisms that—taken to their logical extreme—foster the empty promises of accountability. Whatever benefits might be gained from the shift toward pre-factum mechanisms is likely to be outweighed by the costs when taken too far. The experience of No Child Left Behind is merely one very visible example.

Which leads us to consider an alternative approach based on the one common characteristic of both pre- and post-factum accountability—that is, their focus on the need for social mechanisms (broadly defined) to generate accountable responses. Accountability involves the act (post factum) or expectation (pre factum) of account giving, and either requires the presence or possibility of social mechanisms to be fulfilled. But there is another factor that comes into play and that is the setting within which the mechanisms operate. These settings—whether regarded as cultural contexts, institutional settings, task environments—give form and meaning to account-giving mechanisms. They are the structures, values, norms, beliefs, and attitudes that are requisite for account giving, and they constitute the venues and infrastructure of accountability. Without taking them into consideration, we cannot make sense of account giving successes or failures, nor can we hope to make changes that will enhance or strengthen accountability.

In our effort to unpack and reconceptualize accountability, we would put forward the idea that it is not enough to think of accountability as mechanisms or tools of governance. Instead, we offer the view that *accountability settings* are the core and key factors that must be taken into consideration whenever we evaluate accountability or consider reforms to enhance and strengthen account giving.

In the expanding literature on accountability, the settings are discussed in various ways, usually as cultures of accountability or accountability regimes. These are typically viewed as composites of accountability mechanisms, but we offer a stronger view that treats these contexts as a holistic construct that includes pre- and post-factum mechanisms and more.

What does that "more" involve? That is the critical question that needs answering if we are to advance the cause of improving and strengthening accountable governance. We take our lead from Adam Smith, whose contribution to the study of accountability is rarely acknowledged. For Smith, accountability was sewn into the very fabric of society. "A moral being is an accountable being," he stated in the first edition of his *Theory of Moral Sentiments* (1759, III, 1)—thus articulating a basic premise that informs all his later work, including *The Wealth of Nations*. Contrary to his reputation as an egoist philosopher, Smith believed that any and all social interaction is judged not only by an individual's immediate self-interest, but also by the judgments of an "impartial spectator" that inhabits one's character.[22] In that sense, Smith established nothing less than the first modern theory of accountability, and one is able to see his view of markets in an entirely different light once his idea of moral accountability is taken into consideration.

Cultures are important factors in shaping human choices and behaviors. "Shared values, shared beliefs, shared meaning, shared understanding, and shared sense making are all different ways of describing culture. … These patterns of understanding help us cope with the situations being encountered and also provide a basis for making our own behavior sensible and meaningful" (Morgan 2006, 1348). Building on that idea, we can see several different types of accountability cultures that have a profound influence on accountable governance.[23]

Among the most evident is the culture of professionalism. Sociologists who study professions have always found it difficult to characterize them, but generally they involve a community of actors who possess a body of knowledge and a set of values related to the use of their expertise. Depending on how one defines a profession, their historical role in American government can be traced back to the founding. But there is no question that by the 1960s, professionalism was a pervasive characteristic of the American Administrative State—so much so that it was worthy of close examination by one of the leaders of the field, Frederick Mosher, in a classic 1968 treatment, as well as two symposia in *Public Administration Review* a decade later (see Mosher 1978). Culturally, professions operate as moral communities in the sense that their actions are supposed to reflect an internalized moral obligation that emerges from one's general involvement in a bounded community of professional peers who adhere to an explicit set of shared values and norms.

---

22 "When I endeavour to examine my own conduct, when I endeavour to pass sentence upon it, either to approve or condemn it, it is evident that, in all such cases, I divide myself, as it were into two persons; and that I, the examiner and judge, represent a different character from that other I, the person whose conduct is examined into and judged of. The first is the spectator, whose sentiments with regard to my own conduct I endeavour to enter into, by placing myself in his situation, and by considering how it would appear to me, when seen from that particular point of view. The second is the agent, the person who I properly call myself, and of whose conduct, under the character of a spectator, I was endeavouring to form some opinion. The first is the judge; the second the person judged of. But that the judge should, in every respect, be the same with the person judged of, is as impossible that the cause should, in every respect, be the same with the effect" (TMS, III, 1.6).

23 Analytically, we can talk about the existence of accountability cultures which, for present purposes, can be defined as those *perceived obligations and expectations for account giving that are embedded within the sociocultural environment of an individual or organization*. This is a modified version of a similar concept—accountability web—used in Gelfand et al. (2004). The work of Gelfand and colleagues draws, in turn from Frink and Ferris (1998).

What is key in professional setting are both the shared values and the boundaries that define membership in the community. This setting is where professional accountability cultures thrive, and in government it is exemplified most effectively by those in the uniformed professions (e.g., the military, law enforcement, the U.S. Forest Service).

This response is not surprising since most descriptions of agencies, organizations, or programs as accountable are generally understood to mean accountable to organizational purposes. Following this usage of accountability, we found most public agencies to be generally accountable to their goals or purposes, recognizing, of course, that there are degrees or gradients of public organizational accountability. The degree, level, or quality of public sector accountability, understood within this cultural setting, turns on traditional factors that are thought to determine organizational effectiveness—leadership, commitment to mission, resources, and political support.

There are several excellent examples of so-called high-trust public agencies. Such agencies tend to have leaders who serve for long periods of time. There are several good reasons that the GAO has a well-established high-trust culture, exemplified by the 15-year term of appointment for the Comptroller General (Walker 1986). Admiral Hyman G. Rickover's 63-year tenure in the U.S. Navy no doubt accounted for the high-trust culture of the submarine service and the nuclear navy (Blair 1954). Strong identification with an organization's purposes, so-called mission valence, is still another feature of high-trust organizations: the U.S. Forest Service (Kaufman 1967) is one example, and the Federal Bureau of Investigation was another, until it suffered some reputational damage following the death of J. Edgar Hoover. Symbols like organizational seals, prizes or awards for extraordinary service, uniforms, and social and moral reward systems recognize the sense of duty and public spiritedness of workers (DiIulio 1995). These high culture public organizations have been and are still notably accountable. These forms of accountability are an inherent part of the moral order.

But trust is more than a reflection of long-term leadership. The economist and Nobel Laureate Kenneth Arrow points out that "trust has a very important pragmatic value, if nothing else. Trust is an important lubricant of a social system. It is extremely efficient; it saves a lot of trouble to have a fair degree of reliance on other people's word. Unfortunately this is not a commodity that can be bought very easily. If you have to buy it, you already have some doubt about what you have bought" (Arrow 1974, 23). High levels of trust are fundamental to the functioning of accountable public organizations. At an elemental level, trust is between one person and another. But trust takes on increasing levels of abstraction, as in trust between families, trust in the work group, intraorganizational trust, interorganizational trust, and so forth. Habits of trust are exhibited when drivers instinctively obey street signs and diners instinctively tip servers. In his monumental study of trust, Francis Fukuyama described the United States as a high-trust society, but he covers extensively recent declines in American trust (Fukuyama 1995). Fukuyama's description of trust at the community level is surprisingly close to descriptions of accountability: "As a general rule, trust arises when a community shares a set of moral values in such a way as to create expectations of regular and honest behavior. To some extent, the particular character of these values is less important than the fact that they are shared" (1995, 153).

Our basic contention is that accountability cultures are important to the successful design and maintenance of accountable governance in the traditional Administrative State. The central question we now ask is whether they are also important and appropriate when dealing with the issues and challenges generated by the Extended State. If outsourcing government goods and services reduces accountability, as the evidence presented earlier shows, what can be done to achieve higher levels of contractor accountability? Can there be an idealized, aspirational, and hopeful accountability under conditions of widespread outsourcing and the growing use of third parties to perform government functions?

If outsourcing government goods and services reduces accountability, as the evidence presented earlier shows, what can be done to achieve higher levels of contractor accountability? Can there be an idealized, aspirational, and hopeful accountability under conditions of widespread outsourcing and the growing use of third parties to perform government functions?

The answer can be found in some notable cases where the cultural factor comes into play. The research and development (R and D) functions of the federal government may be regarded as an important exception to the earlier point that in comparative terms outsourced public services tend to be less effective, less well managed, and less accountable than directly operated federal agencies. The operations of federal R and D agencies are particularly challenging in terms of accountability because their tasks tend to be exploratory, the means by which their objectives are to be achieved are often unclear, and genuine performance almost always takes many years rather than a single year. Here we will focus on the operations of the National Institutes of Health (NIH).

NIH has about 18,000 directly employed staff, many of them engaged in contract and grant management. This is a considerably larger grant and contract management staff than one finds in other federal agencies that outsource much of their work. There is little doubt that NIH spends more per dollar on contract management than most other federal agencies. The best estimate is that the annual sum of NIH grants and contracts supports a full time equivalent workforce of approximately 190,000 persons scattered among the leading American medical research centers, medical schools, and universities. Put another way, there are about 11 FTE persons engaged in NIH funded medical and health research for every 1 direct NIH employee (Frederickson and Frederickson 2006). These researchers are scattered across the United States, making NIH one of the most geographically decentralized agencies in the federal government, at least in terms of its grant-funded third parties.

Medical schools and universities organize their research functions on the basis of NIH centers and institutes, with laboratories for cancer, AIDS, and the other diseases around which NIH is organized. In addition, medical schools and health-research organizations are fully geared up to follow NIH grant criteria and protocols. These schools, research organizations, and the various institutes of NIH have interdependent and symbiotic relationships that have evolved over many years. Medical schools and health-research organizations are staffed by highly educated and qualified scientific researchers on the basis of assumptions of a close and continuing relationship based on staff expertise, consistent research success, and the probability of NIH financial support. In turn, NIH serves as a consistent and usually reliable source of financial support. Although NIH grants are made on a competitive basis, using juries and forms of peer review, this competition is within the narrow symbiotic range of established NIH-medical school or health-research organizations.

The NIH initial response to GPRA in the 1990s was interesting and unique. First, NIH chose very general goals that cut across their various institutes, such as "develop new and improved methods for diagnosing disease and disability." Second, research-grant-receiving third parties are given rather wide discretion as they engage in the sometimes untidy and unpredictable pursuit of discoveries. Third, grant-receiving third parties are insulated from the strategic-planning/performance-measurement processes associated with the NIH implementation of GPRA.

Easily, the most interesting feature of the NIH response to the accountability assumptions in GPRA was their initial decision to use qualitative rather than quantitative performance measures. NIH accountability data were "science advances," one-page narratives that detail a particular scientific discovery; "science capsules," one-paragraph narratives that give a snapshot of particular research endeavors; and "stories of discovery," one- or two-page narratives that emphasize the aggregation of long-range processes of medical research. NIH assigned skilled writers to draft these qualitative measures, and they are interesting, engaging, and in some cases, compelling. They are certainly more interesting than standard performance-measure spreadsheets.

According to one NIH official, this is how NIH decided they would implement their performance-measurement responsibilities:

> Well, there was actually a lot of discussion about whether we could use these traditional quantitative methods because, in a way, it would have been easier for us to just use numbers. But we thought that just because we can count it doesn't mean we should, and doesn't mean that it would be meaningful. We felt very strongly that just counting was not going to be useful for demonstrating our accountability to the public and what they're getting for their research dollars. We needed a more descriptive way to say, "Here's the actual outcomes of our research program. Here's what we found, here's what we do, here's what comes from this." So we thought that the publication of our stories is certainly an important vehicle for this—it's the knowledge of new treatments and diagnostics that comes from our research that really needs to be demonstrated (Frederickson and Frederickson 2006, 99).

After several years of the use of qualitative performance measures, NIH changed to standard annual quantitative measures. It is unclear why, but at the time (1999) the Office of Management and Budget was raising the performance-measurement bar.

The case of NIH's initial approach to GPRA serves as a good example of how accountability cultures can be related to Extended State conditions. Those engaged in medical research at NIH and at each of the third-party grant and contract organizations share a common educational background and commitment to scientific health research, and one can classify them as a moral

The case of NIH's initial approach to GPRA serves as a good example of how accountability cultures can be related to Extended State conditions. Those engaged in medical research at NIH and at each of the third-party grant and contract organizations share a common educational background and commitment to scientific health research, and one can classify them as a moral community reflected in a common professional culture rooted in the shared research endeavor.

community reflected in a common professional culture rooted in the shared research endeavor. But there were other accountability cultures at work as well. Although the model is fragmented and decentralized, over time, research cultures flourish and identify the productive, creative, and trustworthy. This arrangement has certain political advantages based on the geographic spread of resources and based on its often compelling search for cures (commitment), and also because some of the best known NIH grant recipients are either prestigious scholars or are associated with prestigious organizations (reputation). As significant, the relationship among NIH and its many third-party partners was a clear example of a high-trust culture of accountability based on organizational interactions.

Another example of extraordinary public accountability in an Extended State context is to be seen in the operation of commercial air travel in the United States. Commercial air travel is a complex, fragmented array of horizontal, vertical, and lateral linkages between multiple jurisdictions at all levels of American government; a wide variety of types of corporations and unions; and a wide range of types of contractors—a system or network rather than a hierarchy (Frederickson and LaPorte 2002). Commercial air-passenger security is part of a unique class of institutional characteristics and challenges that are collectively described as "high-reliability organizations" (HROs). The perspective of high-reliability organizations is based on many years of direct observation of error-intolerant systems, such as commercial air travel, nuclear power generation, nuclear submarine and aircraft carrier operations, production of the components of nuclear weapons, and electricity generation and transmission systems. In the history of American commercial air travel we find outstanding examples of both post-factum and pre-factum accountability. As important, we also find the strong influence of accountability cultures.

The colossal failure of airport and commercial air travel safety on September 11, 2001, resulted in the loss of 3,100 lives and untold billions of dollars in damage to buildings and other physical infrastructure. The events of that day eventually resulted in the wars in Iraq and Afghanistan. Furthermore, the events of that day raised troubling questions about the reliability and security of American commercial air travel. After the terrorism of that day the standard pattern of post-factum account giving played out with legislative hearings, blue ribbon panel and commission appointments, panel and commission reports, and large-scale governmental reorganization. That is traditional accountability.

But the events of 9/11 were the exception in the extreme. Consider pre-factum accountability in the day-to-day functioning of American commercial air travel. On any Friday between 3:00 and 7:00 p.m. about 500,000 Americans will be on airplanes traveling at 500 miles per hour in every conceivable direction, all guided by the air-traffic control system. This unique class of organizational systems works in the context of essential insistence that it be nearly failure free, and, with rare exceptions, it is. That each of these Americans will reach their destinations safely is not regarded as exceptional or miraculous, but a system failure would be regarded as dramatic and horrible.

The density of the commercial air travel HRO is remarkable. Airports are owned and operated by municipalities or local public authorities. Airplanes are built by commercial firms and owned by commercial airline companies responsible for their maintenance and operation. Airline passenger and baggage screening is the responsibility of the Transportation Security Administration, an agency of the federal Department of Homeland Security. Airplane surface travel, take off, landing, and in-flight control is managed by the air-traffic control system of the Federal Aviation Administration, part of the U.S. Department of Transportation. And these are just the larger, more obvious parts of the commercial air-travel HRO system.

Observations of the commercial aviation high-reliability system, as well as other HROs, can be summed-up with the following generalizations:

- First, the physical technologies (radar, hydraulics, and so on) of these systems are tightly coupled, meaning that an important breakdown anywhere along the production process may cause the entire system to fail.
- Second, this tight coupling is characterized by fixed and relatively rigid standard operating procedures, or procedure protocols, that do not ordinarily vary; this means that administrative discretion is sharply reduced.
- Third, humans operating at any point in the production process of high reliability systems require extensive technological training and constant retraining.
- Fourth, such systems are ordinarily funded to a level that will guarantee high efficiency, or, to put it differently, efficiency is much more important than economy in the world of high reliability.
- Fifth, such systems are highly redundant, there being two, three, or even four back-up, or redundant, systems ready to take over should the primary systems fail. One thinks immediately of the redundancy that saved the Apollo 13 space mission.
- Sixth, these systems are highly networked, meaning that many organizations are in the production chain.
- Seventh, these systems are composed of a marvelous mix of governmental, nongovernmental, and commercial organizations, the very definition of high-functioning public-private partnerships.
- Eighth, when the systems are working properly, error reporting is encouraged and not punished; indeed, initiatives to identify flaws in procedures and protocols and thereby avoid failure are rewarded.
- Ninth, ordinarily such systems are rather hierarchical, both within the system and within the organization making up the system. But at times of peak load and emergencies, one finds rule switching by which officials move away from hierarchy and procedures to seek the expertise or experience that might account for or explain an anomaly and suggest possible nonroutine solutions.

These failure-free systems reveal how remarkably effective modern public and private organizations can be if they have adequate resources and are well managed. To be sure, failure-free systems are the subject of intense public scrutiny because of the visibility of failures, however rare.

Every day, we all enjoy the modern miracles of high-reliability systems. And, interestingly, when they fail, it is usually because of human fallibility. It is difficult to imagine modern life without high-reliability commercial air travel. When it works perfectly, nothing appears to happen; in fact, everything happens properly. Commercial air travel is the very definition of successful accountability in an Extended State context.

There will be failures and there will be accidents; simple probability demonstrates that it is so (Perrow 1999). But every day, we all enjoy the modern miracles of high-reliability systems. And, interestingly, when they fail, it is usually because of human fallibility. It is difficult to imagine modern life without high-reliability commercial air travel. When it works perfectly, nothing appears to happen; in fact, everything happens properly. Commercial air travel is the very definition of successful accountability in an Extended State context.

And yet, underlying all the pre-factum accountability mechanisms that make this complex interorganizational arrangement operate are interactive cultural settings, from those emphasizing the professionalism of pilots and other specialists, to a general commitment to safety and industrywide efforts to promote and sustain positive reputation for effective and efficient performance. All these come together in an accountability culture of high trust so pervasive that on most days the system operates without even a thought as to who might be blamable for some tragic mishap since there exists a widespread belief that such an event is highly unlikely.

Viewing accountable governance through the lens of accountability settings and cultures adds considerably to our understanding of how the Extended State can meet the challenges of efficient management, effective performance, and democratic legitimacy. The insights are anecdotal and tentative at this point, and what is required is a closer empirical examination of the dynamics of accountability that takes place in Extended State settings. One initial conclusion is that of the four accountability cultures, the most important in most public sector situations seems to be the high-trust type. In a sense, it is the setting that encompasses the other three and (as we contend in the next section) the one most closely associated with the holy grail of democratic legitimacy.

High levels of trust do not eliminate the need for pre- or post-factum mechanisms, but rather changes the role and functionality of those instruments. This was difficult to achieve under the traditional Administrative State—and it is critical for our successful transition to an Extended State in an atmosphere where public accountability is even more crucial to deal with our complex problems.

This is the major point the message delivered so effectively by philosopher Onora O'Neill in her BBC *Reith Lectures*, titled "A Question of Trust" (O'Neill 2002). In those lectures, O'Neill points to the fact that while we are increasingly dependent on others, we are also increasingly suspicious of them. Whether it be the highly trained health-care professional, the flight attendant demonstrably committed to our safety, or an Internet company executive claiming to be a strict guardian of our privacy, we remain concerned about the reliability of those mechanisms in place within their cultural settings that they claim hold them to account. Trust me, I'm a professional! Trust me, I really care about your well-being! Trust me, I am a person of high integrity who would not risk her reputation by making false promises! But a great many of us respond with skepticism, for our general culture does not promote trusting attitudes toward those who make such claims. For O'Neill, the result is a rush to impose more and more accountability requirements that undermine rather than enhance trust.

O'Neill's observations highlight the importance of focusing on the development of high-trust accountability cultures. High levels of trust do not eliminate the need for pre- or post-factum mechanisms, but rather changes the role and functionality of those instruments. This was difficult to achieve under the traditional Administrative State—and it is critical for our successful transition to an Extended State in an atmosphere where public accountability is even more crucial to deal with our complex problems.

Conclusions: Public Accountability and the Extended State

Although it may seem abstract and somewhat idealistic, a reconceptualization of public administration has emerged in recent years, a reconceptualization that should inform the subject of public accountability. Until rather recently the word *public* in *public administration* was understood to mean essentially the same thing as "government," that public administration is the same thing as government administration. In the early 1980s, David Mathews pointed out that "public" is a pregovernmental concept, a concept much larger than government, and that government is just one of the ways the public or a public expresses itself (1984). Barry Bozeman argued that all organizations are public, just differently public (2004). Frederickson then pressed the claim that governments are unlikely to effectively manage the modern challenges they face without an increased emphasis on the public in public administration (1997). The primary reason is the growing mismatch between social and economic problems on one hand and the declining capacity of governmental jurisdictions to contain or to manage those problems on the other.

Public management is now understood to include government but also all of those organizations and institutions that contract with government to do governmental work. Public administration in the Extended State includes those institutions and organizations that are essentially public serving—the so-called nongovernmental public sector organizations—and the wide range of organizations and institutions that are essentially quasi-governmental in their relationship with citizens, such as privately held utilities. The distinctions between institutions that are essentially public in character and institutions that are private and profit making are now fuzzy (Bozeman 2004). Modern public-management scholars have developed a nuanced conception of institutions that are governmental, nonprofit, and corporate, but also primarily public serving, on the one hand, and institutions that are clearly profit making and in an identifiable market, on the other hand.

The modern field of public administration is increasingly understood to encompass or include government *and* the full range of nonprofit, nongovernmental, and commercial firms engaged in public work. This is especially the case for all the public services delivered by nongovernmental organizations that are financed by governmental revenues. At the center of this reasoning are efforts to make the logic of public accountability and particularly pre-factum public accountability operational in the public service work of nongovernmental organizations. Such efforts may include codes of public service ethics, public service oaths, inclusion of the meaning of public service in training programs, and so forth. Such efforts are especially urgent in nongovernmental organizations operating prisons and other parts of the criminal justice system; nongovernmental organizations operating so-called back office governmental tasks; nongovernmental organizations doing the anti-terrorism work associated with water, food, and drug supplies, air quality and other environmental vulnerabilities; and so forth. However idealistic and optimistic, the aim would be to strengthen public accountability by strengthening the high-trust public service cultures of all of those organizations that do public work.

The rhetoric associated with large-scale outsourcing provides certain clues to contemporary thinking about the increasing role of nongovernmental organizations in the provision of public goods and services. Some phrases used to describe this phenomenon are distinctly questioning or imply concern or disagreement; such as "third-party government," "the shadow bureaucracy," "the state of agents," "hollowed out." These descriptors tend to be used by scholars who study modern public administration. Phrases that are thought to be more neutral or to be distinctly pro-business include "public-private partnerships," "the blended workforce," "relational contracting," "collaborative governance," and "vested outsourcing." These phrases are more commonly used by consultants, corporate executives, and pro-business outsourcing scholars.

The title and subtitle of a recent book on the subject is revealing. John D. Donahue and Richard J. Zeckhauser of the Kennedy School of Government at Harvard, under the title *Collaborative Governance: Private Roles for Public Goals in Turbulent Times,* write:

> All too often government lacks the skill, the will, and the wallet to meet its missions. Schools fall short of the mark while roads and bridges fall into disrepair. Health care costs too much and delivers too little. Budgets bleed red ink as the costs of services citizens want outstrips the taxes they are willing to pay. *Collaborative Governance* is the first book to offer solutions by demonstrating how government at every level can engage the private sector to overcome seemingly insurmountable problems and achieve public goals more effectively (2011, inside flyleaf).

The word *governance,* as distinct from government, has come increasingly to mean the engagement of governmental and nongovernmental organizations in the functioning of the Extended State. The phrase *collaborative governance* strengthens the description of cooperation between governmental and nongovernmental organizations in the Extended State. And the phrase *public goals* as against governmental goals in the subtitle elevates the goals that collaboration with the private sector would help achieve. Note, however, the use of the phrase *private roles,* to describe the roles of private sector rather than a phrase that might have served to describe the *public* roles of the private sector when it engages in public work.

Many nonprofit and corporate organizations are engaged solely, by contracting, in public work. Such organizations may not be governmental but they certainly are public. Many, if not most, of these organizations are trusted and reliable partners in the implementation of public policy. Is it idealistic to expect a public ethos to emerge in these organizations, an ethos of public accountability and particularly pre-factum accountability? Can there be such a public ethos or moral community shared between contracting principals and contractor agents?

Many nonprofit and corporate organizations are engaged solely, by contracting, in public work. Such organizations may not be governmental but they certainly are public. Many, if not most, of these organizations are trusted and reliable partners in the implementation of public policy. Is it idealistic to expect a public ethos to emerge in these organizations, an ethos of public accountability and particularly pre-factum accountability? Can there be such a public ethos or moral community shared between contracting principals and contractor agents?

At the outset of this monograph we highlighted the three problems of governance that accountability addresses—accountability achieved by and through organizational and managerial control, particularly in the context of the extended state; accountability achieved by and through performance measurement; and accountability achieved by and through democratic government. Of the three, it is the last of these that is the most empirically informed. Recall that we found evidence of weak electoral accountability in the practices of traditional democratic politics, the politics of parties, campaigns, campaign financing, interest groups, elections, law and policymaking, and budget making. This form of political accountability is weak because the political process is so often rigged in favor of incumbents and because political accountability exercises tend to be post-factum account-giving hearings that are often political theater rather than serious account giving.

Stronger forms of democratic accountability and particularly accountability based on high levels of trust are to be found in the locally based civic deliberation, social capital, and civil society movement. By any reckoning the evidence is impressive.[24] Using the 50 states as the data base (and controlling for the standard list of socioeconomic and demographic factors), states with higher levels of social capital have children who are healthier, better educated, and safer. Schools work better in high social-capital states. Kids watch less television in high social-capital states. Violent crime is rarer in high social-capital states. The economies of high social-capital states are stronger. People live longer and tend to be happier in high social-capital states. Tax evasion is low where social capital is high. Although the statistical signs are weaker, social capital and tolerance (race, gender, and civil liberties) go together. Greater concern for civic and economic equality is found in high social-capital states (Putnam 2000). Taken together, these findings are evidence of high-trust cultures and reflect high prospects for strong forms of public accountability rooted in civil deliberation and social capital.

Clues in the search for the high-trust accountable cultures are found in:

- the modern reconceptualization of accountability to acknowledge the role played by pre-factum forms of accountability;
- many examples of high culture and high-trust organizations that have been and are successful;
- the National Institutes of Health as an example of operationalizing high-trust accountability cultures in the context of large-scale outsourcing;
- the operations of high-reliability organizations and particularly commercial air travel as evidence of an accountable Extended State; and
- attempts to both clarify and press the claim that nongovernmental organizations engaged by contract to do public work are expected to have a public ethos and to be publically accountable.

24 Measuring social capital in the United States includes measures of community organizational life, levels of engagement in public affairs, measures of community volunteerism, measures of informal sociability, and measures of social trust (Putnam 2000, 287-295).

* * *

Few phrases are as compelling and powerful as *public accountability*. The many uses of that phrase have given it a privileged place in the grammar of public policy and administration. It is not just grammar; the engagement of governmental and policy applications or practices in the name of public accountability have been and continue to be important. As the preceding pages attest, the applications and practices of public accountability often have fallen short of the rhetoric. While such shortcomings disappoint, they do not negate its potential. If anything, both the theories and the applications of public accountability, to date, have been beneath its potential and have not used its privileges to advantage.

Public accountability is neither a fad nor a passing fancy; it is here to stay. Indeed, public accountability and its attendant concepts and practices are so essential to modern public policymaking and policy implementation as to have become the subject.

Public accountability is neither a fad nor a passing fancy; it is here to stay. Indeed, public accountability and its attendant concepts and practices are so essential to modern public policymaking and policy implementation as to have become the subject. The primary reason for the ascendency of public accountability, we suspect, has to do with the limitations of governmental jurisdictions, the extension of governmental jurisdictions by outsourcing, and the growing importance of distinctly public functions exercised by nongovernmental organizations. It is the concepts and practices of public accountability that provide the handles that scholars and public officials grip as they cope with the challenges of modern governance.

The first and most important of these handles is to understand the modern use of the word *accountability* and the concepts and applications that flow from that understanding. Accountability and particularly public accountability is both post-factum account giving after an event, a failure, or a mistake; and pre-factum organizational and managerial procedures, processes, and practices that are designed to cause or result in accountability to public goals or purposes. Given this modern understanding of accountability, it could be said that most governmental and nongovernmental organizations engaged in serving the public are, to some extent, accountable to their public policy purposes and to their clients. Accountability, then, is a matter of degrees and a question of the level of continuing public support for the particular public policy purposes of organizations held to accountability standards.

The second public accountability handle is the notion of the public. Public is a pregovernmental phenomenon. Governments float in a vast public sea. Engagement in the work of governmental jurisdictions is just one of the many ways groups and individuals are publicly engaged. It follows, then, that public accountability is greater than governmental accountability and may take many forms.

The third public accountability handle flows from the second, the prospects and hopes associated with applications of public accountability expectations on the part of nongovernmental organizations engaged in public work, often on the basis of governmental grants of contracts. Workable forms of public-private cooperation that are publicly accountable will require, we suggest, formalized expressions of the public-serving standards of nongovernmental organizations as well as the cultivation of corporate public accountability cultures. The fourth public accountability handle is to recognize that accountability applications and practices are site, place, or context specific. Just as organizations, both governmental and nongovernmental, may have cultures of accountability, so too do neighborhoods, clubs, churches, and other local public organizations. High levels of civic capital are distinctly associated with expectations of public accountability.

The fifth public accountability handle may not always provide a sure grip. Although organizational performance measurement is often justified on premises of enhanced public accountability, the evidence is mixed. Performance measurement, if carefully and prudently done, is helpful, but it is not the same thing as public accountability.

The sixth public accountability handle is the salience of high-trust accountability cultures among the increasingly complex mix of organizations that implement government program in the Extended State. Such organizational arrangements are best when characterized by high levels of internal trust as well as high levels of public trust. Public accountability can also be found where there are high levels of interorganizational trust, built on years of effective so-called public-private partnerships.

Finally, in aspirational terms, public accountability's highest potential is to be found in high-trust environments where governmental and nongovernmental organizations, working individually or together, are dedicated to public service and to accountability to the public. It is the building and maintenance of such an accountability culture that should mark the path forward.

References

Aberbach, Joel D., and Bert A. Rockman. 2000. *In the Web of Politics: Three Decades of the U.S. Federal Executive*. Washington, D.C.: Brookings Institution Press.

Adams, Guy B., and Danny L. Balfour. 2004. *Unmasking Administrative Evil*. Armonk, NY: M.E. Sharpe.

Adserà, Alícia, Carles Boix, and Mark Payne. 2003. "Are You Being Served? Political Accountability and Quality of Government." *Journal of Law, Economics, & Organization* 19: 445-490.

Ammons, David N. 2002. "Performance Measurement and Managerial Thinking." *Public Performance & Management Review* 25: 344-7.

Arendt, Hannah. 1976. *Eichmann in Jerusalem: A Report on the Banality of Evil*. New York: Penguin Books.

Arrow, Kenneth J. 1974. *The Limits of Organization*. New York: W.W. Norton.

Axelrod, Robert. 1984. *The Evolution of Cooperation*. New York: Basic Books.

Behn, Robert D. 2007. *What All Mayors Would Like to Know About Baltimore's Citistat Performance Strategy*. Washington, DC: IBM Center for the Business of Government.

Berman, Evan M. 1998. *Productivity in Public and Nonprofit Organizations: Strategies and Techniques*. Thousand Oaks, CA: Sage Publications.

Bertelli, Anthony, and Craig R. Smith. 2005. "Relational Contracting and Managing Networks: New Public Management and the Search for Gains from Repeated Interaction." *Annual Meeting of the Public Management Research Association*.

Blair, Clay. 1954. *The Atomic Submarine and Admiral Rickover*. New York: H. Holt.

Bok, Derek C. 2001. *The Trouble with Government*. Cambridge, MA: Harvard University Press.

Bovens, Mark. 2005. "Public Accountability." In *The Oxford Handbook of Public Management*, ed. Ewan Ferlie, Laurence E. Lynn, and Christopher Pollitt. Oxford: Oxford University Press.

Boyte, Harry C., and Nancy N. Kari. 1996. *Building America: The Democratic Promise of Public Work*. Philadelphia, PA: Temple University Press.

Bozeman, Barry. 2004. *All Organizations Are Public: Bridging Public and Private Organizational Theories*. Washington, DC: Beard Books.

Bozeman, Barry. 1993. "A Theory of Government Red Tape." *Journal of Public Administration Research & Theory* 3:273-303.

Bozeman, Barry. 2000. *Bureaucracy and Red Tape*. Upper Saddle River, NJ: Prentice Hall.

Breul, Jonathan D. 2007. "Three Bush Administration Management Reform Initiatives: The President's Management Agenda, Freedom to Manage Legislative Proposals, and the Program Assessment Rating Tool." *Public Administration Review* 67: 21-26.

Brooks, Theo. 1995. *Accountability: It All Depends on What You Mean*. Clifton, NJ: Akkad Press.

Bryson, John M. 2004. *Strategic Planning for Public and Nonprofit Organizations: A Guide to Strengthening and Sustaining Organizational Achievement*. San Francisco: Jossey-Bass.

Butler, Judith. 2001. "Giving an Account of Oneself." *Diacritics* 34: 22-40.

Carroll, Archie B. 1998. "The Four Faces of Corporate Citizenship." *Business and Society Review* 100:1-7.

Christensen, Clayton M. 2000. *The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail*. New York: HarperBusiness.

Collins, James C., and Jerry I. Porras. 1994. *Built to Last: Successful Habits of Visionary Companies*. New York: HarperBusiness.

Cooper, Phillip J. 2003. *Governing by Contract: Challenges and Opportunities for Public Managers*. Washington, DC: CQ Press.

DiIulio, John J., Jr., ed. 1994. *Deregulating the Public Service: Can Government Be Improved?* Washington, DC: Brookings Institution.

DiIulio, John J. 1995. *Barbed Wire Bureaucracy: Politics and Administration in the Federal Bureau of Prisons, 1930-1988*. New York: Oxford University Press.

Dodaro, Gene L., U.S. Government Accountability Office. 2011. "Government Performance: GPRA Modernization Act Provides Opportunities to Help Address Fiscal, Performance, and Management Challenges." *Testimony Before the Committee on the Budget, U.S. Senate (March 16)* GAO-11-466T.

Donahue, John D., and Richard J. Zeckhauser. 2011. *Collaborative Governance: Private Roles for Public Goals in Turbulent Times*. Princeton, NJ: Princeton University Press.

Drucker, Peter F. 1954. *The Practice of Management*. New York: Harper and Row.

Drucker, Peter F. 1976. "What Results Should You Expect? A Users' Guide to MBO." *Public Administration Review* 36:12-19.

Dubnick, Melvin J. 1998. "Clarifying Accountability: An Ethical Theory Framework." In *Public Sector Ethics: Finding and Implementing Values*, ed. Charles Sampford, Noel Preston Preston, and C.A. Bois. Leichhardt, NSW, Australia: The Federation Press/Routledge.

Dubnick, Melvin J. 2002. "Seeking Salvation for Accountability." *American Political Science Association*.

Dubnick, Melvin J. 2005. "Accountability and the Promise of Performance: In Search of the Mechanisms." *Public Performance & Management Review* 27: 376-417.

Dubnick, Melvin J., and H. George Frederickson. 2011. "Introduction." In *Accountable Governance: Problems and Promises*, ed. Melvin J. Dubnick, and H. George Frederickson. Armonk, NY: M.E. Sharpe.

Elster, Jon. 1999. "Accountability in Athenian Politics." In *Democracy, Accountability, and Representation*, ed. Adam Przeworski, Susan Carol Stokes, and Bernard Manin. Cambridge, UK; New York: Cambridge University Press.

Etzioni, Amitai. 1993. *The Spirit of Community: The Reinvention of American Society*. New York: Touchstone Books.

Ezzamel, Mahmoud. 1997. "Accounting, Control and Accountability: Preliminary Evidence From Ancient Egypt." *Critical Perspectives on Accounting* 8: 563-601.

Ferlie, Ewan, Laurence E. Lynn, and Christopher Pollitt, eds. 2005. *The Oxford Handbook of Public Management*. Oxford: Oxford University Press.

Follett, Mary Parker. 2003. "The Giving of Orders (1926)." In *Dynamic Administration: The Collected Papers of Mary Parker Follett*, ed. Henry C. Metcalf, and Lyndall F. Urwick. London: Routledge/ Psychology Press.

Forsythe, Dall W., ed. 2001. *Quicker Better Cheaper? Managing Performance in American Government* Albany, NY: Rockefeller Institute Press.

Foucault, Michel. 1991. "Governmentality." In *The Foucault Effect: Studies in Governmentality: With Two Lectures by and an Interview with Michel Foucault*, eds. Graham Burchell, Colin Gordon, and Peter Miller. Chicago: University of Chicago Press.

Frederickson, David G., and H. George Frederickson. 2006. *Measuring the Performance of the Hollow State*. Washington, DC: Georgetown University Press.

Frederickson, H. George. 1997. *The Spirit of Public Administration*. San Francisco: Jossey-Bass Publishers.

Frederickson, H. George. 1999. "The Repositioning of American Public Administration." *Ps: Political Science & Politics* 32: 701-711.

Frederickson, H. George. 2005. "Whatever Happened to Public Administration?: Governance, Governance Everywhere." In *The Oxford Handbook of Public Management*, ed. Ewan Ferlie, Laurence E. Lynn, and Christopher Pollitt. Oxford: Oxford University Press.

Frederickson, H. George, and Todd R. La Porte. 2002. "Airport Security, High Reliability, and the Problem of Rationality." *Public Administration Review* 62: 33-43.

Frederickson, H. George., Gary A. Johnson, and Curtis H. Wood. 2004. *The Adapted City: Institutional Dynamics and Structural Change*. Armonk, NY: M.E. Sharpe.

Frink, Dwight D., and Gerald R. Ferris. 1998. "Accountability, Impression Management, and Goal Setting in the Performance Evaluation Process." *Human Relations* 51: 1259-1283.

Fukuyama, Francis. 1995. *Trust: The Social Virtues and the Creation of Prosperity*. New York: Free Press.

Fukuyama, Francis. 2011. *The Origins of Political Order: From Prehuman Times to the French Revolution*. New York: Farrar, Straus and Giroux.

Fung, Archon, Mary Graham, and David Weil. 2007. *Full Disclosure: The Perils and Promise of Transparency*. New York: Cambridge University Press.

Gelfand, Michele J., Beng-Chong Lim, and Jana L. Raver. 2004. "Culture and Accountability in Organizations: Variations in Forms of Social Control Across Cultures." *Human Resource Management Review* 14: 135-166.

Gilliom, John. 2010. "Lying, Cheating, and Teaching to the Test: The Politics of Surveillance Under No Child Left Behind." In *Schools under Surveillance: Cultures of Control in Public Education*, eds. Torin Monahan and Rodolfo D. Torres. New Brunswick, NJ: Rutgers University Press.

Gilmour, John B. 2007. "Implementing OMB's Program Assessment Rating Tool (Part): Meeting the Challenges of Integrating Budget and Performance." *OECD Journal on Budgeting OECD Journal on Budgeting* 7: 121-160.

Gilmour, John B., and David E. Lewis. 2006. "Does Performance Budgeting Work? An Examination of the Office of Management and Budget PART Scores." *Public Administration Review* 66: 742-52.

Handler, Joel F. 1996. *Down From Bureaucracy: The Ambiguity of Privatization and Empowerment.* Princeton, NJ: Princeton University Press.

Hargrove, Edwin C., and John C. Glidwell, eds. 1990. *Impossible Jobs in Public Management.* Lawrence, KS: University Press of Kansas.

Harwood, Richard C., and John A. Creighton. 2008. "*The Organization-First Approach: How Programs Crowd Out Community.*"

Hatry, Harry P. 2006. *Performance Measurement: Getting Results.* Washington, DC: Urban Institute Press.

Heimann, C. F. Larry. 1998. *Acceptable Risks: Politics, Policy, and Risky Technologies.* Ann Arbor, MI: University of Michigan Press.

Heinrich, Carolyn J. 1999. "Do Government Bureaucrats Make Effective Use of Performance Management Information?" *Journal of Public Administration Research and Theory* 9: 363-393.

Hood, Christopher. 2002. "The Risk Game and the Blame Game." *Government and Opposition* 37: 15-37.

Hood, Christopher, and David Heald, eds. 2006. *Transparency: The Key to Better Governance?* Oxford: Oxford University Press.

Hoos, Ida R. 1972. *Systems Analysis in Public Policy: A Critique, Revised Edition*. Berkeley, CA: University of California Press.

Hoos, Ida R. 1973. "Systems Techniques for Managing Society: A Critique." *Public Administration Review* 33:157-164.

Hummel, Ralph P. 2008. *The Bureaucratic Experience: The Postmodern Challenge*. Armonk, NY: M.E. Sharpe.

Ingraham, Patricia W. 2007. *In Pursuit of Performance: Management Systems in State and Local Government*. Baltimore: Johns Hopkins University Press.

Johnson, Steven. 2010. *Where Good Ideas Come From: The Natural History of Innovation*. New York: Riverhead.

Kamensky, John M., and Albert Morales, eds. 2005. *Managing for Results, 2005*. Lanham, MD: Rowman & Littlefield.

Kaplan, Robert S. 2001. "Strategic Performance Measurement and Management in Nonprofit Organizations." *Nonprofit Management and Leadership* 11: 353-370.

Kaplan, Robert S., and David P. Norton. 1992. "The Balanced Scorecard—Measures That Drive Performance." *Harvard Business Review* 70: 71-79.

Kaufman, Herbert. 1967. *The Forest Ranger: A Study in Administrative Behavior*. Baltimore, MD: Johns Hopkins University Press.

Kearney, Richard C., and Evan M. Berman, eds. 1999. *Public Sector Performance: Management, Motivation, and Measurement*. Boulder, CO: Westview Press.

Kelly, Janet M. 2002. "Why We Should Take Performance Measurement on Faith (Facts Being Hard to Come By and Not Terribly Important)." *Public Performance & Management Review* 25: 375-380.

Kravchuk, Robert S., and Ronald W. Schack. 1996. "Designing Effective Performance-Measurement Systems under the Government Performance and Results Act of 1993." *Public Administration Review* 56: 348-358.

Lerner, Jennifer S., and Philip E. Tetlock. 1999. "Accounting for the Effects of Accountability." *Psychological Bulletin* 125: 255-275.

Light, Paul C. 1999. *The True Size of Government*. Washington, DC: Brookings Institution.

Lowi, Theodore J. 1969. *The End of Liberalism: Ideology, Policy, and the Crisis of Public Authority*. New York: Norton.

Majone, Giandomenico. 1999. "The Regulatory State and Its Legitimacy Problems." *West European Politics* 22:1-24.

Mathews, David. 1984. "The Public in Practice and Theory." *Public Administration Review* 44:120-125.

Mathews, David. 1999. *Politics for People: Finding a Responsible Public Voice*. Urbana, IL: University of Illinois Press.

Mathews, David. 2011. "Foreword." In *Accountable Governance: Problems and Promises*, eds. Melvin J. Dubnick and H. George Frederickson. Armonk, NY: M.E. Sharpe.

Maynard-Moody, Steven, and Michael C. Musheno. 2003. *Cops, Teachers, Counselors: Stories From the Front Lines of Public Service*. Ann Arbor: University of Michigan Press.

McGregor, Douglas. 1960. The *Human Side of Enterprise*. New York: McGraw-Hill.

Milward, Brinton, and Keith G. Provan. 2000. "Governing the Hollow State." Jour*nal of Public Administration Research and Theory 10:* 359-79.

Mintrop, Henrich, and Gail L. Sunderman. 2009. "Why High Stakes Accountability Sounds Good But Doesn't Work—and Why We Keep on Doing it Anyway." *The Civil Rights Project/Proyecto Derechos Civiles at UCLA*.

Mintzberg, Henry. 1996. "Managing Government, Governing Management." *Harvard Business Review* 74: 75-83.

Morgan, Gareth. 2006. *Images of Organization*. Thousand Oaks, CA: Sage Publications.

Mosher, Frederick C. 1968. *Democracy and the Public Service*. New York: Oxford University Press.

Mosher, Frederick C. 1978. "Professions in Public Service." *Public Administration Review* 38: 144-150.

Moynihan, D. P. 2006. "What Do We Talk About When We Talk About Performance? Dialogue Theory and Performance Budgeting." *Journal of Public Administration Research & Theory Journal of Public Administration Research & Theory* 16: 151.

National Academy of Public Administration. 1983. *Revitalizing Federal Management: Managers and Their Overburdened Systems*.

National Civic League Staff. 1999. *The Civic Index: Measuring Your Community's Civic Health*. National Civic League.

Nayak, P. Ranganath, and John M. Ketteringham. 1994. *Breakthroughs*. Amsterdam; San Diego: Pfeiffer.

Newell, Elizabeth. 2009. "OMB Will Allow Agencies to Set Their Own Performance Priorities."

Newland, Chester A. 1976. "Public Personnel Administration: Legalistic Reforms vs. Effectiveness, Efficiency, and Economy." *Public Administration Review* 36: 529-537.

Nichols, Sharon L., and David C. Berliner. 2005. "The Inevitable Corruption of Indicators and Educators Through High-Stakes Testing." EPSL-0503-101-EPRU.

O'Neill, Onora. 2002. *A Question of Trust*. Cambridge: Cambridge University Press.

Orren, Karen, and Stephen Skowronek. 2004. *The Search for American Political Development*. New York: Cambridge University Press.

Perrow, Charles. 1999. *Normal Accidents: Living With High-Risk Technologies*. Princeton, NJ: Princeton University Press.

Peters, B. Guy. 1994. "Managing the Hollow State." *International Journal of Public Administration* 17: 739-756.

Piotrowski, Suzanne J. 2007. *Governmental Transparency in the Path of Administrative Reform*. Albany: State University of New York Press.

Poister, Theodore H., and Gregory Streib. 1995. "MBO in Municipal Government: Variations on a Traditional Management Tool." *Public Administration Review* 55: 48-56.

Popham, W. James. 2005. "How to Use PAP to Make AYP Under NCLB." *The Phi Delta Kappan* 86: 787-791.

Posner, Richard A., and Eric B. Rasmusen. 1999. "Creating and Enforcing Norms, With Special Reference to Sanctions." *International Review of Law and Economics* 19: 369-382.

Prajogo, Daniel I., and Pervaiz K. Ahmed. 2006. "Relationships Between Innovation Stimulus, Innovation Capacity, and Innovation Performance." R&D *Management 36:* 499-515.

Putnam, Robert D. 2000. *Bowling Alone: The Collapse and Revival of American Community*. New York: Simon & Schuster.

Radin, Beryl A. 1998. "Searching for Government Performance: The Government Performance and Results Act." *Ps: Political Science & Politics* 31: 553-555.

Radin, Beryl. 2006. *Challenging the Performance Movement: Accountability, Complexity, and Democratic Values*. Washington, DC: Georgetown University Press.

Ravitch, Diane. 2010. *The Death and Life of the Great American School System: How Testing and Choice Are Undermining Education*. New York: Basic Books.

Rhodes, R. A. W. 1994. "The Hollowing Out of the State: The Changing Nature of the Public Service in Britain." Pol*itical Quarterly 65:*138.

Roberts, Jennifer Tolbert. 1982. *Accountability in Athenian Government*. Madison, WI: University of Wisconsin Press.

Rohr, John A. 1986. *To Run a Constitution: The Legitimacy of the Administrative State*. Lawrence, KS: University Press of Kansas.

Rose, Richard. 1977. "Implementation and Evaporation: The Record of MBO." *Public Administration Review* 37: 64-71.

Rosenbloom, David H. 1983. "Public Administrative Theory and the Separation of Powers." *Public Administration Review* 43: 219-227.

Salamon, Lester M., ed. 1989. *Beyond Privatization: The Tools of Government Action*. Washington, DC: Urban Institute.

Schorr, Lisbeth B. 1997. *Common Purpose: Strengthening Families and Neighborhoods to Rebuild America*. New York: Anchor Books, Doubleday.

Schudson, Michael. 1998. *The Good Citizen: A History of American Civic Life*. New York: Martin Kessler Books.

Selznick, Philip. 1992. The *Moral Commonwealth: Social Theory and the Promise of Community. Berkeley*, CA: University of California Press.

Selznick, Philip. 1995. "Thinking About Community: Ten Theses." *Society 32:* 33-7.

Shearer, Teri. 2002. "Ethics and Accountability: From the For-Itself to the For-the-Other." *Accounting, Organizations and Society* 27: 541-573.

Sherwood, Frank P., and William J. Page, Jr. 1976. "MBO and Public Management." *Public Administration Review* 36: 5-12.

Sinclair, Amanda. 1995. "The Chameleon of Accountability: Forms and Discourses." *Accounting, Organizations and Society 20:* 219-37.

Skocpol, Theda. 2003. *Diminished Democracy: From Membership to Management in American Civic Life*. Norman: University of Oklahoma Press.

Skocpol, Theda, and Morris P. Fiorina, eds. 1999. *Civic Engagement in American Democracy.* Washington, DC: Brookings Institution Press.

Skowronek, Stephen. 1982. *Building a New American State: The Expansion of National Administrative Capacities, 1877-1920. Cambridge* (UK): Cambridge University Press.

Smith, Adam. 1759. *The Theory of Moral Sentiments*. London: A. Millar.

Stone, Walter J., Sarah A. Fulton, Cherie D. Maestas, and L. Sandy Maisel. 2010. "Incumbency Reconsidered: Prospects, Strategic Retirement, and Incumbent Quality in U.S. House Elections." *The Journal of Politics 72:* 178-90.

Stonecash, Jeffrey M. 2008. *Reassessing the Incumbency Effect.* Cambridge (UK): Cambridge University Press.

Uhr, John. 1999. "Three Accountability Anxieties: A Conclusion to the Symposium." *Australian Journal of Public Administration* 58: 98-101.

U. S. General Accounting Office. 1996. "Executive Guide: Effectively Implementing the Government Performance and Results Act." GAO/GGD-96-118:56pp.

U. S. Government Accountability Office. 2005. "Performance Budgeting: Part Focuses Attention on Program Performance, But More Can Be Done to Engage Congress." GAO-O6-28.

U.S Special Inspector General for Iraq Reconstruction. 2009. *Hard Lessons: The Iraq Reconstruction Experience*. Special Inspector General, Iraq Reconstruction.

Verkuil, Paul R. 2007. *Outsourcing Sovereignty: Why Privatization of Government Functions Threatens Democracy and What We Can Do About It*. New York: Cambridge University Press.

Vining, Aidan, and Steven Globerman. 1999. "A Conceptual Framework for Understanding the Outsourcing Decision." *European Management Journal* 17: 645-654.

Walker, David M. 2004. "GAO Answers the Question: What's in a Name?" *Roll Call.* July 19: 3.

Walker, Wallace Earl. 1986. *Changing Organizational Culture: Strategy, Structure, and Professionalism in the U.S. General Accounting Office. Knoxville*: University of Tennessee Press.

West, William F. 1995. *Controlling the Bureaucracy: Institutional Constraints in Theory and Practice*. Armonk, NY: M.E. Sharpe.

White House, Office of Management and Budget. 2010. "Backgrounder: The Accountable Government Initiative: The Obama Administration's Effort to Cut Waste and Modernize Government."

Wildavsky, Aaron B. 1988. *Searching for Safety*. New Brunswick, NJ: Transaction Publishers.

Williams, Daniel W. 2003. "Measuring Government in the Early Twentieth Century." *Public Administration Review* 63: 643-659.

Wilson, James Q. 1989. *Bureaucracy: What Government Agencies Do and Why They Do It*. New York: Basic Books.

Melvin J. Dubnick

Melvin J. Dubnick is Professor of Political Science at the University of New Hampshire and Professor Emeritus at Rutgers University. A former managing editor of *Public Administration Review* and coeditor of *Policy Studies Journal*, he has been writing on issues related to accountability for three decades. A Fellow of the National Academy of Public Administration, he recently co-edited (with H. George Frederickson) *Accountable Governance*.

H. George Frederickson

H. George Frederickson is the Edwin O. Stene Distinguished Professor of Public Administration at the University of Kansas and a Fellow of the National Academy of Public Administration. He is President Emeritus of Eastern Washington University. He is the founding editor of the *Journal of Public Administration Research and Theory*, the author of *The Spirit of Public Administration* and coauthor of *The Public Administration Theory Primer,* and *Measuring the Performance of the Hollow State*.

NATIONAL ACADEMY OF PUBLIC ADMINISTRATION®

Kettering Foundation