Attempting to Answer the "How?" Question:
What Causal Behaviors Could Connect The Behaviors of Public Executives To Improvements in Organizational Performance

Robert D. Behn
Kennedy School of Government, Harvard University
79 John F. Kennedy Street, Cambridge, MA 02138
617-495-9874 redsox@hks.harvard.edu

A Paper Prepared For
The Eleventh Public Management Research Conference
Hosted by the Maxwell School of Syracuse University
June 2-4, 2011

"To explain is to provide a mechanism,
to open up the black box and show the nuts and bolts,
the cogs and wheels of the internal machinery."
Jon Elster[1]

Abstract:

How could a collection of leadership behaviors — such as those engaged in by public executives who are employing the PerformanceStat leadership strategy — affect the thinking and actions of middle managers, front-line supervisors, front-line employees in a way that improves performance? This paper offers hypotheses about the impact of sixteen different causal behaviors that could provide an explanation for the interactions among individuals and aggregates of individuals created by a PerformanceStat leadership strategy — interactions that could improve performance.

Definition of PerformanceStat

A jurisdiction or agency is employing a PerformanceStat leadership strategy if, in an effort to achieve specific public *purposes,* it holds an ongoing series of *regular, frequent, integrated meetings* during which the chief executive and/or the principal members of the chief executive's leadership team plus the director (and the top managers) of different subunits use *current data* to *analyze* specific, previously defined aspects of each unit's past *performance;* to provide *feedback* on recent progress compared with *targets;* to *follow-up* on previous decisions and commitments to produce *results;* to examine and *learn* from each unit's efforts to improve *performance;* to identify and solve *performance-deficit* problems; and to set and achieve the next *performance targets*.[2]

Examples of PerformanceStat include the New York Police Department's CompStat, Baltimore's CitiStat, and the Los Angeles County Department of Public Social Services DPSSTATS.

Does any new policy work? Before one agency or jurisdiction decides whether to replicate a policy employed by another, it wants to know *whether* it works? Fortunately, social science has a method for answering the policy maker's "Whether?" question: the double-blind, randomized, (placebo-controlled) experiment.

Sure, this experiment can be expensive and time consuming. It requires that the policy be clearly specified. It requires that the policy can be implemented for some people and denied to others. It may be difficult to collect the data. It may take a long time to have an impact. Nevertheless, if designed carefully, the experiment can determine *whether* the policy works or not. Then, if the policy works, and if its key, causal characteristics can be clearly specified — that is, the knowledge of the policy is explicit — then other jurisdictions and agencies can faithfully replicated the policy.

The Search for Causal Behaviors: Answering the "How?" Question

"How?" however, does any leadership strategy — including PerformanceStat — work? What is the theory — if not explicit, then at least implicit; if not directly mechanistic, then at least indirectly causal — behind PerformanceStat? *How* might a PerformanceStat strategy help to produce better results? What exactly is the potential causal connection between the behavior of a chief executive and his or her leadership team and the results they are trying to produce?[3]

For a public executive — someone who seeks to produce specific results by implementing a leadership approach employed in another organization — answering the "How?" question is essential. For this executive's circumstances are different, and thus the formal operational machinery of another executive's leadership strategy may not have the same causal effect on employees, subunits, and collaborators, and thus will not automatically produce similar results. Unless, the public executive understand's how another's leadership strategy causes improved performance, he or she will be unable to figure out how to adapt the strategy's leadership behaviors to new circumstances.

Unfortunately, social science lacks an experimental method to answer the public executive's *"How?"* question. Even if the experiment answers the "Whether?" question with a "yes," it usually fails to answer the question posed by Mario Bunge, the McGill physicist whose research has focused on the philosophy of science: "How does it work?" The answer to the "How?" question establishes what Bunge calls "the mechanism that makes the system in question tick."[4] To Bunge, "a mechanism is one of the processes in a concrete system that makes it what it is."

For a policy, a mechanism would provide an explanation that connects key features of the policy to its consequences. It establishes (a) *what* features of the policy have an effect (either deterministically or stochastically) and what ones do not, and (b) *how* these effective features — individually and interactively — contribute to the resulting effects.[5]

Similarly, for a leadership strategy, a mechanism would provide an explanation that connects key behaviors taken by the leadership team to the consequences they are seeking to produce. It establishes (a) *what* leadership behaviors have influence (either deterministically or stochastically) on the actions of organization and what ones do not, and (b) *how* these leadership behaviors — individually and interactively — contribute to these actions. Here the "organization" would refer not only to the formal, governmental entity and the people who work in them, but also to all of those other individuals, informal collections of individuals, formal organizations, and functioning collaboratives whose behavior can influence the desired results.

The Missing "How?" in Medicine and Science

In medicine, answering the "How?" question about a proven treatment is interesting but not essential. All that is required is that the treatment can be specified explicitly. For example, for the polio vaccine, it

is only necessary to specify the components of the vaccine plus the timing and procedures for its use. It would be great to understand how the polio vaccine works, but an M.D. who prescribes it need not understand the science. All the doctors need to know is that they are using in an identical way the identical vaccine that was proven effective in the experiment.[6] The production of the vaccine is based on explicit knowledge, and thus any lab with the required technical competences (people and equipment) can replicate the vaccine. The delivery of the vaccine is also based on explicit knowledge. All very straightforward.

In science and engineering, too, it often isn't necessary to answer the "How?" question. Sir Isaac Newton's law of gravitation was both mathematically precise and operationally useful:

$$F = G\frac{m_1 \times m_2}{r^2}$$

Where F is the force of gravitational attraction between two objects with masses m_1 and m_2 that are separated by the distance r. G is the universal gravitational constant.

Actually G is simply the empirically measured fudge factor that makes the equation work. It connects the three key characteristics of the two objects to the magnitude of the gravitational force that attracts them to each other. There exists, however, no theoretical justification for a specific value of G, and Newton offered none. Indeed, he didn't even calculate a value; that was done by another British scientist, Henry Cavendish. Even today, "the historic difficulty" in specifying G is illustrated by "the inconsistencies among different measurements," which created an need to adjust the 2002 adjustment of G producing the current, 2006 adjustment of 6.67428×10^{-11} $m^3 kg^{-1} s^{-1}$ (just in case you were wondering).[7]

Scientists may find their inability to accurately measure the value of G to be frustrating, and their inability to specify it theoretically disappointing. Still, none of this invalidates the practical utility of Newton's formula. To use it, scientists and engineers do not need to know *how* gravity works.[8]

For public executives, however, answering the "How?" question is absolutely essential. There is no leadership vaccine,[9] no leadership formula. Any leadership strategy involves multiple behaviors each of which is difficult to specify (as are the timing and proficiency of its use). There is no single "treatment" — nothing that can be as easily specified (or as routinely delivered by trained people) as a vaccine or a pill. Moreover, the various behaviors comprising a leadership strategy do not function individually but interact with each other (in ways that may not be clear). Finally, any specific leadership behavior may have different effects with different people, in different circumstances, and in the pursuit of different purposes. When developing a leadership strategy one of the challenges is figuring out which behaviors will have the desired effects on which people, in which circumstances, in the pursuit of which purposes.[10]

Consequently, unless a policy, program, or strategy consists of a well specified components — like a vaccine — that can be implemented simply and consistently in a wide variety of situations, those who seek to adopt it must be able to answer Bunge's question: "Howdoes it work?" Even if an only slightly complicated policy, program, or strategy has produced a definitive "yes" (or even a tentative or conditional yes") in the *whether* experiment, implementing it elsewhere still requires an understanding of the causal *how.* For employing this successful policy, program, or strategy in a different circumstance, organization, or political environment will necessarily involve not just adoption but also some adaptation — perhaps some significant adaptation.

Thus, a public executive who seeks to employ another's leadership strategy needs to understand *how* this strategy works. This executive needs to understands *how* the strategy's leadership behaviors effect the behavior of people that, in turn, produce the desired consequences. Otherwise, this executive will be unable to get his or her organization to "tick" effectively.

The Search for Mechanisms

Scientists and philosophers have long sought both to identify both physical and social mechanisms and to explain how they work. "Science explains in terms of mechanisms," writes Bunge. And recently, observe Peter Hedström and Petri Ylikoski, "mechanisms and mechanism-based explanations have attracted a great deal of attention in the social sciences as well as in the philosophy of science."[11]

Mechanisms, report Hedström and Richard Swedberg, "usually are unobserved analytical constructs . . . that provide hypothetical links between observable events."[12] "A mechanism is an irreducibly causal notion," argue Hedström and Ylikoski; it "provides a how-possible explanation; it tells us how the effect could in principle be produced."[13]

In an effort to explain evolution, Charles Darwin proposed his theory of natural selection containing two important mechanisms: (1) inheritable variations, combined with (2) the struggle for survival. In an effort to explain the wealth of nations, Adam Smith proposed his theory of the invisible hand, for which key mechanism was every individual's desire to be wealthier. In both of these cases, the mechanisms are intentional but individual; each individual seeks survival or wealth develops a strictly personal strategy for achieving it. Thus, the phenomenon that is explained results from the aggregate but uncoordinated impact of multiple individuals carrying out their own, individual intentions; both mechanisms work without anyone's calculated effort at guidance.

Leadership behaviors are also intentional. But the leaders seek to achieve their intentions not only directly through their own actions but also indirectly by consciously attempting to influence the actions of others. The leaders' intention is not to personally survive or become personally wealthier but to achieve some broader purpose that benefits others, including (at least in the case of PerformanceStat) others whom the leadership behaviors are not necessarily trying to influence. The leader is attempting to accomplish a specific result by making a deliberate, conscious, calculated attempt to effect the behavior of others. Thus, there are two steps in the causal explanation: (1) How do specific leadership behaviors have an impact on the thinking and actions of others? and (2) How do these others think and act to help achieve the desired result? Here I seek to identify the first causal mechanism(s): How do the multiple behaviors of a PerformanceStat leadership team explain what individual and organizational reactions?

If the PerformanceStat leadership strategy is to have any validity, there must exist at least one such mechanism connecting leadership behavior to the thinking and actions of others. After all, as Hedström and Ylikoski write, "the absence of a plausible mechanism linking X to Y gives us a good reason to be suspicious of the relation being a causal one."[14] If we cannot identify some plausible explanations about how the behaviors of a PerformanceStat leadership strategy can have an impact on individuals and organizations, we would need to conclude that the strategy has no impact.

The Clunky Metaphor of the Black Box

To understand how a PerformanceStat leadership strategy might produce better results, I have tried to look inside Jon Elster's black box — to identify the "nuts" and "bolts" and to understand the "cogs" and "wheels" of this "machinery." Of course, this metaphor is all wrong. There are no nuts or bolts, no cogs, no wheels. Indeed, there is no machinery — though the analogy is attractive — indeed, reassuring.[15]

After all, the metaphor suggests not only that there exists some machinery but also that, if we are clever enough, our mind's eye can penetrate the blackest of these mechanistic boxes. It permits us to infer that if a leader is only brilliant enough, he or she can somehow (perhaps with the help of some super-power x-ray vision) discern the organization's machinery chugging away, cogs whirling and wheels spinning. Moreover, it permits us to infer that if this brilliant leader spends enough time thinking — and if the leader understands the applicable formulas of mechanical engineering — he or she can figure out what the nuts and bolts hold together and how the turning of specific cogs and the rotating of particular wheels directly affects the thinking and actions of different people.

Indeed, behind the machinery metaphor lies the hint that really smart and resourceful leaders — by selecting carefully and using properly the appropriate nuts, bolts, cogs, and wheels — can create higher performing black boxes. Or, when they are stuck with some organizational machinery designed by others who are less prescient, they can still figure out how to adjust with the cogs and wheels to make it function better. Indeed, all this is reminiscent of Max Weber's description of a bureaucracy as a machine whose "rational calculation . . . reduces every worker to a cog in this [bureaucratic] machine and, seeing himself in this light, he will merely ask how to transform himself from a little into a somewhat bigger cog."[16] In an organizational black box, the humans are the cogs.

Yet, despite our human propensity for mechanical metaphors, we cannot really open up and watch the cogs of a PerformanceStat machine — or any other leadership strategy — in action. For one thing, there is no leadership machine. And to the extent that we are in love with this metaphor, we still need to accept that this machine does not all exist inside a single box — be it black or green or pink. For example, the CompStat "machine" — if you insist on imagining it in this way — is spread all over New York City (and into adjacent jurisdictions as well) and includes not only what goes on inside the "black box" of the New York Police Department. For crime everywhere is affected not only by the individual and collective behavior of the police, but also by the personal behavior of individual citizens and by their collective behavior in their neighborhoods. If there exists a black box for NYPD's CompStat, it is a very large box with millions of cogs and wheels rotating and counter-rotating each other in very complex and obscure ways.

CompStat — indeed, any PerformanceStat — is not a machine. Rather it is a complex, adaptive system[17] — a very complex and continuously adaptive system with multiple feedback loops. Thus, leadership's challenge is not to build some new, fabulous machine. Rather it must work with the existing system — injecting some conscious purposes, creating some specific targets, inspiring with some public recognition, devoting time while refusing to go away . . . . And, all the while, analyzing data, asking questions, scrutinizing reports, in an effort to learn whether these leadership behaviors are introducing new feedback loops that foster adaptions that help to achieve the purposes.

Leadership Behaviors as Plausible Explanations

Thomas Schelling proposes that "a social mechanism is a plausible hypothesis, or a set of plausible hypotheses, that could be the explanation of some social phenomenon, the explanation being in terms of interactions between individuals and other individuals, or between individuals and some aggregate."[18] No nuts or bolts here. No cogs or wheels. No machinery. Just a "plausible hypothesis," or a "set of plausible hypotheses," that "could be the explanation." *"Could be."*

In this spirit, I offer a set of sixteen[19] different hypotheses that could be an explanation for the interactions created by a PerformanceStat leadership strategy that improve performance. Each hypothesis describes a specific interaction between the members of an organization's leadership team and the middle managers, front-line supervisors, front-line employees. Each hypothesis seeks to explain how these interactions change the thinking, conduct, practices, and actions — both of individuals and of the aggregate — that could, in turn, affect organizational performance. Each is one item in what Hedström and Ylikoski call "the toolbox of possible causal mechanisms"[20]

Of course, not even sixteen causal explanations can capture all of the subtle qualities of a leadership strategy let alone its multiple possible effects. Each of the sixteen is but one possible analytical construct. Yet, note Hedström and Swedberg, this is both necessary and desirable:

> It is through abstractions and analytical accentuation, however, that general mechanisms are made visible. But these abstractions also distort by their very nature the description account of what actually happened, by accentuating certain aspects of the situation and ignoring others.[21]

Thus, the only real issue with this (or any other) abstraction is whether what it accentuates (and what it ignores) helps us understand the causal features of the strategy — and thus can help others adopt and adapt their own approach.

Unfortunately, although we can observe an organization's leadership team engaging in these behaviors, we cannot observe *how* these behaviors are affecting others in the organization. This is disappointing but predictable. "Most of reality is unobservable," argues Bunge, who applies this to mechanisms: "Typically, mechanisms are unobservable, and therefore their description is bound to contain concepts that do not occur in empirical data." Indeed, he continues, "most mechanisms are concealed, so that they have to be conjectured."[22]

We can conjecture that "holding meetings can focus everyone's attention on what is most important," but we can't observe any individual's attention being focused (or diffused) by a meeting. We can speculate that "scrutinizing the positive deviants can facilitate everyone's learning," but we cannot peer inside any single individual's brain to observe any such learning. Every one of the individual and organizational reactions that *could be* explained by these sixteen leadership behaviors cannot be directly observed; or if this reaction were to have a visible manifestation, we are still forced to conjecture that the leadership behavior (and not some other influence) provides the explanation.

These hypotheses emerge from my visits to various PerformanceStats, where I have observed a variety of different leadership behaviors that, when employed consciously and creatively, *could* help to improve the results produced by any organization. Given what we know about how humans function, interact, and respond in organizations, these causal behaviors could help us understand *how* the various opaque, enigmatic, ambiguous, and contingent strategies that fit the definition of "PerformanceStat" might influence people in such a way to help improve performance. Specifically, I think that the leadership team of an effective PerformanceStat strategy would adapt many (though maybe not all) of approximately sixteen[23] different leadership behaviors that *could* have a causal effect on people and organizations, and thus on results.[24]

I have observed multiple leadership teams from public agencies and governmental jurisdictions employing each of these behaviors. And I have observed the conduct and culture of subunits, their middle managers, and (sometimes even) their front-line supervisors and front-line workers. Have each of these sixteen always induced precisely the desired changes in the application of formal procedures and in the human effort behind informal practices? Of course not. Sometimes there is resistence, push back, even insubordination. Still, by reproving the recalcitrant while remaining persistent, the insolent and mutinous can be isolated. And by recognizing accomplishments, others can be recruited to the task of producing the desired results.

For any of these behaviors to be effective, the leadership team needs to possess the personal and institutional capacity to undertake the necessary actions, to engage in the essential practices, and to send the specific signals that will induce (though never compel) individuals, teams, and agencies to establish their own performance-focused culture and to engage in the kind of day-to-day results-producing activities that the leadership team seeks.

Sixteen Plausible Hypotheses for Explaining the Impact of PerformanceStat

In particular, the leadership team that implements a PerformanceStat strategy can influence department heads, middle managers, front-line supervisors, and front-line workers to improve performance by consciously and consistently employing the following potentially causal behaviors:

Plausible Hypothesis A: Reiterating the purpose can keep everyone focused on the big picture.

Plausible Hypothesis B: Analyzing data can reveal significant performance deficits.

Plausible Hypothesis C: Creating targets can specify exactly what needs to be accomplished by when.

Plausible Hypothesis D: Making operational assignments can define who needs to fix which performance deficits next.

Plausible Hypothesis E: Devoting resources and time to PerformanceStat can dramatize the chief executive's personal commitment to improving performance.

Plausible Hypothesis F: Holding meetings can focus everyone's attention on what is most important.

Plausible Hypothesis G: Requesting frequent reports on progress can ensure that targets and assignments are taken seriously.

Plausible Hypothesis H: Asking questions of individual subunit managers can promote personal responsibility.

Plausible Hypothesis I: Following up frequently on targets and assignments can create the feedback that can suggest adjustments.

Plausible Hypothesis J: Distributing comparative data widely can help every team appraise, without delusions, its own performance.

Plausible Hypothesis K: Scrutinizing the positive deviants can facilitate everyone's learning.

Plausible Hypothesis L: Recognizing accomplishments can confirm that success is possible and valued.

Plausible Hypothesis M: Reproving the recalcitrant can get everyone's attention.[25]

Plausible Hypothesis N: Telling stories can foster a results-focused culture.

Plausible Hypothesis O: Abetting everyone's implicit evaluation of everyone else can breed individual and team motivation.

Plausible Hypothesis P: Remaining persistent can demonstrate that this isn't going away.

All sixteen of these are *personal behaviors* of the members of the leadership team. They are not systems. They do not include Luther Gulick's seven "functional elements" or "major duties of the chief executive": planning, organizing, staffing, directing, coordinating, reporting, and budgeting" or "PODSCORB."[26] Gulick's administrative functions are certainly important. Directly, however, they contribute little to the organization's performance. Indeed, in many organizations today, these administrative functions have deteriorated into exercises in compliance — dominated by the need to follow formal procedures and abide by reporting requirements — with little attention paid to how they might (or might not) effect actual results.

Emphasizing Purpose; Motivating People

The first four of these sixteen behaviors are designed to dramatize the purposes to be pursued, to specify the results that to be produced, to convey the improvements to be made, and to highlight the performance that needs to be enhanced. The other twelve are designed to motivate individuals, teams, and agencies to make these improvements.[27]

That is, the latter twelve behaviors are designed to motivate people not merely to *work harder* (though for some front-line supervisors and workers that might be the major concern). These twelve behaviors are also designed to encourage individuals, teams, and agencies to experiment with ways to *work creatively* by experimenting with different tactics and by reallocating resources. And they are designed to encourage

everyone to *work differently* — to replace their rule-obsessed behavior with a performance perspective; to get them to focus on accomplish their organization's purpose by achieving their targets.

None of these sixteen behaviors could be sufficient to cause a significant improvement in performance. Nor do these behaviors function independently. Rather, they interact with and reenforce each other. Collectively they serve to ensure that everyone in the organization recognizes both what needs to be done next and what his or her responsibility is for getting it done. All serve to establish and maintain an emphasis on results and performance.

Underlying the first four behaviors is the assumption that the jurisdiction or agency is, indeed, capable of improving its performance. Underlying the other twelve behaviors is the assumption that the leadership team can affect the behavior of subunit managers, middle managers, and (perhaps even) front-line workers. Underlying all sixteen behaviors is the assumption that the leadership team possess the intellectual and interpersonal competencies to implement them.

None of these sixteen behaviors is new. Each has been employed by numerous leaders, in numerous organizations, in numerous circumstances. What makes a public-sector PerformanceStat strategy worthy of our attention is that its leadership team employs them creatively, consistently, coherently, and constantly.

Necessary? Sufficient? Helpful?

If the members of an organization's leadership team who seek to improve performance engage in all sixteen of these leadership behaviors, will performance improve? Maybe. Maybe not. These behaviors are neither individually or collectively, directly or exclusively, deterministic. After all, a multiplicity of factors affect an organization's ability to produce results — many of which the leadership team is unable to influence. Over several years, these managers might be able to make some progress. But, for example, without the support for their superiors, or the flexibility in the deployment of resources, or the cooperation of powerful interests, this progress may be slow. Certainly, even when employed creatively and with diligence, these sixteen leadership behaviors will not be sufficient in all circumstances to create large jumps in performance.

Alternatively, what would happen if an organization's leadership team — in seeking to produce better results — engaged in none of these sixteen behaviors? This counterfactual suggests the relevance of these causal behaviors.[28] For, without all of these leadership behaviors, we would most likely attribute any improvement in performance to outside or random effects. Given what we know about the behavior of humans in organizations, managers who eschew all sixteen of these causal behaviors are unlikely — even if they build a fancy room with the latest technology— to produce better results.

What if a leadership team employed fifteen of the sixteen? Would that improve performance? Maybe. Maybe not. As always, it all depends.[29] Indeed, the answer could depend upon a lot of things — for example, how well the organization was performing before the leadership team decided that they should seek to achieve better results.

Individually and collectively these sixteen behaviors are not sufficient. Still, maybe some of them are necessary. But which ones? Or which ones in what combination? This is not obvious.

Moreover, the quality of each behavior counts. How often do public managers need to "reiterate its purpose" to have an impact on performance? In what ways? This obviously depends upon how well the organization has already internalized this purpose. How many questions do public managers need to ask to get an individual subunit manager to take personal responsibility for subunit performance? Again, it all depends. It depends, for example, upon how much personal responsibility each subunit manager has already assumed? Some may need few (or even no) questions; others may require repeated and pointed questions.

Or suppose that the members of a leadership team only pretend to employ these behaviors. For example, supposed they ask only trivial or irrelevant questions of subunit managers. Or suppose that they are completely incompetent — that they are unable to seriously reiterate their purpose, or to create relevant targets. Or suppose that their persistence consists solely of holding a regular series of ritualistic *pro forma* meetings?

There might be other causal behaviors of leadership that are not on this list of sixteen. Nevertheless, these sixteen do appear to be individually and collectively important. If an organization's managers do not engage in any of them, we would not expect performance to improve, but rather to deteriorate.

Maybe none of these leadership behaviors are absolutely necessary. Certainly none is sufficient. Individually and collectively, however, the sixteen could be very helpful.

Hypothesis A: Reiterating the purpose can keep everyone focused on the big picture.

> By repeating and repeating the public purpose that the organization is responsible for achieving, the leadership team can ensure that, in their pursuit of detailed tasks and specific targets, people do not forget their overarching mission.

The regular PerformanceStat meeting provides a perfect opportunity to remind everyone — yet once again — the organization's mission and the purposes it is trying to accomplish. Too often in large organizations, the daily tasks and operational routines can displace the real, underlying purpose that these tasks and routines are designed to help achieve. Consequently, one of the major challenges of leadership is to ensure that people do not lose sight of the ultimate objective.

Fortunately, one of the core features of this leadership strategy — the regular and frequent meetings — creates a convenient occasion for reiterating the purpose. Effective leaders take every opportunity to ensure that no one engages in, as Nietzsche suggested, the frequent, stupid act of forgetting our objectives.[30]

Reiterating the purpose is, however, more than a defense against forgetfulness. It has a quite positive consequence. For as Richard Hackman has observed: "People seek purpose in their lives and are energized when an attractive purpose is articulated for them."[31] A public organization — almost by definition — has an attractive purpose, perhaps multiple attractive purposes. Reminding everyone of these purposes can, indeed, energized them.

Hypothesis B: Analyzing data can reveal significant performance deficits.

> *By analyzing very current performance data, the leadership team can discern and highlight the performance deficits that the organization needs to eliminate or mitigate.*

Lots of public agencies have lots of performance measures. Dozens of measures. Hundreds of measures. Perhaps thousands of measures. A look at the measures produced by some public agencies, it would appear that: The more measures the better.

As a reporting strategy, this may make some political sense. Some measures will suggest improved performance; for this, you can take credit. Some measures will suggest a deterioration in performance; for this, you can blame others. Then, next year, when different measures improve, you can say: "See, we worked on those and they got getter." As for the ones that got worse, there exist multiple others to blame.

To actually improve performance, however, an organization needs to focus — to focus on fixing the problems that are preventing it from producing better results. No organization can simultaneously produce the results necessary to create improvement on a thousand different measures. Not even on a hundred different. Maybe ten. Most likely not. For most public agencies, attempting to simultaneously

ratchet up performance on ten different dimensions is simply too many. So leaders have to choose. They have to decide which one or two performance deficits need to be fixed next and then set the targets to get them fixed. Indeed, Benjamin Franklin argued that two was too many: "He that pursues two hares at once does not catch one and lets the other go."[32]

So which measures are really important. Which measures need to be improved next. Which measure can reveal — through an increase or decrease — whether or not the organization has eliminated — or, at least, mitigated — one of its key performance deficits?

Time to call out the calvary: the analysts and their data. Actually, you not need an entire regiment. For a small agency, one skilled analyst can work wonders. Larger organizations with multiple responsibilities will need more, but not a regiment. Initially, at least, a small squad will do quite fine.

Give them a large dump of data, and ask: "What are our different performance deficits?" These analysts may not possess the overall, organizational perspective necessary to select the performance deficits on which the organization should focus next. But they should possess the analytical skills necessary to nominate candidates, and perhaps even to suggest strategies for fixing each one.

Identifying performance deficits is an analytical task. Choosing the one(s) on which to focus next is a leadership responsibility.

Hypothesis C: Creating targets can specify exactly what needs to be accomplished by when.

> *By evolving, in consultation with subunit managers, specific, challenging, consequential performance targets for each subunit to achieve, the leadership team can not only identify the results to be produced but also motivate agencies, individuals, and teams to produce them.*

Targets can be inspiring. They can challenge human egos. Targets create very concrete purposes. They aren't merely lofty yet amorphous aspirations. They are very specific. They contain a task that can be completed — and thus can generate a feeling of accomplishment. To the extent (as Hackman suggests) that an attractive purpose energize people, an attractive purpose articulated in a specific target can energize people to undertake the tasks necessary to achieve that target.

To accomplish something, a PerformanceStat leadership team needs to change the behavior of individuals and units — to motivate improvements in both individual and collective performance. Targets — particularly targets for teams — can, if structured well, can do precisely that. That is: both individuals and teams perform better if they are challenged by specific targets. Over the last 40 years, a body of research — primarily the work of Locke and Latham — have revealed that setting targets has a significant impact on performance.[33] About the accumulation of evidence that Locke, Latham, and their colleagues report on the effect of "goal setting" on organizational performance,[34] Jack Duncan of the University of Alabama writes: "No other motivation technique known to date can come close to duplicating that record."[35]

Thus, one motivational tactic of PerformanceStat as employed, for example, by Baltimore's CitiStat is the use of performance targets with four attributes:

(1) *Specific* Targets. These targets create an unambiguous standard of performance for an agency and can create such standards for its sub-units.

(2) *Challenging* Targets. If an agency gets to pick its own target, it may well choose something that is easy — perhaps very easy — to achieve.[36] To motivate better performance, however, requires targets that are challenging. Moreover, report Stephen Harkins of Northeastern University and Richard Petty of Ohio State University, people are more likely to try hard when the task is difficult than when it is easy.[37]

(3) *Consequential* Targets. Targets are particularly relevant if they based on real citizen concerns. In Baltimore, many targets are specific requests that citizens individually create by calling 311. Thus, although managers and employees can debate the citizen requests for which they should create targets and the particular level at which the target should be set, neither managers nor employees can argue that the targets are disconnected from real citizen concerns. People will try harder if the target to be achieved (not matter whether it is difficult or easy to achieve) possesses some social significance.[38]

(4) *Agreed Upon* Targets. Certainly, both the performance measures for which a target should be set and the level at which it should be set were heavily influenced by the mayor and the CitiStat staff. Nevertheless, at least after those agency heads who failed to accept the CitiStat strategy had left city government, the nature of the targets was open to continuous renegotiation between the agency head and the mayor's staff.

Performance targets will not, by themselves, improve performance. Many public agencies establish performance targets and yet fail to pursue them with energy and intelligence.[39] This causal behaviors needs support from others — including operational assignments.

Hypothesis D: Making operational assignments can define who needs to fix which performance deficits next.

By developing an ongoing series of operational tasks that are directly linked to the organization's purpose and its targets, the leadership team can focus the efforts of subunit managers and front-line employees on the next steps for eliminating or mitigating specific performance deficits.

Some performance targets reflect daily operational assignments: In Baltimore, for example, one such target is: Fill all potholes within 48 hours. Data revealing how the organization is doing can be collected frequently, weekly or even daily.

Other performance targets, however, specify what is to be achieved over a longer period of time: In 1975 in New York City, for example, William Bratton sought to reduce crime by 40 percent by the end of three years.

For targets at either ends of this spectrum (and in between too), there are performance deficits that must be fixed before the target can be achieved. To be able to fill every pothole in 48 hours, a city's public works department may need new equipment. It may need to fix its operational routines. Once all of the relevant managers and units agreed to this target, they will begin the analytical process of figuring out what specific performance deficits are preventing them from achieving this operational standard. Then, having identified what needs to be fixed, they will assign a individual, team, and other subunit to fix each. The link between performance target and performance deficit may be quite clear, as can be who should get the task of eliminating or mitigating it.

Some targets, however, are not so directly operational. Yes, to achieve the target, the organization's operations have to be effective. But in many cases, the organization doesn't produce the result that the target measures. Police departments produce neither crime nor the reduction in crime. Police departments do things that directly or indirectly help to reduce crime.

But *who* in the police department should do *what* by *when* (this month, this week, today?) to ensure that after 365 days reported crime is down 10 percent. This is not obvious. It may not be obvious what should be done? And even if the task is obvious, who should complete the task may not be?

Thus setting a macro performance target for an organization — and even monitoring it frequently — is not enough. The organization has to analyze its data to figure out what its key performance deficits

are. Then, it needs to determine the specific tasks that, when completed, will eliminate or mitigate these key deficits. Finally, it needs to give specific people these specific tasks to be completed by specific dates.

Hypothesis E: Devoting resources and time to PerformanceStat can dramatize the chief executive's personal commitment to improving performance.

> *By investing significant resources in an analytic staff and assigning these individuals and other members of his or her leadership team to spend significant time on PerformanceStat, a chief executive demonstrates a serious, personal commitment to improving performance.*

John Ehrlichman, President Nixon's domestic policy advisor, worried that cabinet secretaries, once appointed, would fail to remain loyal to the president's agenda. "We only see them at the annual White House Christmas party," observed Ehrlichman; "they go off and marry the natives."[40] This organizational parochialism of any sub-unit's managers is not limited to the Nixon Administration, to a national government, or, indeed, to a governmental jurisdiction. This is a common challenge in any jurisdiction or agency whose the chief executive has multiple and diverse responsibilities: How does the chief executive get everyone to understand what he or she really wants done?

Sure: The chief executive can write and distribute memos, post a mission statement on the wall of every office, and give speeches. But who pays attention? Who *really* pays attention? Who thinks these communications represent the real concerns of the chief executive (and not the nice sounding words of some intern assigned to the communications office)?

What people really want to know is: Where is the chief executive investing discretionary, allocable resources? In most public organizations, most resources are fixed. Even the elected chief executive cannot easily move them to other purposes. Still, some are. And people want to know what the chief executive is doing with the few scare resources over which he or she has some discretion.

Thus, by investing significant resources in building a talented, dedicated PerformanceStat staff, by purchasing the necessary supporting technology, by constructing a special room for PerformanceStat meetings, and establishing a series of regular, frequent, integrated PerformanceStat meetings, by requiring subunit managers as well as key members of his or her leadership team to participate in these meetings with each agency, a chief executive is sending a clear signal that people can't miss: "This is important." For example, for over a decade, a succession of three mayors have, through the rigorous routine of Baltimore's CitiStat's meetings, dramatized that performance was high on their agenda and thus needed to be high on the agenda of every city agency.

Moreover, everyone in a large organization has a simple test for determining whether the chief executive cares about an issue, a task, a policy, a program:

> If and only if a chief executive spends the most valuable resource of any public official — personal time and the personal time of his or her top staff — does the executive care.

If an executive does not invest any time in an issue or agency, everyone in the government quickly figures out that — despite all of the rhetoric — the executive does not really care.

After all, every public executive — indeed, every human — has exactly 168 hours in every week: no more, no less. And, at midnight on Saturday, those 168 hours are spent — gone. No one can put any of those hours in the bank for another week; no one can loan of a few hours to a friend or colleague. Every week, every human has precisely the same time budget to spend: 168 hours.

Consequently, how an executive spends his or her time — and the time of the organization's leadership team — sends unambiguous signals about what he or she thinks is important. If you want to

know what an executive thinks is important, don't read the memos or listen to the speeches. Just watch how the executive spends his or her most valuable resource: time.

Alfred Ho of the University of Kansas analyzed the value that midwestern mayors place on performance measurement and observed: "If elected executives and legislators are not interested in performance measurement, managers may easily lose their enthusiasm about the tool because they do not see any payoff from their time investment and effort to collect and report the data."[41] Over two decades in Baltimore, not all of the agency managers have necessarily been enthusiastic about the city's performance measures, but they had no doubt about their three mayors' commitment to CitiStat.[42]

To make a PerformanceStat strategy work, the leadership team's commitment has to be high indeed.

Hypothesis F: Holding meetings can focus everyone's attention on what is most important.

> *By holding an ongoing series of regular, frequent, integrated meetings, the leadership team can keep everyone focused both on their important macro purposes and on the essential actions and activities necessary to achieve these purposes.*

No one likes meetings. Yet, every organization has them. For such meetings serve a variety of purposes: they resolve differences, establish priorities, make assignments, check on progress. . . . The list of reasons for holding meetings is long — and so is the list of meetings that anyone has to attend.

Thus, the performance-focused manager doesn't schedule meetings recklessly. After all, by holding a meeting, the organization is consuming one of its most valuable resources: the time of its key people — time that these people could devote other purposes.

By regularly scheduling PerformanceStat meetings, by ensuring that his or her top aides attend these meetings, and by faithfully participating in them, the chief executive of an agency or jurisdiction is sending a powerful message: This is really important.

In Washington, D.C. under Mayor Adrian Fenty, Victor Prince served as director of CapStat (the District's version of CitiStat). When Prince heard the Fenty tell a caller on his VIP Blackberry (Fenty usually carried three) that he couldn't talk because he was in the middle of a CapStat meeting, Prince realized that Fenty had concluded that his CapStat was really effective. The person who called was obviously important enough for the mayor to answer; but the CapStat meeting was even more important.

Hypothesis G: Requesting frequent reports on progress can ensure that targets and assignments are taken seriously.

> *By asking subunit managers to report regularly and frequently in a quasi-public meeting on the status of their assignments, on their progress towards achieving their performance targets, on their strategies for overcoming their performance deficits, the leadership team can remind everyone that these responsibilities are important.*

Many assignments are made, but many fewer are completed. After all, most people working in government have many people — nominal superiors, other high-ranking supervisors, legislators, interest group officers, citizens — who believe that they are able to give these people assignments. They can't complete them all.

So they triage. Perfectly reasonable. But do their triage choices make sense? Have they focused on the assignments that will help their organization fix its important performance deficits? Who knows? No one — unless someone asks for a progress report.

The need to make such a progress report next week or next month can quickly affect an individual's triage decisions. Anyone who has to report back, in front of superiors, peers, and subordinates, on whether they have completed a publicly assigned task, is likely to put that task at the top of his or her list. And, if that task is an essential prerequisite for others to be able to complete theirs, completing it becomes even more imperative.

All bureaucracies produce reports. Bureaucratic superiors use such reports to find out what subordinate units are doing and to affect their behavior. Anthony Downs argues that such reports have three purposes:

(1) "to inform high-level officials about what is happening";

(2) "to remind each subordinate that he must meet certain performance standards";

(3) to encourage subordinates "to carry out the desired performance — or at least to report having done so" through "the fear of punishment for failure."[43]

Of course, if the superiors never look at the reports (or look at them but never do anything to communicate that they have looked at them), the subordinates will quickly learn not to take the reports very seriously, which will neutralize Downs's three purposes for the reports.

An effective PerformanceStat's reporting strategy is, however, significantly different from the usual bureaucratic mechanics. Traditional reports are annual, semi-annual, or at best quarterly, appearing weeks or months after the end of the reporting period. And they come in voluminous documents, so dense that none but a dedicated data wonk will ever read them. Moreover, the sheer volume of the data inhibits anyone's ability to determine whether the organization's performance has improved or retrogressed. (The same is true for many dashboards that governments post on the Web.[44]) The theory behind bureaucratic reports may be based on Downs's three purposes. But the traditional method of implementation neutralizes the impact. "The normal way of functioning," reports an official in Finnish Ministry of Finance, "is that they receive the report and they put it on the shelf."[45] And not just in Finland.

For an effective PerformanceStat, however, reports are quite different. They are very frequent. In Baltimore, for example, they are every two weeks. They contain very current data, usually from a two-week period that ended within the last ten days. They include comparisons — with data from previous periods, with specific performance targets, and, perhaps, among similar subunits — that provide a basis for judging whether performance is improving or deteriorating. And they are personal. Agency managers make their reports personally in response to direct questions from superiors and in front of peers and subordinates. All reports — to one extent or another — help to accomplish Downs's three purposes. But Baltimore's approach heightens the impact of the reporting on each of these purposes.

"Have you completed your assignment?" Knowing that you will asked this two week from today at 8:30 in morning in a room full of people can certainly focus the mind. In Baltimore, a manager can't leave a CitiStat meeting saying: "Whew. I don't have to worry about this for another six months.

Hypothesis H: Asking questions of individual subunit managers can promote personal responsibility.

By directly questioning subunit managers and their subordinates about their unit's progress, the leadership team can establish in these individuals a sense of personal responsibility for completing their assignments, achieving their targets, and improving performance.

Data and performance deficits. Targets and operational assignments. Meetings and reports. For a PerformanceStat leadership strategy, these are all important components. None, however, guarantees that any individual — subunit director, middle manager, front-line supervisor, or front-line worker — feels

any personal responsibility for making progress on either achieving a particular performance target or completing a specific assignment.

Indeed, the leaders of governmental jurisdictions and public agencies who seek to improve performance face a challenge. If they want to produce real results, they need a way to establish *who* is responsible for *what.* Which specific individuals are responsible for which the various specific endeavors that will contribute to the desired results?

At any effective PerformanceStat session, direct questioning establishes this personal responsibility. For these questions are not raised in a private communication (memo, e-mail, or phone call). They are not are not raised in a memo that is circulated to a variety of people. At a PerformanceStat session, each question is addressed directly to a specific individual. This pointed question leaves no doubt who is responsible.

Sometimes, the question is not asked of the subunit director by of another manager of the subunit. This is possible because (at least for large organizations) the subunit is represented not just by its director but by several other managers. (In Baltimore, another two-dozen managers will be in the room.) Thus, questions can be address not only to the agency director who is standing at the podium. Actually, other managers who are called upon to answer a question can also be called upon to step to the podium.

Furthermore, these other agency managers were also in the room at the last meeting, at the meeting before that. They may have been in the room six months ago, often a year ago. The chief executive and the PerformanceStat staff will have learned each manager's name. They know what responsibilities fall within each managers' domain. They remember what questions they asked at the last meeting and at the meeting before that, and at the meeting . . . . They remember who answered which ones satisfactorily, and whose answers require some follow-up.

At effective PerformanceStats, the questions cab be quite pointed — and in two different ways. Each question can be pointed in the sense that it is focused on a very specific, very narrow issue. And each question can be pointed in that it is directed at a particular individual. Indeed, when members of the leadership team ask an individual a question, they may ask the individual to stand — to even come to the podium. When they ask a question, they leave no doubt who they think is personally responsible for answering it.

Moreover, the agency heads often facilitate the process of identifying the responsible person. An agency head may call this individual to the podium. In fact, the agency head may step aside, leaving this individual standing at the podium alone. In doing so, the agency head clearly dramatizes who he or she thinks is responsible.

The podium and the conduct of the session — with specific questions directed at specific individuals — makes the implication clear: It is not only the agency head who is responsible for every aspect of the agency's performance. Every member of the subunit's management team also has some personal responsibility — at least for those aspects of performance for which his or her unit must make a contribution. Effective PerformanceStat sessions focus attention on each individual's personal responsibility for specific aspects of the subunit's performance.

"In Western civilizations it is commonly believed that being identified as the source of one's accomplishments and errors has an important effect on performance," write Kip Williams, Stephen Harkins, and Bibb Latané. Williams and his colleagues conducted a couple of experiments in which the subjects thought that their own, individual effort could or could not be identified. They found that "when individual outputs are always identifiable (even in groups), people consistently exert high levels of effort, and if their outputs are never identifiable (even when alone), they consistently exert low levels of effort." Thus, they conclude, "identifiability seems to be an important, albeit complex, psychological variable and appears to have great implications for human motivation and performance."[46]

An effective PerformanceStat strategy explicitly and carefully seeks to create each manager's own "identifiability."

Hypothesis I: Following up frequently on targets and assignments can create the feedback that can suggest adjustments.

> *By following up regularly on targets and assignments, the leadership team can generate feedback that compares progress with objectives and thus reveals whether the current approach needs to be modified or revamped and, perhaps, what kind of changes might produce better results.*

Regular PerformanceStat meetings provide opportunity for reports and questions. They also provide an opportunity for feedback. The data, the reports, the questions and answers, plus observations by the chief executive's leadership team and the PerformanceStat staff create the basis for providing feedback to the agency about how well it is doing.[47]

Bureaucratic "rigidity" and "ossification," argues Downs, are "most likely to occur in bureaus that are insulated from feedbacks" — from "loud" complaints from "from clients, suppliers, regulatees, rivals and allies." Conversely, he argues, "the rigidity cycle is least likely to occur in bureaus that are under strong and constant pressure from such feedback."[48]

With PerformanceStat, however, each subunit regularly gets such feedback. The leadership team can convey this feedback through complaints — either from their own experience or from criticisms they have received from outside sources — or with the irrefutable data on the agency's latest performance report. Moreover, the impact of this feedback can be significant. Jeffrey Pfeffer, of Stanford Univeristy, reports "impressive evidence that feedback, even in the absence of goal setting, can impact behavior."[49]

Hypothesis J: Distributing comparative data widely can help every team appraise, without delusions, its own performance.

> *By distributing data that compares the results produced by subunits with identical or similar responsibilities, the leadership team can ensure that each subunit has an honest, realistic appreciation of how its performance compares with that of its peers.*[50]

Any large public agency divides its operational responsibilities among a variety of subunits — usually giving each subunit responsibility for a particular geographic area. In medium sized and large cities, the police department divides operational responsibilities among its different precincts (or areas or districts). Similarly, the school department divides the task of teaching students into different schools, and then in to different classes. For its various departments and agencies, the U.S. Government divides responsibilities into (eight to dozen) regions.

Does each of these subunits know how well it is performing? Has each subunit duped itself into believing that it is at least above average — if not absolutely superior? Or does it posses the information necessary to make a clearheaded, realistic assessment of how well it is doing? After all, if the managers for every subunit can able to assume that they are at least above average, they have little reason to press for improvements. If each month or each quarter, however, they get a data that compares the performance of every subunit, they will know where they stand on the organization's performance ladder. They will know which units are performing better, which aren't, and which units are making the biggest improvements.

Comparative data can be displayed in two deceptively simple yet very compelling ways:

> **The Bar Chart** is very simple. It displays accomplishment data for each unit and may include a horizontal line that defines the target for all units. Regardless of whether the target line is there, everyone immediately focuses on the tallest bar (why is this unit doing so well?) and on the shortest bar (why is this unit doing so badly?)

> **The List** is even simpler. On one piece of paper, there are two columns. The first, Column A, contains the names of the units that made their target for the previous period (year, quarter, month). The second, Column B, contains the names of the units that didn't make their targets. When "The List" is distributed at a meeting of all unit directors, behavior is very predictable. People look first for their own unit's name and then for the name of their colleagues' units.[51]

These two quite ordinary ways for providing comparative feedback about results can quickly grab managers' attention. Moreover, such a display of comparative performance data can motivate at least some managers and their units to improve.

These two feedback strategies can help to motivate the behavior of people and organizations both inside and outside of government, as long as everyone has agree on:

(a) the purpose they are pursuing;

(b) the targets to it achieved by when and by whom; and

(c) the indicators that will be used to compare the results produced by different organizations.

And, if the incentives are aligned with the purposes, targets, and indicators (always a big "if"), they can foster learning from the "positive deviants."

The Bar Chart and The List are effective because they provide everyone with three critical pieces of information:

(1) They tell everyone — every manager and every employee — how well his or her unit is doing.

(2) They tell everyone how everyone else's unit is doing.

(3) They tell everyone that everyone else knows how well his or her unit is doing.

Of course, to be motivational, this information must be public — or, at least, quasi-public. That is, the performance feedback must be distributed to everyone who has a role in achieving the overall target as well as achieving the targets for their individual units. It doesn't work if people are only told the height of their subunit's bar. It doesn't work if people are only told if their subunit is in Column A or in Column B. It only works to motivate improvement and foster learning if everyone has access to all three pieces of information. Both The Bar Chart and The List can motivate people because they provide a basis for social comparison.

Both The Bar Chart and The List reward — with recognition and prestige — the high-performing units and their managers. And they both reprove — with peer embarrassment — the low-performing units and their managers. Both of these simple feedback practices can influence the activities of unit managers (who value peer esteem) by motivating them to focus on the results that the organization seeks to produce.

Hypothesis K: Scrutinizing the positive deviants can facilitate everyone's learning.

> *By identifying those subunits that are achieving significant success and by analyzing and explaining the causes of their achievements, the leadership team can help other subunits learn how they too can improve performance.*

The Bar Chart or The List identifies for all to see not just those that are underperforming but also the positive deviants.[52] These are the subunits that have figured something out. Unless, they are purely lucky (which is apt to be exposed the next time the comparative data are collected and distributed), they

have learned something about how to achieve their targets — something that could benefit other subunits with similar targets.

Providing this information may not, however, be enough to improve everyone's performance. For unless these positive deviants are willing — indeed, eager — to share their knowledge, most of the other units may not be able to figure out how to achieve their targets. To improve performance, organizations need both the motivation to improve and the operational capacity for doing so. The positive deviants has, somehow, been able to create this necessary operational capacity.

Thus the nature of the reward system — formal and extrinsic; informal and intrinsic — is important. It influences directly, but perhaps subtly, the willingness of the positive deviants to provide the technical transfer than can help others to improve their operational capacity. If this is a zero-sum game, with the most positive of deviants getting all of the praise, prestige, and esteem (and almost all of any available financial rewards), they will have little incentive to help others. If, however, the rewards are unlimited — either because funding is not a constraint (a highly unlikely occurrence) or because the rewards are not financial (or, at least, not exclusively) — then they can be bestowed on every unit that meets its target. In addition, those positive deviants who help others will implicitly receive (and perhaps will be explicitly given) extra prestige.

Any effort to achieve new performance targets — either for new kinds of performance, or for new, higher levels of existing kinds of performance — requires some learning. After all something new needs to be done — something that the organization has not done before. Thus, the organization needs to create incentives both for subunits to experiment with new ways to achieve these new targets and for the most successful experimenters to share their lessons with others.

These positive deviants may be able to explain what they have done differently. They may be even able to explain why they choose this different approach. But this does not mean that they can explain why it worked. They may not appreciate the cause-and-effect relationships underlying their own, new apparently effective approach. As Carla O'Dell and Jackson Grayson have lamented: "If only we knew what we know."[53]

Thus, it could be that there is a puzzle to be solved. The underlying cause of the positive deviants' successes may not be obvious. Someone may have to analyze the data? Someone may have to investigate what the positive deviants are actually doing.

Hypothesis L: Recognizing accomplishments publicly can confirm that success is possible and valued.

> *By publicly and exuberantly recognizing meaningful accomplishments with intrinsic rewards for subunits, their employees, and their managers, the leadership team can demonstrate that significant successes can be achieved and are valued.*

As Matthew Gallagher, who served as director of both Baltimore's CitiStat and Maryland's StateStat has frequently emphasized, the meetings that he has chaired can "separate the contenders from the pretenders." At many PerformanceStat meetings, I have a difficult time following the conversation. Even if I am familiar with the organization, even if I have previously been to some of its PerformanceStat meetings, the acronyms and code words are still whizzing over my head. Nevertheless, even I can separate the contenders from the pretenders. The answers to the questions combined with the body language tell all. If a subunit has not hit its targets, even a visitor can tell whether this subunit and its manager are trying, whether they have thought analytically and creatively about their performance tasks, and whether they have created an approach that might work.

Questions asked in a quasi public setting are quite revealing. Everyone in the room gets it. Everyone can figure out who is performing and who isn't. Everyone can discern who is trying and who isn't. Everyone can detect who accepts the need to improve performance as legitimate and who believes that everything has been and continues to be going along just fine.

Moreover, the questions and answers themselves provide the first level of reward. For, like most of the significant rewards provided in the public sector, they convey status. The answers to questions reveal — to everyone — who is competent and who isn't. They reveal whose performance is truly outstanding, whose is acceptable, and whose is simply non-existent. Some praise does help, but it doesn't need to be all that effusive. Indeed, one way to provide extra, intrinsic rewards (in terms of internal prestige) to the top performers — the "positive deviants" — is to ask them to explain their strategies in formal training sessions with others. In any organizations, an honest, public complement about what everyone understands is a job well done conveys status. And, in any organization, achieved status provides all sorts of benefits — including social capital that creates deference and warrants flexibility.

Public executives are limited by the number and size of the extrinsic rewards that they can offer to top performers. Intrinsic rewards, however, are available with hardly any constraints. And the quasi-public setting of the PerformanceStat meeting provides an ideal occasion for identifying those who warrant such intrinsic rewards as well as an immediate opportunity to provide them.

Indeed, if they are creative, public executives will discover that they do possess a variety of intrinsic rewards. ways in which to convey status. They often possess a variety of inexpensive opportunities to recognize a job well done. "I was talking with some executives from the business community," Martin O'Malley, then Baltimore's mayor, said to me. And one of them asked: "Were those garbage men in your Orioles box the other night?" "Yes," answered O'Malley.

The mayor of Baltimore has a box for all Baltimore Orioles' baseball games and for all Baltimore Ravens' football games. The Orioles' box comes with food and drink (and not just lemonade). And at CitiStat meetings, O'Malley's leadership team was always looking for outstanding performers to whom they could give tickets to the mayor's box.

So on that night at Orioles Park, while a group of business executives sat in one box, the next box was occupied by an excellent team from the city's Department Public Works? Do you think they kept this game secret from their families and co-workers? Do you think they have forgotten the game?

Hypothesis M: Reproving the recalcitrant can get everyone's attention.

> *By regularly questioning subunit managers about their lack of analytical insight, operational effectiveness, or strategic coherence, the leadership team can ensure that everyone in the organization understands that ineffectiveness (as well as indifference, incompetence, and insubordination) are unacceptable.*

CompStat meetings at the Los Angeles Police Department often discuss the performance of three of the department's Areas (LAPD's equivalent of precincts). Discuss the first Area; take a break. Discuss the second Area; take another break. Discuss the third Area.

At one such meeting, the commander of the first Area, had not done very well. The meeting was being conducted by Charlie Beck, then the chief of detectives (who later succeeded Bratton as LAPD's chief). During the second break, Beck pointed to his Blackberry. He already had an e-mail from the first Area commander apologizing for poor performance.

LAPD's CompStat is quite different in style from NYPD's. During that morning's discussion of the first Area, neither Beck nor any of his leadership team had used any foul language. Neither the words chosen nor the tone with which they were spoken was disrespectful. But Beck and his colleagues did ask questions — lots of questions. These questions were consistently polite.

Yet everyone knew. The commander knew. Beck certainly knew. Even I knew. The Area commander — and the Area — had not performed well.

If an organization's leadership team is making a significant, deliberate effort to improve performance, it cannot tolerate subunit managers who resist either actively or passively. Indifferent,

incompetent, or insubordinate behavior by subunit managers automatically undermines the moral and the motivation of everyone else.

Fortunately, the PerformanceStat strategy has a built-in opportunity both for dramatizing that such behavior is unacceptable and for reproving the recalcitrant and punishing the saboteurs: asking follow-up questions, and more follow-up questions, and then even more follow-up questions. For the inability to provide adequate answers is embarrassing.

Who, however, caused the embarrassment? Was it Beck who asked the questions? Or was it the Area commander who was unable to provide satisfactory answers? Was the question embarrassing? Or was the answer? Or the lack of an answer?

In many organizations, many people believe it is unacceptable to deliberately embarrass a member of the team. They do not approve of a manager who tries to publicly humiliate a subordinate. Although "naming and shaming" is employed in a variety of settings to punish and deter deviant behavior, many people tend to find it slightly unsavory when employed within their own organization.[54] In such a small community, it strikes many as unfair.

Suppose, however, that the subunit managers who have been shamed — if only through their inability to answer a question about their responsibilities — had it coming? Suppose, these subunit managers were incompetent? Or suppose, while others diligent tried to improve performance, they ignored or even undermined these efforts? Do they deserve to be treated identically with those who are improve their sub units' performance?

Indeed, is it fair to treat someone who is performing poorly with the same respect that one treats someone whose performance is exemplary? Is it fair to value someone who is pursuing the organization's mission and targets with commitment equally with someone who ignores them? Doesn't an organization's own integrity require its leadership to ask *all* of the organization's managers about what actions they have taken to complete their assignments, to eliminate their performance deficits, to achieve their targets?

The questioning at a PerformanceStat session need not be relentless. For, as I have frequently emphasized, Jack Maple's focus on "relentless followup" could be interpreted to suggest "hatred, hostility, or vengeance."[55] In reality, however, effective follow-up merely needs to be "persistent." There is nothing inherently unethical about asking a follow-up question — "Has this task been completed?" — and continuing to ask this question until the answer is an undeniable "yes."

Will a manager who answers this question with a "no" — or an equivocation — be embarrassed? I would think so. But, again, who caused the embarrassment? The questioner? Or the answerer?

Specifying the task and its deadline is an ethical undertaking. Is the task *fair?* That is: Is the task one that the manager can reasonably be expected to complete by the target date? Is the assignment of tasks done with *integrity?* That is: Are different managers given tasks of equal difficulty and significance?

Conducting the follow-up is also an ethical undertaking. Is the follow-up questioning done with *respect?* That is: Are the words and the tone of the conversation courteous? Is the inquiry about a missed deadline handled with *compassion?* That is: Are the explanations of extenuating circumstances treated with appropriate consideration for the multiple challenges, pressures, and deadlines facing the manager?

Asking follow-up questions about important tasks designed to fix consequential performance deficits is an essential task of leadership — an ethical task. Indeed, a failure to follow-up would be — given the importance of the underlying purpose — an ethical lapse.

Hypothesis N: Telling stories can foster a results-focused culture.

> *By frequently telling stories about employees whose exemplary work significantly improved performance (and occasionally about the incompetent or indolent who undermine these efforts), the leadership team can validate the meaning underlying everyone's work and foster a results-focused culture.*

Thirty years ago, the idea of organizational culture first invaded the private sector and then government. "This culture stuff is great," said one business executive to another as they left a compelling seminar. "Get me a culture by Monday."[56]

Unfortunately, this executive already had a culture. And like all organizational cultures, it was undoubtedly subtle and deep. This executive wasn't going have a new one by Monday — perhaps not for years. Indeed, speaking of U.S. government agencies, John DiIulio, then a member of the White House staff, observed: "Organizational cultures are among the hardest substances known to humankind." To this general rule, DiIulio offered only one exception: "diamonds."[57]

Also unfortunately, most public agencies have not a results-focused culture, but a rule-obsessed one. After all, if civil servants follow the rules, it is hard to get in trouble. At the same time, if civil servants follow all of the rules, it is hard to improve performance. The rules limit the discretion and flexibility that agencies need to improve their performance.

Edgar Schein defines organizational "culture" as "a pattern of shared basic assumptions that the group learned as it solved its problems of external adaptation and internal integration, that has worked well enough to be considered valid and, therefore, to be taught to new members as the correct way to perceive, think, and feel in relations to those problems."[58] Because, within most governments, there is little opportunity to earn meaningful rewards for positive contributions but multiple opportunities to incur serious penalties for some major or minor error, many civil servants believe that an important, personal problem of their own internal integration is how to stay out of trouble. To solve this problem, the dominant culture teaches them, learn to assiduously follow the bureaucratic rules.

Over half a century ago, the social psychologist Kurt Lewin observed that changing the performance of organizations was difficult:

> A change toward a higher level of group performance is frequently short lived; after a 'shot in the arm,' group life soon returns to its previous level. This indicates that it does not suffice to define the objective of a planned change in group performance as the reaching of a different level. Permanency of the new level, or permanency for a desired period, should be included in the objective.

Thus, Lewin advocated a three-step process for improving performance: (1) "unfreeze" old patterns of behavior; (2) "move" the organization to new behaviors that produce improved performance, and (3) "freeze" the organization at this new level.[59]

Thus, to create a new, results-focused culture, the leadership may need to first destroy the existing, rule-driven mindset before they can move the organization to a new way of thinking about its responsibilities. Then, after creating the new results-focused organization, they will need to figure how to re-freeze the organization so that everyone is committed to the new way of conducting the organization's business.

To Caren Siehl and Joanne Martin, culture is "the glue that holds an organization together" — and includes its "core values" that define its "basic philosophy or mission" as well as the means for transmitting these values, such as stories and rituals. Moreover, add Siehl and Martin, it includes strategies that managers can employ to reinforce underlying values of their preferred culture."[60] From this perspective, a strategy for unfreezing a rule-based culture would be destroy (somehow) the old core values that focuses people and organizations on staying out of trouble by, perhaps, undermining the validity of the old stories (which told how people got into trouble only when they didn't follow the rules and how they

stayed out of trouble when they did) and the rituals that reenforced them. Then, the leaders need new stories and new rituals that can refreeze their new results-focused culture.

In many ways, stories define an organization's culture. They provide the metaphors that explain how people should and should not behave. They illustrate what does and doesn't work. Stories convey the behaviors that are desirable and acceptable, objectionable and intolerable. They are an effective way to indoctrinate new members. Stories that are "told and retold" writes Anne Khademian, newcomers understand what is important . . . and why." They help people "learn the language, rules of thumb, and understandings of the organization."[61]

Everyone tells stories to educate the new comers. Front-line workers use them to prevent their new colleagues from becoming disruptive rate-breakers. But the organization's leadership team can also tell stories to communicate what they are trying to accomplish — or, as Philip Selznick put it — to convey "the institutional embodiment of purpose." Such stories — or what Selznick terms "socially integrating myths" — "are efforts to state, in the language of uplift and idealism, what is distinctive about the aims and methods of the enterprise."[62]

Whether front-line employees are telling stories to their colleagues or organizational leaders are telling them to their employees, they are employing these stories to establish organizational norms. In every organization — including every government jurisdiction, public agency, and all of their subunits — front-line workers and middle managers recount stories whose message contradicts or undermines those they are receiving from top leadership. Thus, these leaders need to be particularly aggressive in establishing the validity and significance of their stories.

Fortunately, PerformanceStat provides a built-in ritual for conveying those stories: the regular PerformanceStat sessions. Although to most visitors, the meetings may appear to be about the data, they are also about the stories. Many of these stories will appear to be about failures of units or individuals to complete an assigned task, or to miss an opportunity to help eliminate a performance deficit. Effective leaders, however, are always looking for stories to tell about middle-managers and front-line workers who did something extraordinary (or even quite ordinary) that made an extraordinary (or ordinary) contribution to the results that the organization is trying to produce.

In the competition over the validity of contradictory stories, the leadership team is at a disadvantage. Others are telling their stories to colleagues every day. The leadership team, however, are able to tell their stories only at the biweekly, or monthly, or even quarterly PerformanceStat sessions. Thus, to keep everyone focused on the big picture, they need to keep telling their stories. They need to identify an ongoing series of individuals and their actions to serve as real examples of how everyone in the organization can contribute to better performance. In addition to these heroes, they need to be sure that the villains are not ignored, that stories about both their sloth and its consequences become another defining myth.

Hypothesis O: Abetting everyone's implicit evaluation of everyone else can breed individual and team motivation.

> *By creating an environment in which everyone present at the PerformanceStat meetings will implicitly (and perhaps explicitly) evaluate everyone else who is present, the leadership team can motivate agencies, individuals, and teams to achieve their targets and engage in results-focused behavior.*

PerformanceStat sessions are not just venues for telling stories. They become stories. Participants easily remember what happened at the last meeting. They remember what happened a month ago or a year ago. And if what happened was particularly remarkable — particularly accomplished or particularly odious — it will, itself, become a story.

Everyone in the room (except for those who are checking their BlackBerries) is observing the discussion: the questions asked; the answers given. At the same time, they are all playing in their heads

the previous set of questions and answers for which the current discussion is a follow-up. They know why today's questions are being asked. And they are attentive to the answers being given. Simultaneously, they are evaluating these answers — and the person offering them.

Indeed, at every PerformanceStat session, every individual's performance is not just being evaluated by the organization's leadership team and its PerformanceStat staff. As they question and probe — praise and fault — everyone else is silently doing the same.

Every individual who is called upon to speak is constantly being evaluated by every other individual in the room. Most of these evaluations are not formal; they are never spoken to the manager. Most are merely implicit. Still, we humans cannot help ourselves; we are always evaluating everyone else. And it is this evaluation — both explicit (as expressed in a follow-up question or comment) and implicit (by everyone in the room who isn't asleep) — that creates the stress for which CompStat, CitiStat, and multiple PerformanceStats are famous.

H. L. Mencken once observed, "conscience is the inner voice which warns us somebody may be looking."[63] At a PerformanceStat meeting, however, no one needs this inner-voice warning. Everyone knows that everyone else is watching — and evaluating.

This is relevant given the research suggesting that evaluation can significantly enhance performance.[64] Indeed, this increase in performance is not dependent on the knowledge that a formal evaluation will be conducted. Rather, the impact comes from the visibility to others of the individual's performance, which introduces the potential for such an evaluation.[65] Clearly, having to stand up and answer questions about what you and your organization have done (and not done) in front of others gives everyone in the room the opportunity to evaluate your performance.

The resulting stress — the desire to be perceived as competent by superiors, peers, and subordinates — can motivate improved performance. But too much stress can also suppress performance. This relationship, first introduced a century ago by Robert Yerkes and John Dodson,[66] is usually described as an "inverted U."[67] As the stress first increases, so does performance; but as the stress continues to increase the performance curve peaks and then declines. Unfortunately, there is no reason to believe that there exists a unique, discoverable peak that is the same for all people, for all tasks, and in all circumstances. Consequently, for any manager, it is not obvious when the evaluations create too much pressure and thus begin to reduce performance. Still, in any organization in which there has been little or no pressure to improve performance, an initial increase in stress is apt to move most people *up* the Yerkes-Dodson curve — to help replace a rule-obsessed culture with a results-focused one.

The evaluations that occur in a PerformanceStat room not only create stress. They also inflict punishment. This punishment may be a sharp rebuke, a mild reprimand, a pointed question, a subtle inquiry, or a sarcastic comment. It may be directly expressed in words or conveyed in tone and body language. Or, as at the Los Angeles Police Department, it may simply come in yet another follow-up question. Rarely will the punishment be a formal critique placed in the employee's file. Still, to move from a rule-obsessed to a results-focused culture, some punishments may essential. Charles O'Reilly and Barton Weitz conclude that the "appropriate use of sanctions may be perceived by [other] employees as legitimate and may be conducive to the development of productive group norms."[68] If indolent behavior goes unpunished — thus permitting everyone to conclude that it is acceptable — it can, itself, become the norm.

Of course, each individual manager is not just being evaluated at the regular PerformanceStat sessions. Every manager is being evaluated every day by co-workers. These co-workers may, however, being using criteria that are not identical — that might actually be antithetical — to those being employed by the organization's leadership team. Consequently, a PerformanceStat leadership strategy will have little impact on front-line workers who never attend the meetings *unless* their agency head also uses the same criteria at internal "AgencyStat" meetings. Unless the leadership's results-oriented culture — completing operational assignments and achieving performance targets — penetrate down to each working team, a PerformanceStat strategy will not create any evaluation or stress among the individuals who do the real work and thus may have little impact on the day-to-day operations.

Hypothesis P: Remaining persistent can prove that this isn't going away.

> *By making PerformanceStat a formal part of organizational life and assiduously adhering to the practices that they have established, the leadership team can demonstrate that their performance strategy is neither superfluous nor ephemeral.*

A new chief executive. A new management scheme. What's new?

The permanent civil service has seen it all before. Indeed, they're anticipating it. New chief executive; new stuff — new people, new rhetoric, new systems, new best practices from the airport bookstore.

And they've dealt with all a series of these newbies. Some, they've educated. Others, they've fought. Still others, they've simply ignored.

In some jurisdictions, the permanent civil servants refer to the political appointees as the "tourists" and to themselves as the "residents." Elsewhere, they talk about themselves as the "We Bes": "We be here before you're here; we be here after you're here."

For they all know that "this, too, shall pass." Indeed, as Abraham Lincoln observed of this maxim about transience: "How chastening in the hour of pride!"[69] For unless the new executive is persistent with his or her new PerformanceStat leadership strategy, the hour of its pride will be short indeed.

This has been true for too many nascent efforts to produce better results with an approach mimicking some of the characteristics of PerformanceStat: There is the initial burst of enthusiasm. Technology is purchased to collect and project data. Analysts are hired to scrutinize these data. Meetings are held. The room is full.

But what are the targets? The assignments? Where are the reports? The questions? The follow-up? Where is the recognition for improved performance? The thanks? Where is the reproof for the recalcitrant? For the inability to produce better results? To even try to produce better results? What is a better result? Indeed, what is the purpose?

Stephanie Hirsch, the first director of SomerStat, observed: "The mayor saying it should happen isn't enough." So what is enough? Nothing really. Producing results is hard, and improving performance is requires continuous effort. Nothing is ever enough. Unless the leadership team is constantly engaged, the focus on results will quickly atrophy.

If the chief executive and his or her leadership team are not persistent — if they don't demonstrate that their PerformanceStat is not going away — none of the other fifteen causal behaviors will have any impact on the organization's performance.

Notes

1. Jon Elster, *Explaining Technical Change: A Case Study in the Philosophy of Science* (Cambridge, U.K.: Cambridge University Press, 1983), p. 24. Elster goes on to note: "Here the term 'mechanism should be understood broadly, to cover intentional chains from a goal to an action as well as causal chains from an event to its effect.

Later, Elster modified his concept of the mechanism, explaining: "In that work [his 1983 book], I advocated the search for mechanisms as more or less synonymous with the reductionist strategy in science." Jon Elster, "A plea for mechanisms," in Peter Hedström and Richard Swedberg, *Social Mechanisms: An Analytical Approach to Social Theory* (Cambridge, U.K.: Cambridge University Press, 1998), p. 47.

2. This is, of course, an ideal-type definition. It will not perfectly describe an one of the particular, operational versions of the PerformanceStat leadership strategy.

3. Baltimore's CitiStat was the first government-wide effort to apply NYPD's CompStat strategy to multiple agencies within an entire jurisdiction (not just a single agency). It is since been adapted by other jurisdictions: cities in the U.S. and elsewhere, plus two U.S. states: Washington and Maryland.

4. Mario Bunge, "How Does It Work? The Search for Explanatory Mechanisms," *Philosophy of the Social Sciences,* vol. 34, no. 2 (2004), p. 182.

5. For many social policies, the mechanism is what Schelling calls the "rational, or at least purposive, individual" whose behavior (often economic) interacting with the behavior of other such motivated individuals creates the consequences. Thomas C. Schelling, "Social mechanisms and social dynamics," in in Peter Hedström and Richard Swedberg, *Social Mechanisms: An Analytical Approach to Social Theory* (Cambridge, U.K.: Cambridge University Press, 1998), p. 32. For examples of such mechanisms see: Thomas C. Schelling, *Micromotives and Macrobehavior* (New York: W. W. Norton & Company, 1978).

6. For analyses of the Salk's polio experiment, see:

Thomas Francis, Jr., Robert F. Korns, Robert B. Voight, Morton Boisen, Fay Hemphill, John A. Napier, and Eva Tolchinsky, "An Evaluation of the 1954 Poliomyelitis Vaccine Trials," *American Journal of Public Health,* vol. 45, no. 5, (May 1955, pt. 2).

Paul Meier. "The Biggest Public Health Experiment Ever: The 1954 Field Trial of the Salk Poliomyelitis Vaccine" in *Statistics, A Guide to the Unknown,* Judith M. Tanur, Frederick Mosteller, William H. Kruskal, Richard F. Link, Richard S. Pieters, Gerald R. Rising (eds.) (San Francisco: Holden-Day, 1975), pp. 2–13.

"The Salk Poliomyelitis Vaccine," *American Journal of Public Health,* vol. 45, no. 5 (May 1955), p. 676.

Marcia Meldrum, "'A calculated risk': the Salk polio vaccine field trials of 1954," *MJB,* vol. 317, no. 7167 (October 31, 1998), pp. 1233-1236.

7. Peter J. Mohr, Barry N. Taylor, David B. Newell, "CODATA Recommended Values of the Fundamental Physical Constants: 2006," *Reviews of Modern Physics,* vol. 80, no. 2 (April-June 2008), pp. 687, 686. CODATA is the Committee on Data for Science and Technology of the International Council on Science

8. Yes, Newton's laws are approximations that happen to work very well, except approaching the speed of light, when scientists and engineers must use Einstein's equations. But many scientists and certainly most engineers don't have to worry about this refinement.

9. Nevertheless, even in social policy, people hope for a "vaccine." The leader of one effort to prevent bullying in schools said: "I don't feel we are in the vaccine stage, where we just know that something is going to work and you just have to line up and take a shot. . . . We are nowhere near there." In fact, we will never get there. Quoted in Sarah Schweitzer, "No easy fix found for bullying," *The Boston Globe,* December 30, 2010, p. A13.

10. Red Auerbach, who created the Boston Celtics' basketball dynasty, was a master at employing differential behaviors to motivate people. Here are two versions of Auerbach's differential motivational strategy:

[Tommy] Heinsohn also came to know that Auerbach never motivated anybody the same way, that he intuitively

> understood which individual buttons to push. He knew never to criticize [Bob] Cousy or [Bill] Russell. He knew never to yell at [Bill] Sharman, who didn't take it well and might just retaliate by punching Red in the mouth. He rarely yelled at [Frank] Ramsey, who might take criticism to hear. Heinsohn was perfect. Auerbach could yell at him, and in doing so could get his message to the entire team. So it was Heinsohn who was too heavy, Heinsohn who wasn't playing defense intensely enough in practice, Heinsohn who wasn't running wind sprints hard enough. Heinsohn who was smoking too many cigarettes, Heinsohn who was always doing something wrong.

Bill Reynolds, *Cousy: His Life, Career, and the Birth of Big-Time Basketball* (New York: Simon & Schuster, 2005), p. 144.

> If Auerbach wanted to chide Russell in the locker room, he did it by osmosis, by yelling at Heinsohn. "He'd say, 'Tommy you gotta do this, Tommy you gotta do that — and that goes for you, too, Russell,'" Heinsohn recalled.

John Powers, "Auerbach's legacy remains unflagging," *The Boston Globe,* December 23, 1999

11. Peter Hedström and Petri Ylikoski, "Causal Mechanisms in the Social Sciences," *Annual Review of Sociology,* vol. 36 (2010), p. 64.

12. Peter Hedström and Richard Swedberg, "Social mechanisms: An introductory essay," p. 13 in Peter Hedström and Richard Swedberg, *Social Mechanisms: An Analytical Approach to Social Theory* (Cambridge, U.K.: Cambridge University Press, 1998).

13. Peter Hedström and Petri Ylikoski, "Causal Mechanisms in the Social Sciences," *Annual Review of Sociology,* vol. 36 (2010), pp. 50, 52.

14. Peter Hedström and Petri Ylikoski, "Causal Mechanisms in the Social Sciences," *Annual Review of Sociology,* vol. 36 (2010), p. 54.

15. Bunge writes that "because most mechanisms are nonmechnical" — including both scientific and social mechanisms, "I prefer to call it *mechanismic"* [emphasis in the original]. Mario Bunge, "How Does It Work? The Search for Explanatory Mechanisms," *Philosophy of the Social Sciences,* vol. 34, no. 2 (2004), p. 203.

Still, the use of "mechanisms" — complete with its cogs and wheels — to explain the observations of social science is well established. In a review of "Causal Mechanisms in the Social Sciences," Peter Hedström and Petri Ylikoski conclude:

> The basic idea of a mechanism-based explanation is quite simple: At its core, it implies that proper explanations should detail the cogs and wheels of the causal process through which the outcome to be explained was brought about.

Annual Review of Sociology, vol. 36 (2010), p. 50.

16. Max Weber, *Economy and Society: An Outline of Interpretive Sociology* (Berkeley, Calif.: University of California Press, 1978), p. LIX.

17. For a discussion of complex adaptive systems, including potential applications to public management, see:

> M. Mitchell Waldrop, *Complexity: The Emerging Science at the Edge of Order and Chaos* (New York: Simon & Schuster, 1992).
>
> Robert D. Behn, "Can Public Managers Usefully Exploit the Principles of Complex Adaptive Systems? No; Yes; Maybe; But Is It Worth It?" A paper presented at the Twenty-First Annual Research Conference of the Association for Public Policy Analysis and Management, Washington, D.C., November 5, 1999, and at the Fifth National Public Management Conference, College Station, Texas, December 3, 1999.
>
> Geert R. Teisman and Erik-Hans Klijn, "Complexity Theory and Public Management: An introduction," *Public Management Review,* vol. 10, no. 3 (2008), pp. 287-297. (This article is an introduction to a special issue of Public Management Review on the relevance of complexity theory to public management, including seven other articles.

18. Thomas C. Schelling, "Social mechanisms and social dynamics," in Peter Hedström and Richard Swedberg, *Social Mechanisms: An Analytical Approach to Social Theory* (Cambridge, U.K.: Cambridge University Press, 1998), pp. 32-33.

19. "Explanations of most concrete social events or states require resort to several elementary mechanisms: One is not enough." Peter Hedström and Richard Swedberg, "Social mechanisms: An introductory essay," p. 21 in Peter Hedström and Richard Swedberg, *Social Mechanisms: An Analytical Approach to Social Theory* (Cambridge, U.K.: Cambridge University Press, 1998).

20. Peter Hedström and Petri Ylikoski, "Causal Mechanisms in the Social Sciences," *Annual Review of Sociology,* vol. 36 (2010), p. 61.

21. Peter Hedström and Richard Swedberg, "Social mechanisms: An introductory essay," p. 15 in Peter Hedström and Richard Swedberg, *Social Mechanisms: An Analytical Approach to Social Theory* (Cambridge, U.K.: Cambridge University Press, 1998).

22. Mario Bunge, "How Does It Work? The Search for Explanatory Mechanisms," *Philosophy of the Social Sciences,* vol. 34, no. 2 (2004), pp. 207, 200, 186.

23. I am sure that others can identify additional causal behaviors. Still others will argue that some of these sixteen behaviors could and should be collapsed into fewer, but more potent ones. I simply argue that each of these sixteen can, in many circumstances, be directly causal and this list is a helpful way to conceptualize the cause-and-effect relationships of a PerformanceStat leadership strategy.

24. A variety of observers had noted the absence of any underlying theory to help support particular management strategies:

Drucker writes that one "explanation for the tendency of so much public administration today to commit itself to policies that can only result in non-performance is the lack of concern with performance in public administration theory." Peter F. Drucker, "The Deadly Sins of Public Administration, Public Administrative Review, vol. 40, no. 2 (March/April 1980), p. 106.

Talbot writes: "What is striking about most of the policy and practitioner, and even academic, literature is the absence of theoretical justification for particular models of performance proposed." Colin Talbot, "Performance Management," in *The Oxford Handbook of Public Management,* Ewan Ferlie, Laurence E. Lynn, Jr., and Christopher Pollitt, eds. (Oxford, U.K.: Oxford University Press, 2005), p. 508.

I am not, however, arguing that anyone in the New York Police Department (or any other police department) or in the Baltimore mayor's office (or any other CitiStat office), or in any other PerformanceStat office specified any of these sixteen causal behaviors. Instead, these sixteen leadership behaviors are my effort to suggest (given what we know about human behavior in organizations) why particular aspects of a well conceived PerformanceStat leadership strategy might help to foster (individually and collectively) improved organizational performance.

25. No motivational system can be all rewards. But in the specific context of a governmental organization seeking to improve performance, what is the opposite of "recognizing accomplishments"? It can't simply be "punishing failure." The punishment of failure may not help to produce any better results. Firing the incompetent who produced the failure will not, itself, help. Will the replacement be any better?

For this behavior, I have carefully chosen two words: The verb to "reprove" and the noun "recalcitrant."

For "reprove," Webster's New Dictionary of Synonyms, lists "rebuke," "reprimand," "admonish," "reproach," and "chide." Other possible verbs include: "blame," "castigate," "censure," "chasten," "chastise," "condemn," "correct," "denounce," "discipline," "punish," "reprehend," . . . . The list is quite long. And, indeed, some PerformanceStats (NYPD's CompStat, in particular) have created a reputation that they do "castigate" or "condemn" or . . . .

"Reprove," however, has what I think is the right connotation: "To reprove is to blame or censure, often kindly or without harshness and usually in the hope of correcting the fault." Ah: "correcting the fault." In PerformanceStat, that is certainly "the hope." Philip B. Gove, ed., *Webster's New Dictionary of Synonyms* (Springfield, Mass.: G. & C. Merriam Company, 1968), p. 685.

Now, what kind of people deserve being "reproved"? Here I have chosen to focus on the "recalcitrant" — those who are actively rebellious or passively resistant. For at least when a leadership team is launching their PerformanceStat, these are apt to create the biggest personnel problems.

Later, the leadership team will have to deal with the merely "ineffective," for which synonyms include "ineffectual," "inefficient," and "inefficacious" [Gove, p. 436] — all of which "mean not producing or incapable of producing results," which is, of course, the concern of any PerformanceStat.

But what about the "irresponsible" or the "incompetent"? They should not be merely reproved. They ought to be assigned to a less demanding job that is suitable for their competence, or fired — and replaced by someone who has demonstrated skills and success as a manager and leader.

26. Luther Gulick, "Notes on the Theory of Organization," in Luther Gulick & L. Urwick (eds.), *Papers on the Science of Administration* (New York: Institute of Public Administration, 1936), p. 13.

27. I could have collapsed the first four behaviors and the latter twelve into just two behaviors: Behavior I & Behavior M. Behavior I would focus on the approaches employed to identify and convey the Improvements that had to be made and would contain three sub-theories: I1, I2, I3, and I4. Behavior M would focus on the Motivational strategies employed to achieve the improvements and would contain twelve sub-theories: M1 . . . M12. I wanted, however, to emphasize the motivational affects of these twelve different (though obviously reenforcing) behaviors.

28. This is consistent with the counterfactual standard of the causal-mechanism approach. Hedström and Ylikoski write:

> A mechanism-based explanation describes the causal process selectively. It does not aim at an exhaustive account of all details but seeks to capture the crucial elements of the process by abstracting away the irrelevant details. The relevance of entities, their properties, and their interactions is determined by their ability to make a relevant difference to the outcome of interest. If the presence of an entity or of changes in its properties or activities truly does not make any difference to the effect to be explained, it can be ignored. This counterfactual criterion of relevance implies that mechanism-based explanations involve counterfactual reasoning about possible changes and their consequences.

Peter Hedström and Petri Ylikoski, "Causal Mechanisms in the Social Sciences," *Annual Review of Sociology,* vol. 36 (2010), p. 53.

29. For the definitive explanation of why in organizations "it all depends," see, Harvey Sherman, *It All Depends* (University, Ala.: The University of Alabama Press, 1966).

30. The full quote from Friedrich Nietzsche is: "Forgetting our objectives is the most frequent act of stupidity." Friedrich Nietzsche, *The Wanderer and His Shadow* (1880). In Friedrich Nietzsche, *Human, All Too Human: A Book for Free Spirits,* translated by R.J. Hollingdale (Cambridge, U.K.: Cambridge University Press, 1996), p. 360.

31. J. Richard Hackman, "The Psychology of Self-Management in Organizations," in *Psychology and Work: Productivity, Change, and Employment,* Michael S. Pallack and Robert O. Perloff (Washington, D.C.: American Psychological Association, 1986), p. 102.

32. Quoted in, Walter Isaacson, *Benjamin Franklin: an American Life* (New York: Simon and Schuster, 2003), p. 99.

33. The research on the motivational impact of goals is long and significant. Here are just a few publications:

> Gary P. Latham and Edwin A. Locke, "Self Regulation through Goal Setting," *Organizational Behavior and Human Decision Processes,* vol. 50, no. 2 (1991), pp. 212-247.
>
> Gary P. Latham and Craig C. Pinder, "Work Motivation Theory and Research at the Dawn of the Twenty-First Century," *Annual Review of Psychology,* vol. 56 (2005), pp. 485-516.
>
> Edwin A. Locke and Gary P. Latham, *Goal Setting: A Motivational Technique That Works* (Englewood Cliffs, N.J.: Prentice Hall, 1984).

Edwin A. Locke and Gary P. Latham, *A Theory of Goal Setting and Task Performance* (Englewood Cliffs, NJ: Prentice-Hall, 1990).

Edwin A. Locke and Gary P. Latham, "Building a Practically Useful Theory of Goal Setting and Task Motivation: A 35-Year Odyssey," *American Psychologist*, vol. 57, no. 9 (2002), pp. 705-717.

Edwin A. Locke and Gary P. Latham, "What should we do about Motivation Theory," *The Academy of Management Review,* vol. 29, no. 3 (2004), pp. 388-403.

Terence R. Mitchell and Denise Daniels, "Motivation," in *Handbook of Psychology,* Vol. 12, *Industrial and Organizational Psychology,* Walter C. Borman, Daniel R. Ilgen, and Richard J. Klimoski, eds. (New York: John Wiley and Sons., 2003), pp. 225-254.

Robert Rodgers and John E. Hunter, "Impact of Management by Objectives on Organizational Productivity," *Journal of Applied Psychology,* vol. 76, no. 2 (1991), pp. 322-336.

34. Edwin A. Locke, Karyll N. Shaw, Lise M. Saari, and Gary P. Latham, "Goal Setting and Task Performance," *Psychological Bulletin* vol. 90, no. 1 (1981), pp. 125-152.

35. W. Jack Duncan, *Great Ideas in Management: Lessons from the Founders and Foundations of Managerial Practice* (San Francisco: Jossey-Bass, 1989), p. 172.

36. Christopher Pollitt, "Performance Management in Practice: A Comparative Study of Executive Agencies," *Journal of Public Administration Research and Theory,* vol. 16, no. 1 (2006), p. 35.

37. Stephen G. Harkins and Richard E. Petty, "Effects of Task Difficulty and Task Uniqueness on Social Loafing," *Journal of Personality and Social Psychology* vol. 43, no. 6(1982), pp. 1214-1229.

38. Most of the research in social and personal psychology is based on laboratory experiments with undergraduates. For such experiments, the consequence to the student subject is almost always trivial, of much less consequence for the subjects (or for anyone else), than any course quiz.

Sometimes, however, these undergraduate subjects are tricked; the formal task that they have been given is not the behavior that the laboratory experiment is attempting to examine. See, for example:

John M. Darley and Bibb Latané, "Bystander Intervention in Emergencies: Diffusion of Responsibility," *Journal of Personality and Social Psychology,* vol. 8, no. 4 (1968), pp. 377-383.

Bibb Latané and John M. Darley, "Group Inhibition of Bystander Intervention in Emergencies," *Journal of Personality and Social Psychology,* vol. 10, no. 3 (1968), pp. 215-221.

A significant exception to the tradition of the undergraduate psychology lab, and thus to the external validity of the conclusions, is research of Adam Grant. See, for example:

Adam M. Grant, "Relational Job Design and the Motivation to Make a Prosocial Difference," *Academy of Management Review,* vol. 32, no. 2 (2007), pp. 393-417.

Adam M. Grant, Jane E. Dutton and Brent D. Rosso, "Giving Commitment: Employee Support Programs and the Prosocial Sensemaking Process," *Academy of Management Journal,* vol. 51, no. 5 (October 2008), pp. 898-918.

Adam M. Grant and Kimberly A. Wade-Benzoni, "The Hot and Cool of Death Awareness at Work: Mortality Cues, Aging, and Self-Protective and Prosocial Motivations," *Academy of Management Review,* vol. 34, no. 4 (2009), pp. 600-622.

39. Christopher Pollitt, "Performance Management in Practice: A Comparative Study of Executive Agencies," *Journal of Public Administration Research and Theory,* vol. 16, no. 1 (2006), pp. 25-44.

40. Richard P. Nathan, *The Administrative Presidency* (New York: John Wiley & Sons, 1983), p. 30.

41. Alfred Tat-Kei Ho, "Accounting for the Value of Performance Measurement from the Perspective of Midwestern Mayors," *Journal of Public Administration Research and Theory,* vo. 16, no. 2 (2006), p. 219.

42. Ho assessed the interest of midwestern mayors in performance measurement and observes: "mayors of larger cities were less likely to take personal initiatives to launch performance management reforms." Yet in Baltimore, larger than any city Ho surveyed, three mayors did make performance measurement — and *performance leadership* — a personal initiative. Alfred Tat-Kei Ho, "Accounting for the Value of Performance Measurement from the Perspective of Midwestern Mayors," *Journal of Public Administration Research and Theory,* vo. 16, no. 2 (2006), p. 224.

43. Anthony Downs, *Inside Bureaucracy* (Washington, D.C.: Brookings, 1967), p. 145.

44. Robert D. Behn, "On why public executives need to avoid Dashboards as Data Dumps," *Bob Behn's Public Management Report,* vol. 5, no. 11 (July 2008).

45. Christopher Pollitt, "Performance Management in Practice: A Comparative Study of Executive Agencies," *Journal of Public Administration Research and Theory,* vol. 16, no. 1 (2006), p. 34.

46. Kipling Williams, Stephen Harkins, and Bibb Latané. 1981. "Identifiability as a Deterrent to Social Loafing: Two Cheering Experiments," *Journal of Personality and Social Psychology,* vol. 40, no. 2 (1981), pp. 303 & 311. See also: Kipling D. Williams, Steve A. Nida, Lawrence D. Baca, and Bibb Latané, "Social Loafing and Swimming: Effects of Identifiability on Individual and Relay Performance of Intercollegiate Swimmers," *Basic and Applied Social Psychology,* vol. 10, no. 1 (1989), pp. 73-81.

47. Downs observes that the creation of "a separate monitoring agency" (which a PerformanceStat office certainly is) "tends to increase the rigidity of the operating bureau" because (a) the monitoring agency "imposes ever more complex and ever more restrictive regulations," (b) the operating agency must devote "more and more of its resources to satisfy the increasing demands of the monitors for information and written reports," and (c) the operating agency "tends to devote ever more resources to figuring out ways of evading or counteracting the monitors' additional regulations." Anthony Downs, *Inside Bureaucracy* (Washington, D.C.: Brookings, 1967), pp. 158-159.

For PerformanceStat, the first of these three behaviors doesn't (quite) apply. Instead of regulations, PerformanceStat imposes targets and assignments, which can certainly become "ever more complex and every more restrictive" or demanding. Downs's second argument does indeed hold; the leaders of each subunit must show up for their regular meetings and devote time and other resources in preparing for them. Over time, however, the chief executive's leadership team can reduce significantly the effort that the managers of subunits devote to evasive tactics, if only by replacing those agency heads who engaged in them. But they may not have the same success in reducing the evasive behavior of front-line supervisors and workers.

48. Anthony Downs, *Inside Bureaucracy* (Washington, D.C.: Brookings, 1967), pp. 163-164.

Downs suggests two ways that organizations can break the ossification: by creating a new organization to deal with a complex and novel "urgent task"; or by reorganizing (pp. 160-161 & 165-166). But the tasks conducted by a centuries-old city such a Baltimore are hardly novel. And Downs's second way for a city to shake-up its ossified agencies — through reorganization — has its own, well-known disadvantages.

49. Jeffrey Pfeffer, "Organizations and Organization Theory," in *The Handbook of Social Psychology,* Third Edition, Vol. I (Theory and Method), Gardner Lindzey and Elliot Aronson eds. (New York: Random House, 1985), p. 390.

See also: Clive Seligman John M. Darley, "Feedback as a Means of Decreasing Residential Energy Consumption," *Journal of Applied Psychology,* vol. 62, no. 4 (1997), pp. 363-368; Edward L. Thorndike, Edward L. 1927. The law of effect. *The American Journal of Psychology* vol. 39, no. 1/4 (December 1927), pp. 212-222.

50. Obviously, this behavior doesn't work in agencies that are not large enough to have multiple subunits whose performance data can be compared.

51. Robert D. Behn, *Leadership Counts: Lessons for Public Manag*ers (Cambridge, Mass.: Harvard University Press, 1991), p. 71. Robert D. Behn, "On the Motivational Impact of The List," *Bob Behn's Public Management Report,* vol. 1, no. 2 (October 2003).

52. Gretchen M. Spreitzer and Scott Sonenshein, "Toward the Construct Definition of Positive Deviance." *American Behavioral Scientist,* vol. 47, no. 6 (February 2004), pp. 828-847.

53. Carla O'Dell and C. Jackson Grayson, "If Only We Knew What We Know: Identification and Transfer of Internal Best Practices," *California Management Review,* vol. 40, no. 3 (Spring 1998), pp. 154-174.

54. For a discussion of when naming and shaming might work, see: Ray Pawson, "Evidence and Policy and Naming and Shaming," *Policy Studies,* vol. 23, no. 3/4 (2002), pp. 211-230.

55. One dictionary reports that relentless implies "an absence of pity or of any feeling that would cause one to relent and to restrain through compassion the fury or violence of one's rage, hatred, hostility, or vengeance." Philip B. Gove (ed.), *Webster's New Dictionary of Synonyms* (Springfield, Mass.: G. & C. Merriam Company, 1968), p. 383.

56. James Lincoln and Didier Guillot, "A Durkheimian View of Organizational Culture," in Marek Korczynski, Randy Hodson, Paul K. Edwards, eds., *Social Theory at Work* (Oxford, U.K.: University of Oxford Press, 2006), p. 104.

57. John Aloysius Farrell, "Officials' anti-church bias alleged," *The Boston Globe,* August 17, 2001, p. A2. See also: The Brookings Institution, "A Governance Studies Event: White House Issues Report on Barriers to Faith-Based Organizations," August 16, 2001,

http://www.brookings.edu/events/2001/0816faith-based-initiatives.aspx

58. Edgar H. Schein, *Organizational Culture and Leadership,* 2nd ed. (San Francisco: Jossey-Bass, 1992), p. 12.

59. Kurt Lewin, "Frontiers In Group Dynamics: Concept, Method and Reality in Social Science; Social Equilibria and Social Change," *Human Relations,* vol. 1, no. 1 (June 1947), pp. 34-35.

60. Caren Siehl and Joane Martin, "The Role of Symbolic Management: How Can Managers Effectively Transmit Organizational Culture?" In J. G. Hunt, D. Hosking, C. Schriesheim and R. Stewart (Eds.), *Leaders and Managers: International Perspectives on Managerial Behavior and Leadership* (Vol. 7). Elmsford, NY: Pergamon Press, 1984, 227-239.

61. Anne M. Khademian, *Working with Culture: How the Job Gets Done in Public Programs* (Washington, D.C.: CQ Press, 2002), p. 3.

62. Philip Selznick, *Leadership in Administration: A Sociological Interpretation* (New York: Harper & Row, 1957), chap. 4, 151. For more on the role of stories in creating an organizational culture, see:

Joanne Martin and Melanie E. Powers, "Truth or Corporate Propaganda: The Value of a Good War Story," in *Organizational Symbolism,* Louis R. Pondy, Peter J. Frost, Gareth Morgan, and Thomas Dandridge, eds. (Greenwich, CT: JAI Press, 1983), pp. 93-107).

Ian I. Mitroff and Ralph H. Kilmann, "Stories Managers Tell: A New Tool for Organizational Problem Solving," *Management Review*, vol. 64, no. 7 (1975): 18-28.

Edgar H. Schein, *Organizational Culture and Leadership,* Second Edition (San Francisco: Jossey-Bass, 1992), pp. 89-92, 182-184, & 25.

Alan L. Wilkins, "The Creation of Company Cultures: The Role of Stories and Human Resource Systems," *Human Resource Management,* vol. 23, no. 1 (1984): pp. 41-60.

63. Henry Louis Mencken, *A Mencken Chrestomathy: His Own Selection of His Choicest Writings* (New York: Knopf, 1949), p. 617.

64. Unfortunately, this potential for evaluation can have conflicting effects. It can facilitate the performance of simple or well-learned tasks, while inhibiting the performance of complex or novel tasks. Stephen G. Harkins, "Social Loafing and Social Facilitation," *Journal of Experimental Social Psychology,* vol. 23, no. 1 (1987), pp. 1-18.

65. Stephen G. Harkins, "Social Loafing and Social Facilitation," *Journal of Experimental Social Psychology,* vol. 23, no. 1 (1987), pp. 1-18.

66. Robert M. Yerkes and John D. Dodson, "The Relation of Strength of Stimulus to Rapidity of Habit-Formation," *Journal of Comparative Neurology and Psychology,* vol. 18 (1908), pp. 459-482.

67. Yaniv Hanoch and Oliver Vitouch, "When Less is More: Information, Emotional Arousal and the Ecological Reframing of the Yerkes-Dodson Law," *Theory and Psychology,* vol. 14, no. 4 (2004), pp. 427-452.

68. Charles A. O'Reilly, III, and Barton A. Weitz, "Managing Marginal Employees: The Use of Warnings and Dismissals," *Administrative Science Quarterly,* vol. 25, no. 3 (1980). p. 467.

69. Abraham Lincoln, "Address Before the Wisconsin State Agricultural Society, Milwaukee, Wisconsin" (September 30, 1859), in Roy P. Basler (ed.), *The Collected Works of Abraham Lincoln,* vol. III (New Brunswick, N.J.: Rutgers University Press, 1953): pp. 481-482. The more complete quotation is:

> It is said an Eastern monarch once charged his wise men to invent him a sentence, to be ever in view, and which should be true and appropriate in all times and situations. They presented him the words: "And this, too, shall pass away." How much it expresses! How chastening in the hour of pride! How consoling in the depths of affliction!

This research is supported by a grant from the Smith Richardson Foundation