Applying System Dynamics Modeling To Foster a Cause-and-Effect Perspective in Dealing with Behavioral Distortions Associated with a City’s Performance Measurement Programs

Carmine Bianchi\textsuperscript{a} & Daniel W. Williams\textsuperscript{b}
\textsuperscript{a} University of Palermo
\textsuperscript{b} Baruch College
Published online: 25 Mar 2015.

To cite this article: Carmine Bianchi & Daniel W. Williams (2015) Applying System Dynamics Modeling To Foster a Cause-and-Effect Perspective in Dealing with Behavioral Distortions Associated with a City’s Performance Measurement Programs, Public Performance & Management Review, 38:3, 395-425, DOI: 10.1080/15309576.2015.1006471

To link to this article: http://dx.doi.org/10.1080/15309576.2015.1006471

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the “Content”) contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and
APPLYING SYSTEM DYNAMICS MODELING TO FOSTER A CAUSE-AND-EFFECT PERSPECTIVE IN DEALING WITH BEHAVIORAL DISTORTIONS ASSOCIATED WITH A CITY’S PERFORMANCE MEASUREMENT PROGRAMS

CARMINE BIANCHI
University of Palermo

DANIEL W. WILLIAMS
Baruch College

ABSTRACT: This paper aims to show how applying system dynamics methodology to performance management can provide a powerful modeling perspective enabling public sector organizations to prevent, detect, and counteract behavioral distortions associated with performance measurement. A dynamic performance management approach is able to support performance management system designers in outlining and implementing a consistent set of measures that can allow public sector decision-makers to pursue sustainable organizational learning and development. This perspective implies a major shift from a static to a dynamic picture of organizational processes and results. It means framing delays between causes and effects, feedback loops, and trade-offs in time and space associated with alternative scenarios. It also means understanding how different policy levers impact the accumulation and depletion of strategic resources over time, and determining how performance drivers affect end results. An exemplar application of this perspective is outlined in relation to municipal crime-control policies. Unintended behavioral consequences generated by the implementation of the CompStat program (New York Police Department) on reward and performance management systems are framed through the “lenses” of dynamic performance management.


Address correspondence to Daniel W. Williams, Baruch College, D901, 1 Bernard Baruch Way, New York, NY 10010, USA. E-mail: daniel.williams@baruch.cuny.edu
Throughout the last century, but particularly in the last two decades, public management and public policy initiatives have focused on performance through the use of targets, benchmarks, and other relatively narrow managerial goals. Common practices include the use of benchmarks (Ammons, 1996, 1997, 1999; Coe & O'Sullivan, 1993; Eimicke, 1998; Hall & Holt, 2003; Hatry, 2006; Keohley, 1997; Keohley & Abercrombie, 2008; McNair & Leibfried, 1992; Rivenbark & Carter, 2000); the publication of comparison data (Hood, 2012; Micheli & Neely, 2010), the use of PerformanceStat practices (Abramson & Behn, 2006; Behn, 2008), the use of performance budgeting (Buck, 1949; Curristine, 2007; Ebdon & Franklin, 2006; Gross, 1969; Hall & Holt, 2003; Harr & Godfrey, 1992; Hatry, 2007; Lu, 1998; Melkers, 2006; Musell, 2009; Parsons, 1957; Sherwood, 1954; van Nispen & Posseth, 2006; Willoughby, 2004), and the setting of performance targets within public policies, such as George W. Bush's signature program, No Child Left Behind.

Yet these practices, or some of them, are commonly said to be associated with inappropriate and counterproductive behavior (Birnberg, Turopolec, & Young, 1983; Bohte & Meier, 2000; Bouckaert & Balk, 1991; De Bruijn, 2006; Hood, 2006; Kelman & Friedman, 2009; Leeuw, 2003; Meyer & O'Shaughnessy, 1993; Radnor, 2005, 2010; P. Smith, 1993, 1995; Van Dooren, Halligan, & Bouckaert, 2010). This paper examines the consequences of using performance measurement practices for management and policy when the dynamic conditions in which the practices are used are not fully considered by the policy or management decision-makers when designing and implementing performance practices. While the conceptual model developed in this paper is at an early stage of analysis and is, therefore, of more interest to an academic audience, the method of analysis used and the principle of understanding the process while setting measures and targets is of critical significance to practitioners who are in the day-to-day practice of setting actual performance targets.

Although there are many differences between both programs and implementation in various countries, they all exhibit some common characteristics. In particular, when programs focus on output or outcome targets that are to be met without building on an understanding of the process and factors that lead to those outputs and outcomes, they are likely to jeopardize the improvement of service-delivery efficiency and effectiveness. For example, New York City's CompStat may be associated with controversial activities such as stop-and-frisk, which has been challenged in federal court (Eterno & Silverman, 2006; Judge Won't Delay, 2013). Another example is that the performance targets of No Child Left Behind may be associated with reported cheating scandals (Rich, 2012, 2013; Rich & Delaquérière, 2012; Winerip, Severson, & Brown, 2013).
The design of a balanced and comprehensive set of performance standards, based on the understanding of cause-and-effect relationships among the key components of management strategies and related policies, is a fundamental step to avoid the risk of an “illusion of control.” In this context, the dysfunctional implications associated with the behavioral consequences of performance measurement are relevant issues to overcome the risk of inverting means with ends, or of adopting an only short-term view in implementing such programs. Both researchers and practitioners have advocated an outcome-and-learning-oriented perspective, fostering leadership and consensus building in performance management.

The goal of this paper is to develop an initial qualitative dynamic performance management insight model, showing how this model can enhance a better understanding of the causes and effects related to adopted policies, undertaken actions, and targeted results in the public sector. Insight system dynamics modeling is an established practice that can be used to inform the understanding of processes and is highly dependent on graphic demonstration (Warren, 2000, 2008; Winch & Joyce, 2006; Wolstenholme, 1999). It should not be confused with quantitative parameter-setting modeling, which sometimes occurs as a second stage of analysis. The point of such modeling is to identify areas where dynamic factors may have important influence on the way a process occurs. For performance measurement, such conceptualization may lead to a realization that factors not directly within the control of an organization nevertheless require observation, and also may lead to a view that some factors within the control of the organization can be produced in too great a quantity. For the rest of this paper, “dynamic” means changing in time and interactive with the process in which it occurs (i.e., having a feedback characteristic). Its opposite, “static,” means conceptualized as abstract from the process; thus, its value can change in time, but that leads to no consequences for the process in which the static factor is found.

After an analysis of the literature on unintended behavioral consequences of performance management, the paper will demonstrate the benefits arising from the use of a dynamic performance-management view in order to overcome such risks. To apply this modeling to a specific performance practice, the paper examines the CompStat program, adopted in 1994 by the New York Police Department. However, CompStat serves as a model for the application of the method; it is not the primary focus of the paper. CompStat itself has spread to many other localities and many other functions. Baltimore adopted CitiStat, and when the mayor became the governor of Maryland, the state adopted StateStat and derivative programs such as BRACStat (Gibson & Julnes, 2010). New York has ParkStat and NYCHAStat. On the federal government website Performance.gov, there are similarly labeled programs, such as...
FEMAStat and HUDStat. In particular, we show how dynamic performance management applied to performance management programs can strengthen the design of a consistent and comprehensive set of performance measures, and move out of a static picture of organizational processes and results.

**Unintended Consequences of Performance Measurement: An Introduction**

Robert Merton (1940) warned that employees will focus their attention on whatever is actually set as their work expectations. Chris Argyris (1952) gave rise to a new area of research, which is referred to as “reliance on accounting performance measures.” Around this research stream, several scholars have contributed in the last decades.

Hopwood (1972) examined the effects produced by different evaluation styles, such as focus on short- vs. long-term, or accounting data vs. nonaccounting data oriented. He found that a leadership style characterized by a strong emphasis on budget-related performance is significantly associated with job-related tension, which is dysfunctional for the organization. Otley (1978) could not confirm some of these results; he found that subordinates’ dysfunctional behavior, such as invalid data reporting, does not depend on superiors’ reliance on accounting performance measures. Instead, it relates to the consistency between adopted performance measures and the performed task. Brownell (1982) and Hirst (1983) also carried out a similar research.

Related studies emphasize the ambiguity of performance measurement (Vakkuri & Meklin, 2006). They have developed around “bounded rationality” (March & Simon, 1958), organizational behavior (Cyert & March, 1963), ambiguities in the link between information and decision-making (March, 1987), and around rationality and cause-and-effect relationships in decision-making (Simon, 1955). In this regard, cognitive limitations, conflicting interests, uncertainty, paradoxes, and ambivalences often make performance measurement a difficult task. Otley (1978) shows that dysfunctional behavior in using information provided by management accounting systems often originates from an inconsistency between information and the complexity of the task environment, or from technical inadequacy of the information itself (Mintzberg, 1975). This implies that various organizational participants often ignore or even manipulate or falsify information in order to achieve their own personal goals, regardless of the organization's needs.

Hofstede (1967, 1978, 1981) demonstrated that the traditional cybernetic management-control process model may generate unintended and undesirable effects when applied to contexts where objectives are missing, unclear, or shifting, and the outcomes from achieved results are not measurable. “Budget
“gaming” and opportunistic behavior are among the most evident results emerging from a budget process and information whose structure does not fit with the task environment. Hofstede recommended the use of political control as an alternative to cybernetic control, which may fit with those contexts where managerial processes and expected results can be parameterized through standards. In the political-control framework, decisions are based on negotiations and judgment. Decisions often deal with policies, which are not composed only of rational elements, but also consist of cultural values, which are shared by different people belonging to a group.

Ouchi (1979) identified three different logics for management controls: bureaucratic, market, and clan control. Again, such a distinction was based on the characteristics of the task environment, concerning the ability to measure outputs, and the level of knowledge of the transformation process. A “clan” form of control, based on ritual and ceremony, was suggested for those contexts where the ability to measure outputs is low and the knowledge of the transformation process is imperfect (Birnberg et al., 1983). Lack of consideration of the need to adopt a contingency approach in the design of management-control systems, to take into account the characteristics of the organization, may generate an “illusion of control.” This would give rise to a number of dysfunctional consequences for individuals, groups, and the organization as a whole (Dermer & Lucas, 1986).

Though not explicitly labeled performance management, a similar approach can be found in the conceptual framework outlined by Mintzberg (1979, chap. 1), based on the search for consistency between the design of coordinating mechanisms with the characteristics of tasks. As Mintzberg remarks, for complex tasks characterized by unknown processes and the difficulty to measure outputs, coordinating through standardization may generate dysfunctional effects. Therefore, he suggests “mutual adjustment.” Although adopting a task-coordinating mechanism may preclude adopting alternative mechanisms, still a single unit carrying out heterogeneous tasks can use a proper mix of mechanisms to coordinate with other units. Merchant (1982) outlined a similar conceptual model, distinguishing action control from results and personnel control.

**On the Dysfunctional Behavioral Effects Produced by Improper Use of Performance Measures in the Public Sector**

Beginning in the 1990s, Peter Smith (1993) shows that the specific complexity of the public sector may amplify the dysfunctional behavior effects of improperly designed performance measures. The public sector is characterized by a number of relevant attributes that sharply differentiate it from the private...
and “for-profit” sector, including difficulties in (1) achieving consensus on the outputs of delivered services and on the metrics to adopt to measure them, (2) identifying and measuring the outcomes corresponding to such outputs, (3) communicating with diverse stakeholders (citizens, service users, elected officials, contractors, managers) to involve them in using performance measures and adapting their behavior to the improvement of generated outcomes over time.

Other important complexity factors in the public sector are related to the need to:

1. Pursue a proper fit between the profile of performance management systems and the profile of the institutional and political/cultural systems characterizing the society where a regional public sector organization operates (Bianchi et al., 2010; Borgonovi, 1996, p. 105);

2. Search for a sustainable performance development which is based on both an organizational level and an interinstitutional perspective; assessing performance sustainability requires not only a focus on the single organization’s results, but also on how such results contribute to the wider system’s performance, a factor that will affect the single organization in the long run (Bianchi, 2012, pp. 146–149). This implies a need to adopt a joined-up whole-of-government approach (Boyle, 1999; Christensen & Lægreid, 2007, pp. 33–35; Johnson, 2005; Pollitt, 2003). Such a perspective requires collaboration between different agencies, ministries, and other administrative units and different independent public and private institutions within a given region to achieve successful design, implementation, and evaluation of public policy outcomes.

The need for joined-up government is particularly strong when cross-cutting issues are hot topics in the political agenda, as with social policies, encompassing different interconnected sectors, such as immigration, education, healthcare, safety, welfare, and housing. Since public sector units and institutions acting to satisfy such social needs may play different roles, with a substantial level of autonomy, a major threat for government-wide performance is that the focus on agency-level output measures by each player may not lead to the achievement of the wider community outcomes (P. Smith, 1993).

Such a potential inconsistency between outputs and outcomes in performance management systems may result from a failure of decision-makers to take a systemic view of the phenomena they address (Bianchi, 2012). As a result, policy and management decision-makers may often adopt a bounded, short-term, and static view when selecting performance measures. Therefore, the risk of dysfunctional behavioral effects from the improper design and use of measures in public sector performance management may arise not only because of an inconsistency between the adopted performance measures and the characteristics of the task environment, but also because of a static, nonsystemic view taken by performance management systems designers.
With inconsistency between task and measurement, dysfunctional behavioral effects are primarily detected at individual level. As described in the next section, this topic has been widely studied. Such behavior is associated with goal displacement, that is, deliberate actions of employees to pursue individual or group goals that diverge from organizational goals, using such strategies as misrepresentation and gaming.

With failure to consider systemic factors, dysfunctional behavioral effects are detected at organizational level. In subsequent sections, we show that this behavior is characterized by an emerging and gradual rather than deliberate action. It arises from a lack of consistency in the design of performance management systems with other coordination mechanisms, such as the rewards and career systems. This phenomenon gives rise to a weak span of support (Simons, 2005), that is, a low propensity of the organization to produce collaborative actions toward a group of shared goals whose achievement is considered a means to attain both group and individual goals.

If we consider the relationship between outputs and outcomes related to a responsibility area, the relevant boundaries of the dynamic feedback structure generating the observed outcomes associated with the investigated problems are often much broader than those related to the “levers” on which each organizational unit acts. Consequently, a lack of awareness of the mismatch between the system boundaries perceived at an organizational level and those of the broader system generating the outcomes is a major cause of dysfunctional behavior. This is an emergent organizational dysfunction intrinsic to performance management systems, rather than a behavioral dysfunction.

For example, in order to be effective in the long run, crime prevention requires collaboration and policy coordination between different institutions. In fact, the effectiveness of a police department is, in the long run, subordinate to its capability not only to prevent or repress crime, but also the capability of the wider system—schools, courts, other public safety institutions, nonprofit-sector organizations—to keep crime under control by reducing new and reiterated crime inflow and increasing repressed and solved crime outflows.

Figure 1 demonstrates that both prevention and suppression policies are relevant leverage points to sustainable crime control. For example, if the police only focus on dealing with uncontrolled—that is, unsolved—crimes, and were made accountable to a performance measure such as the number of crimes solved, which is an output measure, even attaining a target for this measure alone may not lead to a reduction in the stock of uncontrolled crime in the region, which is an outcome. Though an increasing pattern of solved crimes might signal police efficiency, it would not necessarily imply that the police are effective. Figure 1 shows that crime is kept in control through two balancing loops labeled B1 and B2, which produce effective crime prevention and
suppression policies, allowing a steady reduction to a lower crime level over time.\(^3\) Without synergy between the different actors involved in crime control policymaking, there is a risk that the reinforcing loop originated by reiterated crime would prevail over the balancing loops.\(^4\)

The lack of framing linkages between outputs and outcomes in performance measurement can lead to a breakdown in the coordination between elected officials and administrators and between different independent agencies, and generate a fracture between managerial and political control. They may also tend to encourage a sectorial and departmental view of administration.

Performance management systems designers, policymakers and managers must be aware of such mismatches in order to avoid ritualistic superficial adoption of budgeting and performance measures, leading to an illusion of control and opportunistic behavior. Conversely, broadening the observed system's boundaries, as previously described, can support a shift of focus from measurement to management, from data collection to systematic use of information, from an input or output to an outcome view of organizational results (Matheson, Scanlan, & Tanner, 1997; Moynihan, 2008), thereby using performance management systems as a learning tool (Moynihan & Landuyt, 2009; Vakkuri & Meklin, 2006, pp. 240–242).

Current research into the use of performance information, as compared with its production, is mixed, but frequently suggests that it is limited at best (Behn,

Based on the analysis conducted in this section, to move toward this direction it can be useful to combine an “external” view (i.e., primarily outcome-oriented) with an “internal” view (i.e., primarily output-oriented) in the design of performance management systems for a public agency or any other autonomous public sector organization unit (Bianchi, 2010, pp. 373–375).

**Profiling the Causes Underlying Dysfunctional Behavioral Effects Produced by Improper Use of Performance Measures in the Public Sector**

It is possible to distinguish three major limitations of performance measurement, that is, attribution, representation of quality, and goal displacement (de Lancer Julnes, 2006):

**Attribution** refers to the possibility that the causal connection between outputs and outcomes and corresponding actions might be erroneous, biased, or partial. This error can undermine accountability, performance evaluations, objectives or actions, and motivation.

**Representation of quality** refers to an intrinsic problem in any attempt to measure the results of accomplished actions, and related outcomes (Bouckaert & Halligan, 2008, chap. 8). On the one hand, use of a single performance measure will have the high risk of a simplistic and partial analysis of causation. On the other hand, use of a broad and detailed set of performance measures leads to a less selective analysis and unclear interpretation of emerging results (van Dooren et al., 2010, p. 160).

Flamholtz (1996) defines **goal displacement** as “a lack of goal congruence created by motivation to achieve some goals sought by the organization at the expense of other intended goals.” It may be caused by several phenomena, including suboptimization, selective attention to organizational goals, and inversion of means and ends. **Suboptimization** occurs when the performance of an organizational unit is achieved at the expense of another unit, or the organization as a whole. **Selective attention** occurs when certain goals are pursued at
the expense of other goals. Birnberg et al. (1983, pp. 121–122) labeled such phenomenon “focusing.” Inversion of means and ends occurs when performance measures and reward systems motivate people to achieve intermediate results rather than the overall organizational goals. The last two phenomena have also been labeled “tunnel vision” and “myopia,” respectively (P. Smith, 1993, 1995).

Bohte and Meier (2000) define goal displacement as individual’s tendency to maximize certain outputs at the expense of related outcomes. They identify three major forms of organizational cheating: cutting corners, lying, and biasing samples. Cutting corners occurs when decision-makers focus on quantity rather than quality. Lying occurs when decision-makers try to take advantage of information asymmetries between their agencies and external evaluators, providing information that may put them in a positive light. Biasing samples implies that decision-makers select or report only those cases that lead to mostly positive evaluations. For instance, agency activity may only be directed toward easy cases requiring a minor effort, rather than other more difficult cases.

Other similar dysfunctional behaviors encouraged by an excessive use of outcome-related performance indicators (ORPIs) include (P. Smith, 1993, 1995):

- **convergence**: “an emphasis on not being exposed as an outlier on any Outcome-Related Performance Indicators (ORPI), rather than the desire to be outstanding” (P. Smith, 1993, p. 141);
- **ossification**: “a disinclination to experiment with new and innovative methods” (P. Smith, 1993, p. 141);
- **misrepresentation**: a “deliberate manipulation of data so that reported behavior differs from actual behavior” (P. Smith, 1995, p. 292). Birnberg et al. (1983, pp. 121–124) identified a number of ways to distort the information system, such as: “smoothing,” “biasing,” “filtering,” and “illegal acts”;
- **gaming**: a “deliberate manipulation of behavior to secure strategic advantage” (Radnor, 2005; P. Smith, 1995, p. 298). Van Dooren et al. (2010, pp. 162–165) have qualified similar phenomena as “measure fixation” and “cream skimming,” which is selecting at intake (Behn & Kant, 1999).

Based on this framework, Vakkuri and Meklin (2006, pp. 244–246) observe that the list of possible dysfunctions originates from heterogeneous theoretical and methodological underpinnings, including the theory of accounting measurement (e.g., tunnel vision, misrepresentation, and convergence), optimization theory (e.g., suboptimization), the theory of organizational behavior and learning (e.g., myopia, ossification), and game theory (e.g., gaming). The heterogeneity of the research field may suggest the need for a more theoretical discussion of ambiguity. Vakkuri and Meklin claim that if it is true that “the users may deliberately employ performance measurement systems for their unique individual, organizational and political purposes” (2006, p. 244), it is also true that “performance measurement systems set the conditions for organizational actors to adapt to” (2006, p. 245).
Distortions in interpreting performance information can also result from characteristics of the adopted measurement system rather than to the behavior of people. For instance, this may happen when phenomena are measured more frequently and accurately than in the past, which may generate a feeling of an increase of its intensiveness in comparison to when the phenomena were not so closely monitored (Bouckaert & Balk, 1991).

Detecting, Preventing, and Explaining Unintended Effects of Performance Measurement in the Public Sector: The “Performance Paradox” Theory

The term “performance paradox” (Leeuw, 2003; Meyer & O'Shaughnessy, 1993) frames dysfunctional effects of performance management where phenomenon are characterized by a weak capability of performance indicators to measure and affect performance (Meyer & Gupta, 1994). The performance paradox implies that adopted indicators lose the capability to discriminate between good and bad performers because organizations adapt their performance (i.e., actions) to them without actual improvement (Van Dooren et al., 2010, p. 165; Van Thiel & Leeuw, 2002, p. 271).

These researchers describe positive learning, which is an increased ability over time to reach performance targets without necessarily improving actual organizational performance, particularly in terms of outcomes. They also describe perverse learning, in which actual performance is misrepresented through overstatement or understatement of performance, thus challenging the quality and effectiveness of a performance management system. The underlying cause is often a focus on too few measures, perhaps only one; or on only short-term measures, which should rather be considered as “drivers” (together with other measures) for the attainment of outcome measures that describing the value generated by the public sector to service users and the community. Therefore, while positive learning can be considered a physiological characteristics of the “performance paradox,” requiring periodic revision of adopted performance measures and standards, a disconnect between performance drivers and end-results (and, more broadly, a static approach) as well as a too tight scope of adopted indicators can be considered as recurring pathologies of the investigated phenomenon.

Van Thiel & Leeuw (2002, p. 271) provide an interesting example of the risks of under- or overstatement of performance in the case of the Dutch police. They discuss the use of an aggregate performance measure of the percentage or number of crimes solved by the police. If we treat the percentage of crimes solved as a single critical performance measure, a decrease in it might suggest that police performance is deteriorating. However, the reduction in the
percentage of crimes solved could result from an increase in the stock of crimes because of an increase in the volume of new crimes. The absolute number of crimes solved might remain constant or even increase, yet the ratio of inflow (new crimes) to outflow (solved crimes) could also increase. These circumstances reflect a potential shortage of resources or a simple need to adjust to new circumstances, not poor performance per se. Likewise, an increase in the crimes solved over a number of years should normally lead to a reduction in the stock of criminals; thus making it more difficult to continue improving or even to keep the absolute number of crimes solved. However, a reduction in this measure resultant from a drop in the stock of criminals would not indicate performance deterioration.

This example suggests that where a change in an environmental variable is an intended program objective, targets should be linked to changes in the stock of that environmental variable, particularly when that variable is associated with the change in a single adopted performance target. In the analyzed case, this would mean that the absolute target of crimes to be solved might be reduced as the corresponding stock of potential criminals declines. A problem in implementing this correction could be that it is difficult to estimate the number of criminals, particularly if one considers that there is often a number of unknown crimes, and of potential criminals.

In this case, focusing on a single performance measure—and in such a static manner—can generate unintended consequences. For example, targeting a constant or rising number or percentage of solved crimes, regardless of the reduction in the stock of criminals, might lead to gaming (i.e., perverse learning). The police might increase (even beyond the socially tolerable levels) pressure on the community in search of a number of crimes to solve. This is a typical example of inversion between means and ends.

The problem could be exacerbated by the use of a single aggregate measure of crime, rather than by differentiating crimes, such as felonies from misdemeanors. Moreover, if one considers the number of crimes rather than the number of criminals, performance evaluation might be affected by the percentage of crimes committed by several people together, as, for instance, in the case of vandalism. If we take this phenomenon into account, the adopted performance measure should also have to consider the number of arrested criminals per crime solved.5

Extending the boundaries of the observed system might reduce the risk of performance paradox. For example, rather than focusing on only the crimes solved, the police might also be made accountable on a set of measures that would describe the effectiveness of their crime-prevention efforts. Obviously, one should properly estimate the different delays characterizing the connection between the adoption of crime-prevention policies and their effects on the two inflows related to new committed crimes and new criminals. Although focusing
on output measures may give a feeling of certainty and clarity in performance measurement, combining output with outcome measures is likely to better support purposeful action, coordination between actors, and reporting to different stakeholders. A focus on crime prevention, combined with suppression, would allow the police to take an active role in designing and implementing broader social policies, in coordination with others, such as schools, social services, nonprofits, and associated stakeholders.

Another measure that may help prevent a performance paradox is using mechanisms to track inertial phenomena that gradually deplete resources affecting organizational outputs and outcomes. Such phenomena are characterized by slow continuous flows depleting relevant stocks, which after perhaps long delays suddenly generate destructive effects and are difficult to perceive by static performance measurement systems. Though the destructive effects appear unexpected and unpredictable, they might be tracked in advance, allowing corrective or preventive measures, if the weak signs of change in a number of (often intangible) resources are detected by the performance management system.

Van Thiel and Leeuw (2002, p. 276) provide the example of a fireworks factory explosion that killed more than 20 people in Enschede, Netherlands, in 2000. The tragedy might have been avoided if the monitoring process by the local and central governments, inspectorates, and the fire department over a number of “small” illegal activities had been regularly and effectively carried out. Lack of supervision over these activities allowed automatic renewal of a number of licenses to operate in a residential area. The gradual accumulation of small problems inertially reduced the level of strategic resource of public safety of the area. Although the potential problem was constantly rising, it was not evident to the community, until the tragedy manifested it as a completely new event. The explosion of a fertilizer plant in West, Texas, in the spring of 2013, reflecting storage of anhydrous ammonia without proper federal reporting, may reflect a similar condition (M. Smith, 2013).

This evidence suggests that public performance management processes will benefit from a more systems-oriented and dynamic approach. A performance management system will be more sustainable when it considers how organizational activities contribute to the buildup and deployment of tangible and intangible stocks of strategic resources, with a view to affect a balanced set of outputs and outcomes. This perspective should also consider whether a public sector organization is affected by the activities of other organizations, public or private, when it delivers service users and the community.

In the next section, we will examine how to build and implement dynamic performance management systems to prevent, detect, and counteract the risks of performance paradox in public sector organizations.
Dynamic Performance Management: A Modeling Perspective to Support Organizations to Prevent, Detect, and Counteract Behavioral Distortions Associated with Performance Measurement

An approach to overcome the myopic view in designing and using performance measures is system dynamics (SD). SD is a modeling methodology to map system structure to capture and communicate the behavior driving processes and the quantification of the relationships to produce a set of equations that form the basis for simulating possible system behaviors over time. Insight (qualitative) modeling focuses on mapping that emphasizes approximate graphical representation of the system. The principle is, that if process structure determines system behavior, and system behavior determines organization performance (Davidsen, 1991; Richardson & Pugh, 1981; Sterman, 2000, pp. 28–29), then the key to developing sustainable strategies to improve performance is acknowledging the relationship between processes and behaviors and managing the leverage points (Ghaffarzadegan, Lyneis, & Richardson, 2011). The advantage of using this approach is it places performance measures within the broader context of the system (Bianchi & Rivenbark, 2014), responding to the reality that even simple policy and process changes to impact specific outputs and outcomes are not likely to be that “simple” in organizations (Bianchi, Winch, & Tomaselli, 2008).

A key for applying SD to performance management is the instrumental view of performance (Bianchi, 2012, pp. 153–155), which makes alternative means for improving performance that explicitly identify both end-results and their drivers. To affect such drivers, each responsible entity must build up, preserve, and deploy strategic resources that are systemically linked to each other. Strategic resources are modeled as stocks of available tangible or intangible factors in a given time. Their dynamics depend on the value of corresponding inflows and outflows. Such flows are modeled as valves on which decision-makers can act through their policies, in order to influence the dynamics of each strategic resource, and, through them, performance (Bianchi, 2010).

Sustainable development is attained by maintaining an appropriate balance between strategic resources. Each strategic resource should provide the basis to sustain others in the same system. For instance, both workers and equipment provide capacity, which influences perceived service quality. This affects regional attractiveness, which, in turn, influences population dynamics. A change in regional population will affect workload and perhaps the stock of available financial resources, and eventually capacity and service. The feedback loops underlying the dynamics of the different strategic resources imply that the flows affecting such resources are measured over a time lag. Therefore, understanding how delays influence strategic resources and achieved results becomes a key issue for managing performance in complex dynamic systems.
Managing strategic resources to affect performance is a dynamic and complex task. In fact, intangible resources (e.g., organizational climate, trust, knowledge, and image) are difficult to identify and measure. Furthermore, processes of accumulation and drain affecting the dynamics of strategic resources are inertial, since delays underlying them are difficult for decision-makers to perceive, and also because effects generated by actions taken, or not taken, in the recent or remote past are intertwined with each other, and single causes cannot be easily matched to related effects. A critical tipping point in managing strategic resources to affect organizational performance is associated with the capability of policymakers to (1) identify those strategic resources that most determine success in the environment, (2) ensure that the endowment of such resources is satisfactory over time, and (3) keep a proper balance between the different strategic resources.

Figure 2 examines the Dutch police case using an instrumental dynamic performance management (DPM) model. It divides performance into three linked levels: strategic resources → performance drivers → end-results, which are separated by dotted-line boxes. On the left at the top is a key for the various symbols, and below that is an outline of the conceptual framework. To the right of the outline, these concepts are applied as they may relate to functions and dysfunctions generated by the performance paradox. Focusing on a single static

Figure 2. Dutch Policing Dynamic Performance Management (DPM) Model
performance measure might lead the police to increase the number or percentage of solved crimes regardless of any reduction in the stock of criminals, which would make it more difficult to attain the same performance level in succeeding years. In the top box, Level 1, the police focus on inflows of strategic resources, such as staff and equipment, to offset outflows of the same. They deploy these to address inflows of crimes and criminals, converting them to outflows through solved crimes and identified criminals. The purpose of these actions is to investigate and control crime (a performance driver, shown at Level 2 of Figure 2) and thereby solve crimes (an end-result shown at Level 3).

To counteract the problem of diminishing returns, the police might increase the pressure on the community in search of a number of crimes to solve, eventually reaching an intolerable level of intrusion into the lives of community members. So in this example, the risk of gaming is associated with the possibility that without a deliberate purpose, the police staff and other strategic resources might be deployed to pursue performance drivers, such workload measures as the number of controls or investigations undertaken, with the intent to achieve numerical targets.

If solving crimes and reducing the corresponding stock of criminals contribute to the community's quality of life, an excess of controls and investigations may offset in the long run the positive effects of the described policy, as shown by the second effect on the right side of Level 3 in Figure 2. This might reduce the community's quality of life. This could generate both a reduction in the population level and an increase in the inflow of new crimes. In other words, an excess of focus and intensive efforts on a single policy might generate specific dysfunctional effects over time. In this case, such effects could be related to an increase of the number of crimes per inhabitant (shown on the right side of Level 2 in Figure 2), which would further reduce the community's quality of life, a vicious reinforcing loop.

This analysis suggests a possible way to counteract the behavioral distortions arising from the performance paradox by extending the boundaries of the observed system to adopt policies that might reduce the risk of gaming and unintended behavioral effects. For example, the police might also be made accountable on a set of measures that would gauge the effectiveness of crime prevention. A focus on crime prevention combined with crime suppression would also allow the police to engage in designing and implementing broader social policies in coordination with other actors.

Figure 3 shows how instrumental DPM modeling might extend the system's boundaries. It shows that when considering additional end-results related to the stock of criminals, two more flows are taken into account: the effect on new criminals and the effect on new potential criminals. There are many reasons why addressing criminals leads to new criminals; for example, it produces...
opportunity for those who would otherwise be deterred by competition, and it removes criminals the police already know from the environment. Correspondingly, the stock of potential criminals is added as a relevant strategic resource to include when formulating policies that may contribute to sustainable community security. In addition, a second set of performance drivers is included in this alternative context: the number of preventive actions.

In the next section, we use DPM to frame a number of problematic behavioral issues and related implications on performance management emerging from the CompStat program of the New York Police Department and other similar programs adopted over the last two decades.

Applying Dynamic Performance Management to Analyze the Behavioral Distortions Associated with Performance Measurement: Insights from the CompStat Program of the New York Police Department

In 1994, the New York City police introduced CompStat, a statistics-based performance measurement system aimed at motivating employees, with a specific focus on precinct commanders, to sharply reduce crime (Behn, 2003, p. 591; Silverman, 1999, pp. 88–89, 101). It is claimed that in New York City,
after CompStat was introduced, major crimes declined over 76% (Eterno & Silverman, 2010, p. 428). Following the successful results of CompStat, similar programs or other police performance monitoring programs have been widely adopted by other cities in the United States (Weisburd, Mastrofski, McNally, Greenspan, & Willis, 2003; Willis, Kochel, & Mastrofski, 2010), Australia (Chilvers & Weatherburn, 2004; Mazerolle, McBroom, & Rombouts, 2011; Mazerolle, Rombouts, & McBroom, 2007), the United Kingdom (Hamelin & Spenlehauer, 2013), and elsewhere (Davis, 2010).

Similarly, the “Stat” model has been adopted well beyond policing; for example, CitiStat is

a data-driven management system designed to monitor and improve the performance of city departments in real-time. Implemented in Baltimore in 2000 by then Mayor Martin O’Malley, CitiStat uses basic, inexpensive computer software to track a myriad of government performance indicators. Managers of each city department report to City Hall every two weeks to present their performance data and answer questions from the mayor’s office. The mayor's office uses this data to identify underperformance and press for improvements. (Perez & Rushing, 2007, p. 3)

After Mayor O’Malley was elected governor of Maryland, Baltimore may have lost its enthusiasm for CitiStat (Broadwater, 2012); however, the mayor, now governor, took the practice with him to the statehouse, creating StateStat.8 These various programs draw on the “broken windows” theory; Wilson and Kelling (1982) theorized that to prevent major crime, police must attend to minor quality-of-life offenses and use assertive methods. According to this theory, unattended broken windows are a symptom of neglect in a district, and neglect leads to social disorder, inertial decay of daily quality of life, and more severe property damage, ultimately leading to a spiral of urban decay (Eterno & Silverman, 2012, p. 192). Eterno and Silverman remark that “CitiStat’s timely data permits the assessment and coordination of diverse social services dealing with graffiti, abandoned vehicles, vacant housing, lead paint abatement, urban blight, drugs, and drug treatment” (p. 15).

According to Boba Santos (2012, p. 43), the success of CompStat is explained by two reasons: (1) a very efficient information system enables decision-makers and performance evaluators to rely on up-to-date data on computerized crime, arrest, and “quality of life,” and (2) this information can regularly be used for interactive crime prevention and reduction strategy meetings, where commanders are held accountable for a number of output measures related to crime in their districts. As described by Eterno & Silverman (2012, pp. 24–25), during the weekly meetings at police headquarters, “charts and maps compare index crime … numbers to the same period last year at three levels: weekly, 28 days, and year to date …. Always being compared to last year's numbers can be a real burden, especially if your previous year was exceptional.”9
During a meeting, each commander stands at a podium and presents statistics on crime control in his precinct to the top management (police commissioner, first deputy commissioner, or others). The commander and his collaborators are held accountable for the achieved outputs. His career and rewards are affected by the measured and reported outputs.

It may happen that “commanders are berated and embarrassed in front of their peers because the numbers do not look ‘correct’ ” (Eterno & Silverman, 2012, p. 25), that is, they do not correspond to the performance targets. According to Robert Zink (as cited in Eterno & Silverman, 2012, p. 62), this “has degenerated into a situation where the police leadership presses subordinates to keep numbers low by any means necessary.” Manipulating or “fudging” crime data has been mentioned as a recurring practice adopted by the police to generate the expected numbers to report; for instance: “misclassify crimes from felonies to misdemeanors, under-value the property lost to crime so it's not a felony” (Eterno & Silverman, 2012, p. 27). Therefore, despite reports of success in fighting crime, CompStat and other similar programs can be a cause of dysfunctional behavior and performance paradox.

Behn (2011) defines “Campbell's law” as, “The more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures, and the more apt it will be to distort and corrupt the social pressures it is intended to monitor” Campbell (1969). As this rule is relevant to crime control performance measurement, where Eterno and Silverman (2012, p. 11) argue that focus on the short run is detrimental to the intended long-run effect, “False arrests have been identified as the result of arrest quotas … arrest quotas may encourage police to focus on less difficult and important arrests at the expense of more significant and arduous arrests.”

The remaining figures show how system dynamics can be used to diagnosis performance management systems dysfunctionality and help identify the causes behind a “performance paradox” so that decision-makers can act to avert them. Figure 4 illustrates the distorted effects on reported crime produced by performance measures and by a reward system focused on only short-term and “easy to achieve” output measures (quotas). The balancing loop B1 would be the functional response of the performance management system to the actual crime level. An increase of this level would imply a higher effort to suppress crime, leading (after a delay) to a reduction of crime. However, the use of an unbalanced set of performance measures and of short-term “output-only-based” reward mechanisms could hamper the dominance of the described balancing loop by fostering two other loops: the reinforcing vicious loop R and the balancing loop B2.

The reinforcing loop R is generated by an increasing effort to solve “easy” crimes due to the existing set of performance measurement and rewards systems and, therefore, to report an increasing pattern of such kinds of crimes. This
generates an increase in the distributed rewards, which in turn gives rise to a “search” for new “easy-to-solve” crime to deal with and report. Though such a vicious loop can contribute to reduce crime (in absolute terms), it might not generate the desired effects in dealing with “hard-to-solve” crimes. On the other hand, the described system could also affect the statistics regarding such crimes through data fudging. The distribution of rewards may lead the police to reduce reported hard-to-solve crime and shift part of it to the lower rank of easy-to-solve crime. This may generate a further effort towards the suppression of easy-to-solve crimes. So while the loop B2 dominance replaces the loop B1, it is also the basis for the loop R to further and progressively generate its dysfunctional effects.

In practice, the unbalanced condition may be equated to New York City’s intense focus on small crimes using stop-and-frisk practices, having the possible effect of producing crimes where none exist. Possibly New York does not adequately attend to serious crimes—by, for example, downgrading serious crimes to misdemeanors (Ruderman, 2012). Thus, it may allow the more professional criminal population to grow.
Figure 5 adopts an instrumental DPM approach to detect the dysfunctions caused by lack of alignment between performance management and rewards systems. The figure shows that when performance measurement and rewards generate “false reporting” on easy-to-solve cases, the performance management system becomes the means for the organization to reach the desired rewards. In other words, the performance management system here takes an ancillary role in respect to the rewards system, rather than using it to pursue sustainable organizational development. As illustrated in the area labeled Level 1, the use of strategic resources is primarily diverted by the rewards system to the execution of controls and preventive actions on easy cases. This effort contributes to meet the output targets, shown as flows of easy cases solved.

There are two effects. First, it allows the organization to distribute rewards based on the achievement of the target number of cases to solve (shown in the top box on the right). The large volume of solved easy cases reinforces the process and leads to an expectation of more similar cases, which implies that organization redirects efforts to such cases. On the other hand, although an increase in the outflow of easy cases solved should lead to a reduction in the stock of easy-to-solve crimes, this does not happen, because of the inflow of new false cases. This variable is shown as an output of the rewards system at

Figure 5. A DPM Model to Detect Distortions
the bottom right of Figure 5. The volume of false cases is proportional to effort, which is proportional to the stock of targeted number of crimes to solve.

Therefore, an abnormal behavior of the rewards system here generates a set of end-results, which may allow the rewards regardless of organizational results. On the contrary, the results measured by the performance management system become the internalized method to pursue the rewards. A further vicious effect can be that unattended hard crimes might allow increase of hard-crime criminals who become skilled, leading to a gradual long-run reduction in the city's quality of life.

The insights emerging from the previous examples suggest possible measures to counteract these potential problems. Figure 6 shows that improving the performance management system may lead to greater reporting promptness and selectiveness. This implies investing in performance measurement and information systems and also in organizational design and human capital. Behn (2008, p. 5) argues that one of the main errors committed in implementing CompStat, CitiStat, and similar programs has been the lack of dedicated analytic staff to design proper performance management systems and support decision-makers in interpreting the results emerging from performance measurement.

For instance, as shown in Figure 4, improving the promptness and selectiveness of reporting may allow decision-makers to design policies that better balance

Figure 6. An Improved DPM Model
capacity allocation in solving hard and easy crimes. Balancing the mix of achieved end-results may contribute to the more sustainable development of a public sector organization and the region. This is shown in Figure 4 through the outcome measure for “change in community quality of life.” This variable is, in turn, affected by a performance driver, the “crime seriousness ratio,” which is significantly affected by the stock of hard-to-solve crimes as a share of total crimes.

Conclusions

This paper has shown how dynamic performance management can be used to analyze and diagnose the performance paradox, that is, the weak capability of adopted performance measures to affect performance, especially because of behavioral distortions resulting from the use of available data and information. Although such behavioral distortions are a product of human nature, particularly when quantitative indicators are used to affect decision-making, they can be amplified by improper design of performance management systems.

On the one hand, a possible cause of this phenomenon may be an inconsistency between specific adopted performance measures and the corresponding task environment. In this case, behavioral distortions in using performance measures tend to be rooted at the individual level, and are due to the use of performance measures that do not fit the characteristics of the activities performed within an organization.

On the other hand, behavioral distortions in using performance measures can be primarily related to a bounded and static perspective in the design of overall performance management systems. This phenomenon can imply discontinuity between (1) outputs and outcomes, (2) the political and administrative levels, and (3) different institutions playing complementary roles in pursuing social policies (i.e., lack of joined-up government). In this case, unintended behavioral phenomena that are relevant to performance management are primarily rooted at the organizational or group, rather than individual, level. In fact, the main responsibility for the dysfunctional behavior can be attributed to the inconsistent design of the performance and related systems (e.g., rewards and career), which is the primary cause of a flow of organizational actions in a direction that is not sustainable in the long run.

This second cause of behavioral distortions in performance measurement has been examined through dynamic performance management, with an exemplar case of municipal crime control policies, particularly the CompStat program. Though we believe that this paper has shown the usefulness of dynamic performance management as a method to foster a cause-and-effect perspective in dealing with behavioral distortions associated with city's performance measurement programs, we also believe that further research will be needed to better frame,
and possibly outline the main archetypes related to different typologies of the recurring dysfunctions associated with the analyzed problem contexts.

This paper provides a conceptual basis for empirical examination of performance “Stat” practices using dynamic modeling. Future research should focus on empirical examination of this model to determine its applicability, and more specifically should tune the identified characteristics to criminal justice or other “Stat” style performance management systems and later seek to calibrate such models. Empirical modeling will measure the relationships, verify their positive or negative signs, and possibly identify other relationships that more completely describe the system. When such models are determined, policymakers and managers are in a better position to influence the system to attain the desired results.

Notes

1. New York City Housing Authority Stat, currently offline beginning with the emergency response to Hurricane Sandy.
2. It is not clear how closely the federal practice follows the local practices it is partly modeled after.
3. The signs on the arrows in Figure 1 show direct and inverse relationships between variables. A reinforcing loop is determined by a positive polarity, while a negative polarity implies a balancing loop. While a reinforcing loop generates multiplicative growth in the affected variable's behavior mode and implies instability in the system, a balancing loop fosters a stable behavior (Forrester, 1961).
4. The hypothesized reinforcing loop paradoxically increases crime because the police may lose the main point of their work, as discussed later in this paper.
5. This discussion is not intended to suggest that any existing measurement system exhibits all the potential errors identified. It describes risks and sources of risk.
6. Though the described model is intentionally simple, it is worth remarking that the stock of criminals should be divided into several classes related to the seriousness of committed crimes. Lack of clarity in the method for distinguishing and categorizing crimes might encourage gaming behavior by police decision-makers, with the intent to maximize outputs, performance rating, and related rewards. In fact, police might simultaneously oversolve easy-to-solve crimes (e.g., through stop-and-frisk techniques) and undersolve hard-to-solve crimes. Of course, categorizing crimes may also have a perverse effect if it permits gaming by reclassifying serious crimes to less serious categories. In this regard, it has been emphasized that there is a need to design clear rules to avoid catching the “dolphins” rather than the “sharks” (Maple & Mitchell, 1999, p. 155).
7. To avoid excessive complexity, Figure 3 does not show the link from Level 3 backwards to Level 1 for potential criminals and criminals.
9. Index crime includes: murder, rape, robbery, felony assault, burglary, grand larceny auto, and grand larceny.
11. Rouwette et al. (2004) have recently developed an interesting study on system dynamics applied to crime control.
References


Carmine Bianchi is Professor of Business & Public Management at the University of Palermo (Italy), where he is also the scientific coordinator of CED4 System Dynamics Group. He is the director of the PhD program in “Model Based Public Planning, Policy Design, and Management”, and of the Masters course in “Managing Business Growth through System Dynamics & Accounting Models. A Strategic Control Perspective”. Bianchi is a member of the Steering Committee of the European Master in System Dynamics, an Erasmus Mundus funded program, patronized by the European Commission. He has published in several academic and professional journals, where he also serves as a member of the Scientific Committee. Bianchi has research and consulting experience with public and private sector organizations in designing policies and outlining consistent programs to link strategy and implementation. Consulting and education projects that he has conducted cover strategic planning and control, performance management and reporting, as well as System Dynamics modeling for performance improvement and crisis prevention (Dynamic Performance Management). In the last decade Bianchi has been strengthening an international network related on “Dynamic Performance Management”. He has been collaborating with a number of universities all over the world, ranging from Europe to Asia and the United States.

Daniel W. Williams has taught at Baruch College since 1995, after nearly 20 years with the Virginia Medicaid program. He teaches budgeting, performance measurement, and ethics. His research focuses primarily on budgeting, performance measurement, and the history of public administration. He has shared the Outstanding Paper Award from the International Institute of Forecasters, and the Jesse Burkhead Award from the Board of Directors of Public Finance Publications, Inc.