

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

Carmine Bianchi

Professor of Business & Public Management
Department of European Studies
University of Palermo, Italy
Via Maqueda, 324, 90100 Palermo
Phone: +39.338924446
bianchi.carmine@gmail.com

Daniel W. Williams

Associate Professor
Baruch College
One Bernard Baruch Way, D901
New York, NY 10010
Phone: +1.646-469-0345
Daniel.Williams@baruch.cuny.edu

Abstract

This paper aims to show how applying the system dynamics methodology to performance management can provide public sector organizations a powerful modeling perspective to prevent, detect and counteract behavioral distortions associated to performance measurement. A dynamic performance management approach is able to support performance management system designers to outline and implement 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. This 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 application of this perspective is outlined, in relation to crime control policies at Municipal level. Concerning this, unintended behavioral consequences generated by the implementation of the Compstat program (at the New York Police Department) on reward and performance management systems are framed through the 'lenses' of dynamic performance management.

Key-words: Performance Management, Performance measurement, Behavioral distortions of performance measurement, Municipal performance, Crime Control, Compstat.

1. Introduction.

Particularly in the last decade, increasing efforts have been produced to improve efficiency and effectiveness in the provision of public services. *Doing more with less* has been a recurring label adopted by elected officials and public administrators to remark the goal of various formal performance measurement/management programs undertaken by government jurisdictions and public agencies in the world.

Although the premises of such programs (e.g., in terms of legislative issues, involved roles and responsibilities) have to some extent differed from country to country, it is possible to envisage a number of common crucial issues behind their effective implementation. In particular, a formal compliance to procedures and a narrow focus on the achievement of performance targets – rather than on the understanding of the factors affecting outputs and outcomes – are likely to jeopardize the achievement of the ultimate goals, i.e.: improving efficiency and effectiveness in service delivery. On this regard, 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, has been emphasized as a fundamental step to avoid the risk of an “illusion of control”.

In this context, the dysfunctional implications associated to 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 illustrate how a dynamic performance management view¹ can enhance a better understanding of the causes and effects related to adopted policies, undertaken actions and targeted results in the public sector.

After an analysis of the literature on unintended behavioral consequences of performance management, the paper will illustrate the benefits arising from the use of a dynamic performance management view, in order to overcome such risks.

Then, we will provide an application of such approach to a number problematic issues emerging from the Compstat program, adopted since the year 1994 by the New York Police Department.

¹ The term “dynamic”, associated to “performance management”, is here used to identify a method to support decision makers in an organization to better understand and learn about cause-and-effect relationships between relevant factors affecting organizational outputs and outcomes *over time*. In particular, framing how (both information and material) delays affect such causal relationships is a characterizing attribute of a *dynamic* view of performance management. Furthermore, such view is systemic. In fact, it implies that delays in cause-and-effect relationships are framed by indentifying the relevant system’s structure, which generates observed behavior (i.e. performance). Also, this analysis implies that system’s boundaries are defined. Such boundaries identify the context in relation to which a number of interconnected variables (and decision makers affecting them) are recognized as major determinants of performance outputs and outcomes. Usually, within such boundaries, one may identify not only those variables and decision makers pertaining to a single organization, but also others, which – though they may be external in respect to a given organization – are relevant to affect its performance.

In particular, we will show how dynamic performance management applied to Municipal programs can strengthen the efforts towards the design of a consistent and comprehensive set of performance indicators, and move out of a static picture of organizational processes and results.

2. Unintended consequences of performance measurement: an introduction.

Dysfunctional behavioral effects in using budgets and performance measures to affect people and group behavior towards desired goals have been investigated since a relatively long time ago.

A pioneering study on the field was provided by Argyris (1952), who gave rise to a new area of research, which was referred to as “Reliance on accounting performance measures”. Around this research stream, several scholars have contributed in the last decades (Hopwood, 1972; Otley, 1978; Brownell, 1982; Hirst, 1983).

Such studies emphasized the ambiguity perspective of performance measurement (Vakkuri & Meklin, 2006). They have been both developed – on the one hand – around the research on ‘bounded rationality’ (March & Simon, 1958), on organizational behavior (Cyert & March, 1963) and on the ambiguities in the link between information and decision making (March, 1987), and – on the other hand – around the research on rationality and cause-and-effect relationships in decision making (Simon, 1955). On this regard, it has been remarked that cognitive limitations, conflicting interests, uncertainty, paradoxes and ambivalences often make performance measurement a difficult task.

Otley (1978) remarked how dysfunctional behavior in using information provided by management accounting systems is often originated by lack of consistency between information and the underlying complexity of the task environment, or by 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 the organization’s needs.

Another important stream of research on behavioral distortions in using budgets was started following to the seminal studies of Hofstede (1967; 1978; 1981). His work demonstrated how the cybernetic approach through which the management control process was traditionally modeled – when applied to contexts where objectives are missing, unclear or shifting, and the outcomes from achieved results are not measurable – may generate unintended and undesirable effects. “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. On this regard,

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. According to such alternative framework (i.e., political control), decisions are based on negotiations and judgment. Decisions often deal with policies, which are not composed by only rational elements, but also consist of cultural values, which are shared by different people belonging to a group.

Ouchi (1979) proposed a similar framework. He identified three different logics for management controls, i.e.: 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 to individuals, groups, and the organization as a whole (Dermer & Lucas, 1986).

Though not explicitly referred as performance management or measurement, a similar approach can be found in the conceptual framework outlined by Mintzberg (1979, chapter 1), based on the search of a consistency between the design of coordinating mechanisms with the characteristics of tasks. As Mintzberg remarks, for those complex tasks, characterized by unknown processes and the difficulty to measure outputs, adopting standardization as a coordinating mechanism may generate dysfunctional effects. Therefore, “mutual adjustment” is suggested in these cases. It is worth remarking that, though the adoption of a given coordinating mechanism for the fulfillment of a certain task is alternative to the adoption of others, the use of a proper mix of mechanisms can be appropriated to coordinate a single unit carrying out heterogeneous tasks with other units.

Merchant (1982) outlined a similar conceptual model, distinguishing *action* control from *results* and *personnel* control.

3. On the dysfunctional behavioral effects produced by improper use of performance indicators in the public sector.

In the early ‘90s, based on the described findings developed since the beginning of the ‘50s, a new stream of research started to point out how the specific public sector complexity may amplify the challenges imposed to performance management system designers by the dysfunctional behavioral

effects arising from the improper use of indicators. In fact, the public sector is characterized by a number of attributes that sharply differentiate it from the private and ‘for-profit’ sector.

Among the most relevant specific factors in this context are the difficulties by a public sector organization in: 1) achieving consensus on the outputs of delivered services and on the metrics to adopt to measure them; 2) identifying the outcomes corresponding to such outputs, and measure them; 3) communicating with stakeholders covering different roles (i.e.: citizens, service users, elected officials, and managers) to involve them in using performance measures and adapt their behavior to the improvement of generated outcomes over time (Smith, 1993).

Other important specific 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 territorial public sector organization operates (Borgonovi, 1996, p. 105; Bianchi *et al*, 2010);
- 2) search for a sustainable performance development which is both based on an institutional (or organizational) level, and on an interinstitutional (or territorial) perspective. According to this last 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, p. 146-149). This implies a need to adopt a joined-up – or *whole-of-government* – approach (Boyle, 1999; Christensen & Laegreid, 2009, p.33-35; Johnson, 2005; Pollitt, 2003). Such a perspective requires that – for the successful design, implementation, and evaluation of public policy outcomes – different units (e.g. agencies or ministries) related to a public sector organization and different independent public (and private) sector institutions operating in a given territory should establish collaboration.

The need of joined-up-government is particularly strong when the so-called *cross cutting issues* are hot topics in the political agenda. For example, this is the case of social policies, encompassing different interconnected sectors, such as: immigration, education, health care, safety, welfare, and housing. Since the public sector units and institutions acting to satisfy such social needs (which are referred to individuals, groups, and the community) may play different roles, with a substantial level of autonomy, a major threat for performance management systems is that the focus on single output measures by each player may not guarantee that the achievement of such outputs will lead to the achievement of the wider pursued outcomes to the benefit of the community (Smith, 1993).

Such a potential inconsistency between outputs and outcomes in designing and implementing performance management systems can be described as an effect of a lack of systemic view of the

phenomena around which decision makers' actions and results should be focused, in order to pursue sustainable development (Bianchi, 2012). A bounded, short-term, and static view is often adopted when performance measures are set in the described contexts. Therefore, the risk of dysfunctional behavioral effects from the improper design and use of indicators 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 and non-systemic view taken by performance management systems designers (fig. 1).

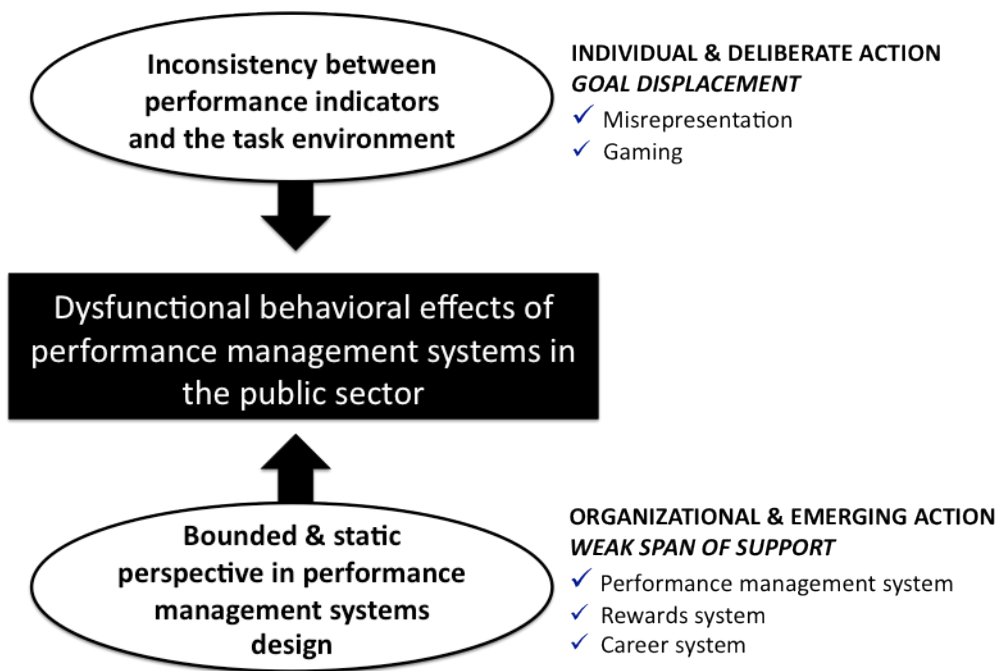


Figure 1: Primary factors leading to dysfunctional behavioral effects of performance management systems in the public sector.

In the first case, dysfunctional behavioral effects are primarily detected at individual level. As described in the next section of this paper, this is the research area on which most of the literature on this topic seems to be focused. Such behavior is associated to goal displacement, i.e. deliberate actions of employees to pursue individual or group goals, which diverge from organizational goals. Regarding such behavior, misrepresentation and gaming are recurring phenomena.

In the second case, dysfunctional behavioral effects are, instead, detected at organizational level. As described later (sections 6 and 7 of this paper), this behavior is characterized by an emerging and gradual – rather than deliberate – action, which arises from a lack of consistency in the design of performance management system with other coordination mechanisms, such as the rewards and career systems. This phenomenon gives rise to a weak span of support (Simons, 2005), i.e. a low

propensity of the organization to produce collaborative actions towards a group of shared goals, whose achievement is considered as a means to attain both group and individual goals.

From the discussion here developed emerges how, if we consider the relationship between outputs and outcomes – related to a given responsibility area – the relevant boundaries of the dynamic feedback structure generating the observed behavior (i.e., outcomes) associated to the investigated problems are often much broader than those related to the ‘levers’ on which each unit is enabled to act. A lack of awareness of the mismatch between the system boundaries perceived at a local (or organizational) level and those characterizing the broader system generating the observed outcomes, which will provide the inputs to the next budget cycle for each agent, is a major cause of dysfunctional behavior, which is primarily intrinsic to performance management systems, rather than to the behavior of specific groups or individuals (fig. 2).

For example, crime prevention/detection/repression, in order to be effective in the long run, would require collaboration and policy coordination between different institutions. In fact, the effectiveness of a police department – in the long run – is subordinate to its capability not only to prevent or repress crime, but also to the capability of the wider system (e.g. schools, courts, other public safety institutions, non-profit sector organizations) to keep crime under control, i.e. to reduce its inflow (new and reiterated crime) and to increase its outflows (repressed and detected crime).

Figure 3 illustrates how both prevention and repression policies are relevant leverage points to enhance sustainable crime control.

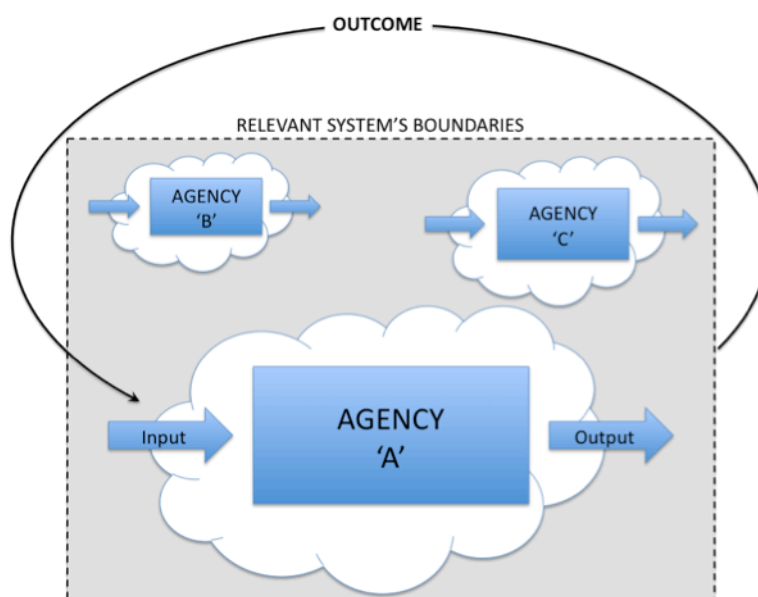


Figure 2: A major cause of dysfunctional behavioral effects in public sector performance management: the mismatch between the relevant system’s boundaries and the boundaries of each agency’s performance management systems.

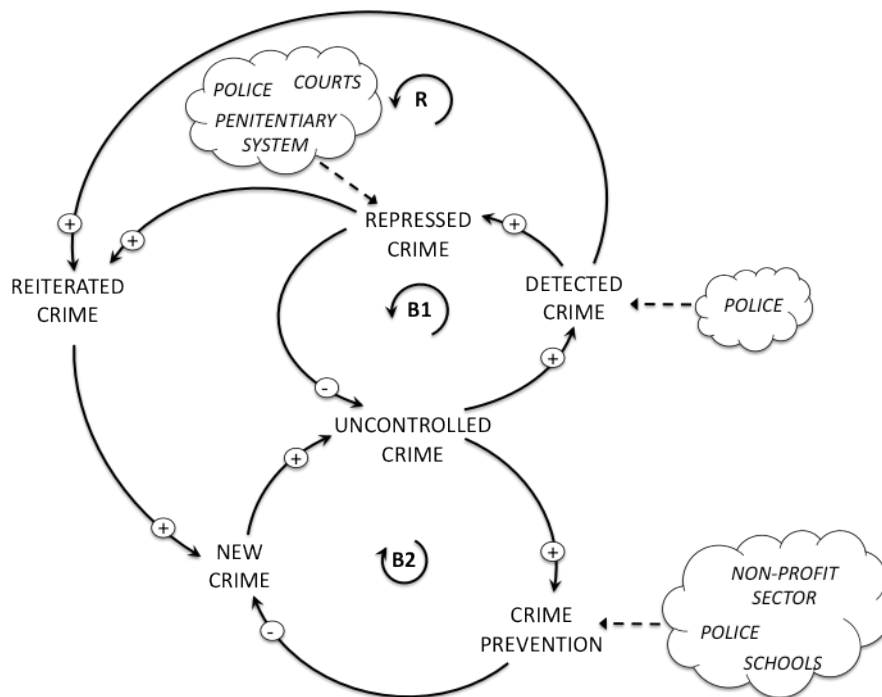


Figure 3: Balancing (or counteracting) and reinforcing (or vicious) loops behind sustainable crime control policies.

For example, if the police would only be focused on dealing with uncontrolled (i.e. undetected) crimes, and it would be made accountable on a main performance indicator, e.g. the number of detected crimes (output measure), even attaining this only measure might not lead to a reduction in the stock of (uncontrolled) crime in the territory (outcome measure). Though an increasing pattern of detected crimes by the police might signal efficiency in using staff and means, this would not necessarily imply that the police are effective too. In fact, without a synergy between different actors involved in crime control policy making, there would be the risk that the reinforcing loop originated by reiterated crime would prevail over the two balancing loops (B1 and B2), from effective crime prevention and repression policies, that would allow one to reduce and keep steadily low the crime level, over time ².

² It is worth remarking that the ‘+’ and ‘-’ signs of the arrows in fig. 3 mean a direct and inverse relationship between a variable and another one, respectively. A reinforcing loop is determined by a positive polarity (if we multiply the algebraic signs of the links in the loop). A balancing loop implies a negative polarity, which is again determined by the multiplication of the algebraic signs of the links in the loop. While a reinforcing loop generates exponential growth in the affected variable’s behavior mode (so it implies an instability in the system), a balancing loop fosters a stable behavior. The principles behind the study of the structure and behavior of systems over time, through the identification and analysis of dominant feedback loops, refer to system dynamics (Forrester, 1961). More details on System Dynamics applied to performance management will be illustrated in the section 6 of this paper.

The described phenomenon of lack of framing linkages between outputs and outcomes in performance measurement is often a major cause of lack of efforts towards the development of formal and/or informal coordinating mechanisms between elected officials and administrators and between different independent agencies. Such phenomena may generate a fracture between managerial and political control. They may also tend to encourage a sectoral and departmental view of administration.

The perception of such mismatches by performance management systems designers, policy makers and managers in public agencies is fundamental to avoid a ritual and superficial adoption of budgeting and performance indicators, leading to an ‘illusion of control’ and opportunistic behavior. Conversely, in order to make use of performance management systems as a learning-oriented tool (Moynihan – Landuyt, 2009; Vakkuri & Meklin, 2006, p. 240-242), 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 *et al*, 1997; Moynihan, 2008).

Current use of performance measurement is surprisingly limited (Behn, 2002, 2003; de Lancer Julnes, 2008; de Lancer Julnes & Holtzer, 2001). According to Sanger (2012, p. 185), “performance measurement rarely leads to improvement government performance or more efficient and accountable municipal management”. Committed leadership (Bianchi & Rivenbark, 2012-a, p. 521; Moynihan, 2008, p. 78) is the key to implement performance management systems, according to a learning-oriented approach (Sanger, 2008). Performance management ought to change behavior.

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

4. Profiling the causes underlying dysfunctional behavioral effects produced by improper use of performance indicators in the public sector.

It is possible to distinguish three major limitations of performance measurement, i.e. attribution, representation of quality, and goal displacement (De Lancer Julnes, 2006).

“Attribution” refers to the possibility that the causal connection between a given set of outputs and outcomes, and a corresponding set of actions, might be erroneous, biased, or partial. This

phenomenon can be a major problem to pursue and implement the accountability principle, to evaluate performance, correct objectives or actions, motivate and reward people.

“Representation of quality” refers to an intrinsic problem in any attempt to measure the results of accomplished actions, and related outcomes (Bouckaert & Halligan, 2008, chapter 8). On the one hand, the more one refers to only a single performance measure, the higher the risk to make a simplistic and partial causation analysis will be. On the other hand, the broader and detailed the adopted performance indicators set is, the less selective the analysis and interpretation of emerging results will be (Van Dooren *et al*, 2010, p. 160).

A third important limitation of performance measurement is “goal displacement”.

Flamohltz (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”. According to Flamohltz, goal displacement may be caused by several phenomena, including: sub-optimization, selective attention to organizational goals, and inversion of means and ends. *Sub-optimization* occurs when 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 *at al* (1983, p. 121-122) define a similar phenomenon as “Focusing”. *Inversion of means and ends* occurs when performance measures and reward systems tend to motivate people to only achieve certain intermediate results, rather than also the overall (i.e. final) organizational goals. The last two phenomena have also been defined as “Tunnel vision” and “Myopia”, respectively (Smith, 1993, 1995).

Bohte and Meier (2000) define goal displacement as individual’s tendency to maximize certain outputs at the expense of related outcomes. In respect to this, they identify three major forms of organizational cheating, i.e.: “cutting corners”, lying, and biasing samples. The “Cutting corners” practice occurs when decision makers focus on quantity, rather than quality. “Lying” is a practice that occurs when decision makers try to take advantage of information asymmetries between their agencies and external evaluators, with a purposeful action oriented to provide information that may put them in a positive light. “Biasing samples” is a third possible goal displacement technique implying that decision makers select and report only those cases that may mostly lead to positive evaluations. For instance, agency activity may only be directed towards those easy cases, requiring a minor effort than other more difficult cases.

Other similar dysfunctional behavior modes, which are encouraged by an excessive use of outcome-related performance indicators (ORPIs) in the public sector, have been identified as (Smith, 1993, 1995):

- *convergence* i.e. “an emphasis on not being exposed as an outlier on any ORPI, rather than the desire to be outstanding” (Smith, 1993, p. 141);
- *ossification*, i.e. “a disinclination to experiment with new and innovative methods” (Smith, 1993, p. 141);
- *misrepresentation*, i.e. a “deliberate manipulation of data so that reported behavior differs from actual behavior” (Smith, 1995, p. 292). Concerning this phenomenon, Birnberg *et al* (1983, p. 121-124) identified a number of corresponding methods of distorting the information system, such as: “Smoothing”, “Biasing”, “Filtering”, and “Illegal acts”;
- *gaming*, i.e. a “deliberate manipulation of behavior to secure strategic advantage” (Smith, 1995, p. 298; Radnor, 2005). Van Dooren *et al* (2010, p. 162-165) have qualified similar phenomena as: “measure fixation” and “cream skimming”. This last case is related to the possibility that – when confronted with output measures – organizations may be tempted to select the intake (Behn & Kant, 1999).

Based on the described conceptual framework provided by the literature on the dysfunctional behavioral effects produced by performance indicators in the public sector, Vakkuri & Meklin (2006, p. 244-246) propose a thought provoking theoretical extension of the existing body of knowledge on the field. They observe that the list of possible dysfunctions originates from heterogeneous theoretical and methodological underpinnings, which range from theory of accounting measurement (e.g. tunnel vision, misrepresentation, and convergence), to optimization theory (e.g. sub-optimization), to 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 of a more theoretical discussion of ambiguity. Such logical process of analysis has led the authors to identify two main contexts related to the described phenomena. On this regard, the authors claim that, if it is true that “the users may deliberately employ performance measurement systems for their unique individual, organizational and political purposes” (Vakkuri & Meklin, 2006, p. 244), it is also true that “performance measurement systems set the conditions for organizational actors to adapt to” (Vakkuri & Meklin, 2006, p. 245).

In other cases, distortions in interpreting performance information can be mainly due to the characteristics of the adopted measurement system, rather than to the behavior of people. For instance, this happens when a phenomenon is measured more frequently and accurately than in the past, generating the feeling of an increase of its intensiveness or frequency over time, or in comparison to other contexts where the same phenomenon is not so closely monitored (Bouckaert & Balk, 1991).

5. Detecting, preventing, and explaining unintended effects of performance measurement in the public sector: the “performance paradox” theory.

The several conceptualizations shown in the last section about the possible causes of dysfunctional effects of performance management in the public sector have been effectively framed by the so-called “Performance paradox” theory. Based on a private sector research, Meyer and O’Shaughnessy (1993) coined the term “performance paradox” (Leeuw, 2003) to mean a phenomenon characterized by a weak capability of performance indicators to measure and affect performance (Meyer & Gupta, 1994). The performance paradox implies that adopted indicators lose their own capability to discriminate good and bad performers, because organizations adapt their performance (i.e. people actions) to them (Van Dooren *et al*, 2010, p. 165).

Later, Van Tiel & Leuw (2002, p. 271), provided relevant examples of how the deterioration of performance indicators may affect also the public sector.

Apart from the so-called *positive learning* – implying an increased capability over time by people to systematically reach performance targets, although this may not automatically mean an improvement in the overall organizational performance (especially in terms of outcomes) – the authors describe the phenomenon of *perverse learning*. Regarding this, overrepresentation and underrepresentation of performance provide a real challenge to the quality and effectiveness of a performance management/measurement system, generated by the “performance paradox”. In these two cases, reported performance is actually – respectively – worse and better than the actual one. The underlying cause is often a focus on a single indicator or on a too bounded set of them, and/or on only single short-term indicators, which should rather be considered as “drivers” (together with other measures) for the attainment of outcome (or “end-result”) indicators, describing the value generated by the public sector to service users and the community.

Therefore, while positive learning can be considered as a physiological characteristics of the “performance paradox”, requiring periodical revision of adopted performance measures and standards, a disconnection between performance drivers and end-results (and, more broadly speaking, 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 & Leuw (2002, p. 271) provide an interesting example of the risks of *under* or *overrepresentation* of performance by discussing the case of the Dutch police. They observe how focusing the only and aggregate performance measure of the percentage or the number of crimes solved by the police in a given time may give rise to wrong conclusions. In fact, if we consider the percentage of crimes solved as a single performance measure, a decrease in such indicator would suggest that police performance is deteriorating. However, this might not be necessarily true if –

although more perpetrators have been arrested, prosecuted and penalized than before – the reduction in the percentage of crimes solved would result from an increase in the stock of crimes, due to an inflow of new committed crimes which is higher than the outflow of crimes solved in a give time period. Likewise, an increase in the crimes solved over a number of years (which would indicate a better performance) is likely to determine – other things being equal – a reduction in the stock of perpetrators. This would make more difficult for the police to further improve – or even to keep stable – the number of crimes solved, in the future. However, a reduction in this indicator, which is primarily due to a drop in the stock of perpetrators – generated by the past good police performance – would not indicate a performance deterioration.

This example suggests, first of all, that – particularly when a single performance measure is adopted – a special attention should be paid in setting and explaining the value of the standard. In the analyzed case, this would mean that the absolute standard (i.e., the number of crimes solved) might be reduced in consideration of the reduction of the corresponding stock. A problem in implementing this correction could be due to the difficulties in estimating the number of perpetrators, particularly if one considers that – beyond the notified crimes – there is often a number of unknown crimes, and of potential perpetrators (which can be very hard to estimate), who might commit crime in the future.

In this case, focusing on a single performance indicator – and in such a static manner – might generate unintended consequences. In particular, for example, forcing the police to keep the number of solved crimes (or the percentage of solved crimes on total) stable – regardless the reduction in the stock of perpetrators – might give rise to gaming phenomena (i.e. *perverse learning*). In fact, the police might be inclined to increase (even beyond the socially tolerable levels) the pressure on the community in the search of a number of crimes to deal with and to report. This is a typical example of inversion between means and ends.

Furthermore, the problem of evaluating police performance in the described context could be exacerbated by the use of a single and aggregate measure of crime, rather than by differentiating – for example – felonies from misdemeanors. Also, if one considers the number of crimes – rather than the number of criminals too – performance evaluation might be affected by the percentage of crimes committed by several people together, like it may happen – for instance – in the cases of vandalism. If we take into account this phenomenon, the adopted performance measure should have also to consider the number of arrested criminals per crime solved.

Extending the boundaries of the observed system, in terms of adopted measures, 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

its 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 the new committed crimes and the new perpetrators. Though it is true that focusing on output measures may give a feeling of certainty and clarity in performance evaluation, it is also true that 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 to crime repression would also allow the police to take an active role in designing and implementing broader social policies, in coordination with other actors, especially those operating in the public and private “not-for-profit” sector (for instance, schools and associated stakeholders).

Another possible measure that may help public sector organizations to detect and prevent a “performance paradox” is introducing in the performance management system proper mechanisms that can enable decision makers to perceive and track *inertial* phenomena that gradually generate the depletion of resources primarily affecting organizational outputs and outcomes. Such phenomena are difficult to perceive by static performance measurement systems. In fact, they are characterized by slow – but continuous – flows depleting relevant stocks, which (after even long delays) will suddenly generate destructive effects. Though the negative events produced by such phenomena may often eventually seem unexpected – and, therefore, unpredictable – they might have been tracked well in advance and corrective or preventive measures might have been timely adopted if the weak signs of change in a number of (often intangible) resources would have been detected by the organization performance management system.

Van Thiel & Leuw (2002, p. 276) provide another clear example of this phenomenon, in relation to a tragic event that took place in the Dutch city of Enschede in the year 2000. By that time, a firework factory explosion killed more than 20 people. The tragedy would have, probably, been avoided if the current monitoring process by the local and central governments, inspectorates, and the fire department over a number of ‘small’ illegal activities would have been regularly and effectively carried out. Conversely, lack of supervision over these activities gave rise to automatic renewal of a number of licenses to operate in the residential area. Therefore, the gradual accumulation of small problems inertially reduced the level of safety (strategic resource) of the area. Though the potential problem was constantly rising, it was not evident to the community, until the tragedy manifested it as a completely new event ³.

³ The recent explosion of a fertilizer plant in West Texas, reflecting storage of anhydrous ammonia without proper federal reporting may reflect a similar condition. See: Smith (2013) Records: Texas plant hadn't told feds about explosive fertilizer. Retrieved June 11, 2013, from www.cnn.com/2013/04/25/us/texas-explosion-plant

The discussion developed so far provides the logical premise to claim for a more systems-oriented and dynamic performance management/measurement process in public sector organizations. A proper method to frame performance sustainability should have to consider how the activities carried out by the organization contribute to build up and deploy a bundle of – tangible and intangible – stocks of strategic resources, with a view to affect a balanced set of outputs and outcomes. Such perspective of performance should also consider whether a single public sector organization is contributing with other organizations to generate and deliver value to service users and the community.

In the next section of this paper we will illustrate how to build and implement dynamic performance management systems to prevent, detect, and counteract the risks of performance paradox in public sector organizations.

6. Dynamic performance management: a modeling perspective to support organizations to prevent, detect and counteract behavioral distortions associated to performance measurement.

One approach to overcome the myopic view in designing and using performance indicators to affect decision makers' behavior in the public sector is System Dynamics (SD).

SD is a modeling methodology that is adopted to map system structure to capture and communicate an understanding of 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. The underlying principle is that, if process structure determines system behavior, and system behavior determines organization performance (Davidsen, 1991; Richardson – Pugh, 1981, p. 15; Sterman, 2000, p. 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.

The advantage of using this approach is placing performance measures into the broader context of the system (Bianchi – Rivenbark, 2012-b), 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 *et al.*, 2008).

A key for applying SD to performance management is the so-called “instrumental” view of performance (Bianchi 2012, p. 153-155). The instrumental view implies that alternative means for improving performance be made explicit. In this regard, it is necessary to identify both end-results

and their respective drivers. To affect such drivers, each responsibility area must build up, preserve, and deploy a proper endowment of 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 therefore—through them—performance indicators (Bianchi, 2010).

The maintenance of an appropriate balance between strategic resources is the key to sustainable development. 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 territory attractiveness, which, in turn, influences population dynamics. A change in the population that a municipality must serve 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 to manage performance in dynamic complex 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 a 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 policy-makers to: a) identify those strategic resources that most determine success in the environment where an organization or different organizations operate; b) insure that the endowment of such resources is satisfactory over time, and c) keep a proper balance between the different strategic resources.

Figure 4 frames the logical structure of a generic dynamic performance management (DPM) model, according to the “instrumental” view. It shows how the *end-results* provide an endogenous source in an organization to the accumulation and depletion processes affecting strategic resources. In fact, they can be modeled as *in-* or *out-*flows, which change over a given time span the corresponding stocks of strategic resources, as a result of actions implemented by decision makers. End-results that most synthetically measure the overall organizational performance are flows affecting the

accumulation of corresponding strategic resources that cannot be purchased. These are: 1) resources generated by management routines, and 2) financial resources (Bianchi, 2012).

Figure 4 also highlights that *performance drivers* are a measure of factors on which to act in order to affect the final performance. They are measured in relative terms, i.e. as a ratio between organizational or territorial (inter-institutional) performance perceived by the community or specific groups of service users and a benchmark, or target. Such denominator must be gauged in relation to either perceived past performance, or users’ expectations, or even competitors’ (e.g. other territories) performance. For instance, if related to an end-result such as the number of new business initiatives undertaken in a urban area in a given time span, corresponding performance drivers could be associated to the (financial and socio-political) perceived stability of a region, and to the perceived transparency and promptness of the public sector (e.g., in terms of authorization protocols or supply of various services, such as those related to security, transportation, social assistance, housing).

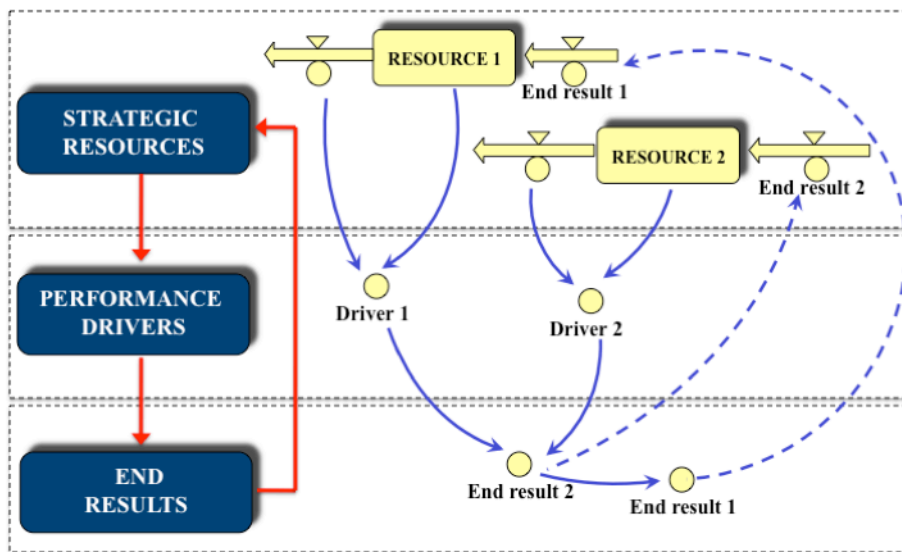


Figure 4: A dynamic performance management view framed through an “instrumental” perspective.

In order to affect such drivers in the desired direction, each decision maker must build up, preserve and deploy a proper endowment of tangible and intangible strategic resources systemically linked each other.

The growth of a single organization and of a territorial community (e.g., a urban area) embracing different institutions can be sustainable if the rate at which end-results change the endowment of corresponding strategic resources is balanced. This implies that each institutional decision maker is able to increase the mix of strategic resources and that this increase is not obtained by reducing the endowment of the wider strategic resources in the territory.

End-results can be measured over a sequential chain and positioned on several layers. “Last layer” end-results are those changing the endowment of strategic resources that cannot be purchased. To affect the results positioned on this “last layer,” further layers must be identified. For example, cash flows can be affected by the current income and net working capital flows. These more detailed financial measures are, in turn, affected by non-monetary, end-results. So, activity volumes affect revenues and the net working capital flow. They also affect purchase volumes, which impact on purchase costs and (through purchases on credit and the change in inventory) on the net working capital flow. Therefore, activity volumes can be located on a first layer of end-results. Such results can be affected through performance drivers.

An application of DPM through the instrumental view to the case of the Dutch police, described at the end of the last section, is framed in figure 5.

The three-level framework (strategic resources → performance drivers → end-results) here depicted contributes to identify and to counteract possible dysfunctions generated by the ‘performance paradox’.

As commented in the last section, focusing performance measurement on a single and static indicator might force the police to generate actions to increase the number of solved crimes (or the percentage of solved crimes on total) – regardless the possible consequential reduction in the stock of perpetrators, that would make more a more difficult to attain the same performance of the previous years. To counteract the problem, the police might be inclined to increase (even beyond the socially tolerable levels) the pressure on the community in the search of a number of crimes to deal with and to report. In other words, in this example the risk of gaming is associated to the possibility that – although without a deliberate purpose – the police staff and other means (strategic resources) might be deployed to pursue a set of performance drivers (mainly represented by workload measures, such as the number of controls or investigations undertaken in a given time), with the intent to “make the numbers” needed to achieve the final targets.

If solving crimes and reducing the corresponding stock contributes to improve the community’s quality of life, an excess of controls and investigations – perhaps beyond socially tolerable levels – may offset in the long run the positive effects of the described policy. This might reduce the community’s quality of life. This might also generate a reduction in the level of population and an increase in the inflow of new committed crimes. In other words, an excess of focus and intensive efforts on a single policy might generate side effects in the long run. In this case, such effects could

be related to an increase of the number of crimes per inhabitant, which would generate a further reduction in the community's quality of life (vicious reinforcing loop)⁴.

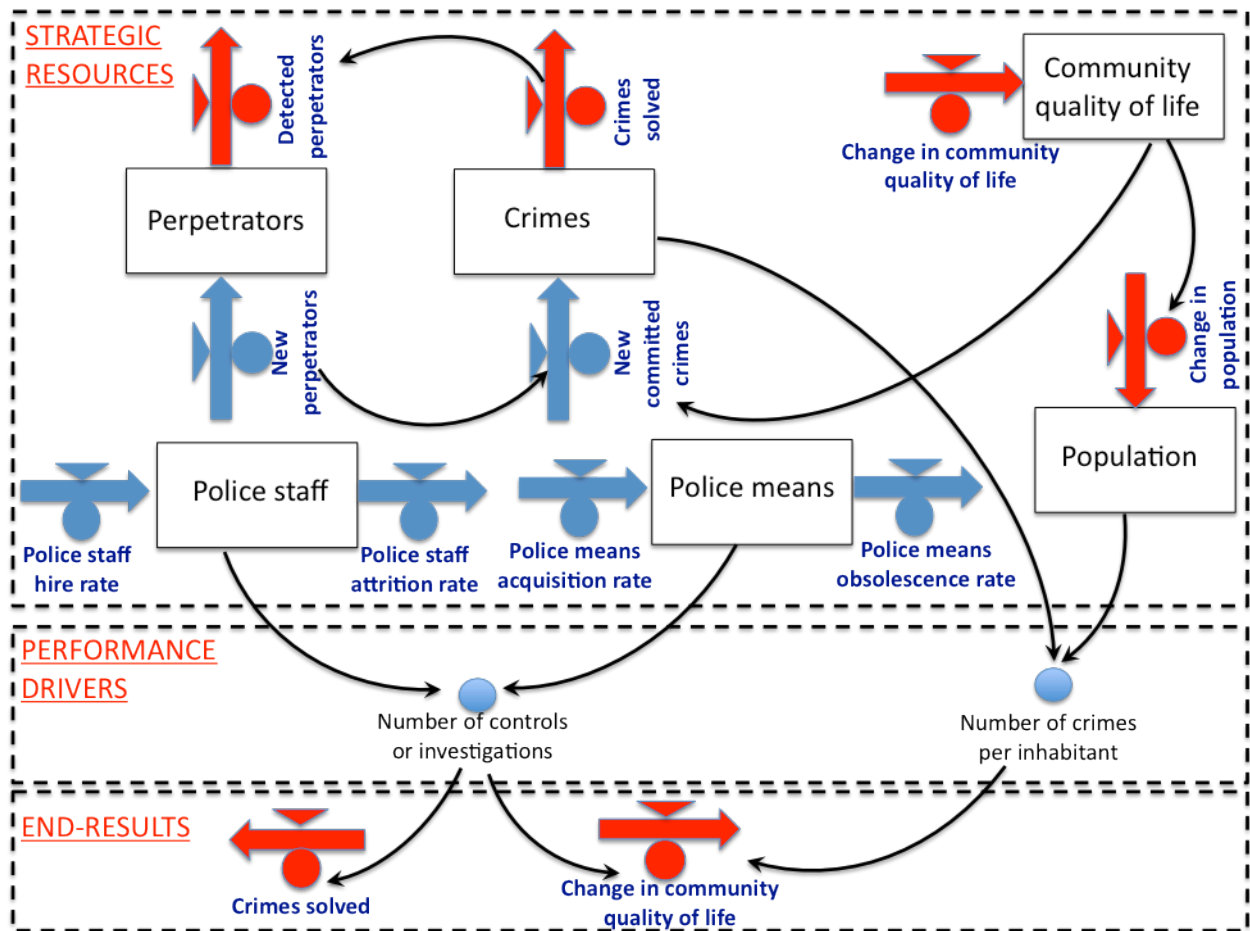


Figure 5: Using ‘Dynamic Performance Management’ to model the unintended effects produced by the performance paradox: the Dutch police example.

This analysis suggests a possible a possible way to counteract the behavioral distortion that a performance paradox might generate in this case. As also described in the previous section of the paper, extending the boundaries of the observed system, in terms of adopted measures and corresponding policies, 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 to crime repression

⁴ Though the described model has been intentionally kept simple, it is worth remarking that the stock of criminals should have been here divided into several classes, related to the level of seriousness of committed crimes. As it will be commented with more detail in the next section of this paper, lack of clarity in the method for distinguishing and categorizing crimes, according their level of seriousness, might originate ‘gaming’ behavior by the 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 & frisk’ techniques) and undersolve ‘hard to solve’ crimes. On this regard, it has been emphasized the need of designing clear rules to avoided to catch the ‘dolphins’, rather than the ‘sharks’ (Maple 1999, p. 155).

would also allow the police to take an active role in designing and implementing broader social policies, in coordination with other actors.

Figure 6 shows how this extension of the explored system’s boundaries might be made possible by adopting a modeling approach based on DPM, according to an instrumental view. The figure shows how – among the end-results – two more flows are taken into account for performance evaluation, i.e.: the new perpetrators and the new potential perpetrators. Correspondingly, also the stock of potential perpetrators is added as a relevant resource to which to refer in order to formulate policies that may contribute to a sustainable community development. Also, a second set of performance drivers is included in this alternative context, i.e.: the number of controls and preventive actions.

Based on the analysis developed so far, in the next section we will frame, through a DPM approach, a number of problematic behavioral issues (and related implications on performance management/measurement) emerging from the Compstat program of the New York Police Department, and other similar programs adopted later in order to control crime, by other cities of the world.

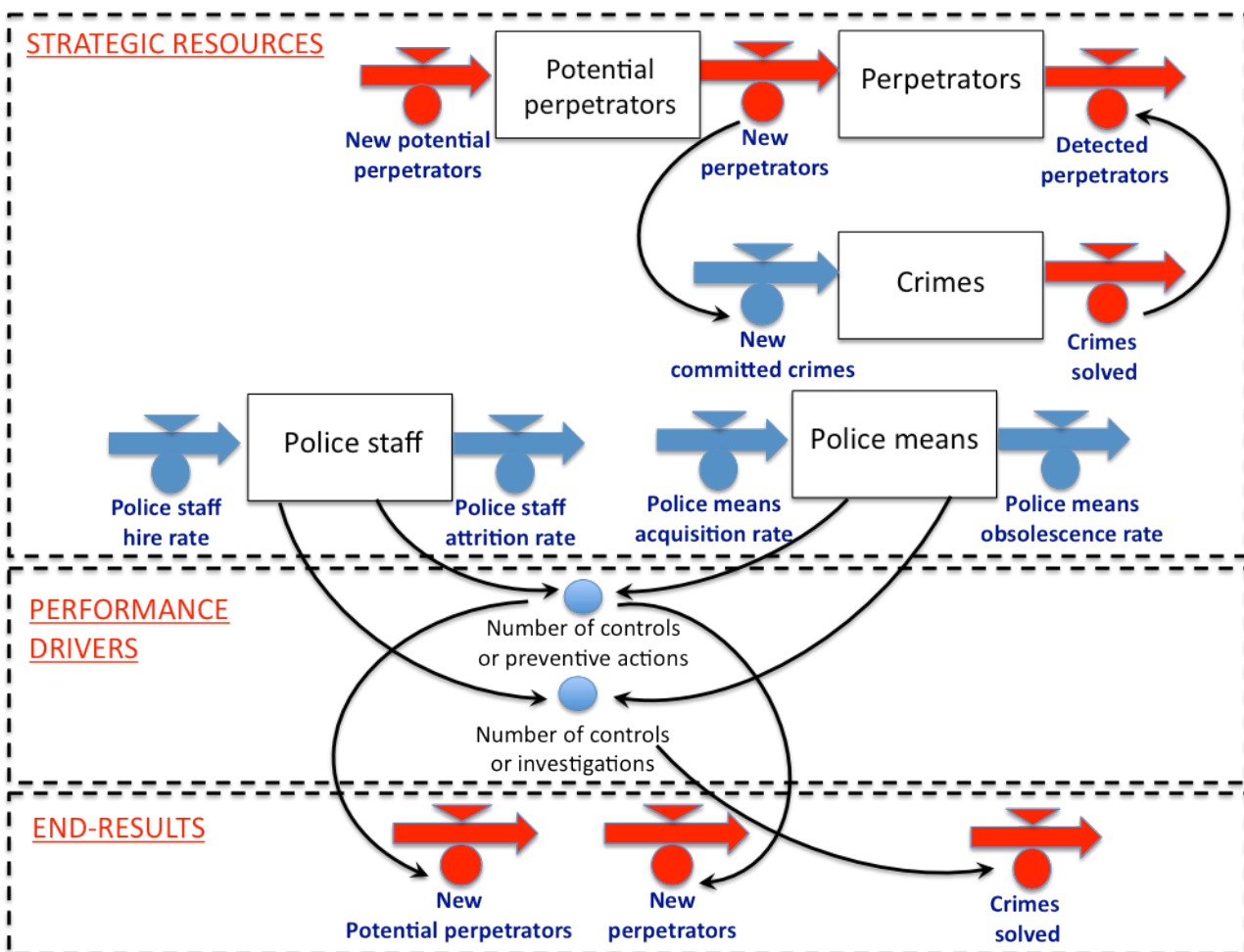


Figure 6: Using ‘Dynamic Performance Management’ to counteract the unintended effects produced by the performance paradox: the Dutch police example.

7. Applying dynamic performance management to analyze the behavioral distortions associated to performance measurement: insights from the “Compstat” Program at New York Police Department.

In the year 1994, the New York Police Department introduced CompStat, a statistics-based performance measurement system aiming to motivate employees, with a specific focus on precinct commanders, with the purpose to sharply reduce crime (Silverman, 2001, p. 88–89, 101; Behn, 2003, p. 591). It is claimed that in New York City, after Compstat was introduced, major crimes have declined over 76 % (Eterno & Silverman, 2010, p. 428).

Following to the successful results of Compstat, similar programs have been subsequently adopted by many other city agencies, not only in the USA (e.g. Australia, Scotland, Canada).

For instance, a similar program is Citistat, i.e. “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). Both Compstat and Citistat draw many of their conceptual underpinnings from the so-called ‘Broken Windows’ theory. Wilson & Kelling (1983) firstly theorized that – in order to prevent major crime – police must be attentive also to minor quality of life offenses, and use assertive methods, starting from avoiding leaving a broken window unattended. According to this theory, unattended broken windows are a symptom of negligence for a district. This would give rise to disordered behavior, inertial decay of daily quality of life, and more severe property damage, leading to a spiral of urban decay (Eterno & Silverman, 2012, p. 192). Based on this assumption, as Eterno & Silverman remark (2012, p. 15), “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”.

According to Boba Santos (2013, p. 43), the success of Compstat can be 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”; 2) this information can be used on a regular basis for interactive crime prevention and reduction strategy meetings, where commanders are held accountable on a number of output measures related to crime data in their districts. As described by Eterno & Silverman (2012, p. 24-25), during the weekly meetings at police headquarters, “charts and maps compare index crime (i.e., murder, rape, robbery, felony

assault, burglary, grand larceny auto, and grand larceny) 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". During a meeting, each commander stands at a podium presenting to the top management (e.g. police commissioner, first deputy commissioner, chief of department, and deputy commissioner of operations) statistics on crime control in his precinct. The commander and his collaborators are held accountable on the achieved outputs. His career and rewards will be 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), i.e. they do not correspond to the pre-set performance standards. According to Zink (2004), 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, in spite of the reported success in fighting crime, Compstat and other similar programs are also considered as a major cause of dysfunctional behavior and performance paradox.

By quoting Campbell (1976), Behn (2011) remarks how "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". The so-called "Campbell's law" is considered as relevant to the experience of various performance measurement programs in crime control policy implementation. Eterno and Silverman (2012, p. 11) emphasize how the focus on the short run to the detriment of the long run is an effect of the commented phenomena. They report as follows: "False arrests have been identified as the result of arrest quotas". They also remark how "arrest quotas may encourage police to focus on less difficult and important arrests at the expense of more significant and arduous arrests"⁵.

Based on the analysis developed so far, figure 7, 8, and 9 show how System Dynamics, when applied to the analysis and diagnosis of dysfunctionalities in performance management/measurement systems, can support the designers of such systems and other key decision makers in identifying the causes behind a 'performance paradox', and to act on them⁶.

⁵ "In 2005, New York's city's patrolmen's union reported that productivity arrest quotas resulted in the arrest of an 80-year-old man for feeding pigeons and pregnant woman for sitting down to rest on a subway station" (Murray, 2005).

⁶ Rouwette *et al* (2007) have recently developed an interesting study on System Dynamics applied to Crime Control.

In particular, fig. 7 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 figure shows that the balancing loop “B1” would be the functional response of the performance management/measurement system to the actual crime level. An increase of this level would imply a higher effort to repress crime, leading (after a delay) to a reduction of crime. However, the figure also shows how the use of an unbalanced set of performance measures and of short-term “output-based-only” reward mechanisms would hamper the dominance of the described balancing loop. In fact, the dysfunctional behavior that they would generate might foster two other loops, i.e. the reinforcing vicious loop “R” and the balancing loop “B2”.

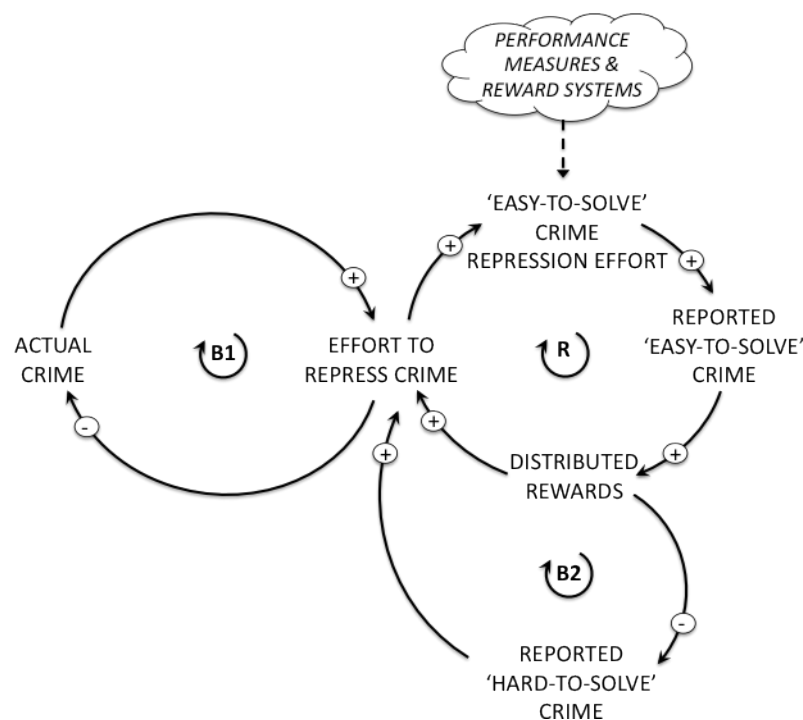


Figure 7: Distorted effects on reported crime produced by a rewards and a performance measurement system focused on only short-term output measures.

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 kind of crimes. This generates an increase in the distributed rewards, which in turn gives rise to the ‘search’ of new “easy-to-solve” crime to deal with and report. Though such a vicious loop would contribute to reduce crime (in absolute terms), it might not generate the desired effects on dealing with “hard-to-solve” crimes. On the other hand, the described system would also affect the statistics regarding such crimes (data fudging). In fact, the

distribution of rewards based on the described logics, would lead to a lower tendency by police to classify and report “hard-to-solve” crime and a higher inclination to ‘shift’ or declassify part of it to the lower rank of “easy-to-solve” crime. This would generate a further effort towards the repression of “easy-to-solve” crimes. In other words, while the loop “B2” dominance replaces the loop “B1”, it also provides the basis for the loop “R” to further and progressively generate its dysfunctional effects.

Figure 8 adopts an instrumental view of the DPM approach to detect the dysfunctionalities caused by lack of alignment between performance management and rewards systems. The figure shows how, 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 as a means to pursue sustainable organizational development.

As illustrated in fig. 8, the use of police staff and means (strategic resources) is primarily diverted by the rewards system to the execution of controls and preventive actions on ‘easy cases’. This effort contributes to generate the desired ‘numbers’, in terms of outputs, here depicted as flows of ‘easy’ cases solved.

The effect of this phenomenon is twofold.

On the one hand, it allows the organization to distribute rewards, based on the achievement of the target number of cases to solve. A consolidated habit of the organization to accumulate rewards – as a consequence of a large volume of fulfilled “easy” cases – reinforces the described process, which implies that organizational efforts are primarily allocated to such cases.

On the other hand, although an increase in the outflow of “easy” cases solved in a given time (end-result) should have to determine a reduction in the stock of “easy-to-solve” crimes, this does not happen, because of the inflow of new false ‘easy-to-solve’ reported cases. This variable has been depicted in fig. 8 as an output of the rewards system, whose size is proportional to the effort on reporting false data, which is – in turn – proportional to the targeted number of crimes to solve (information stock).

Therefore, an abnormal behavior of the rewards system here generates a set of end-results, which may allow the growth of the rewards system, regardless the achievement of the organizational results through the performance management system. On the contrary, the results measured by such system become a means to pursue the rewards system’s goals.

A further vicious effect can be referred to unattended hard crimes. In fact, this phenomenon could allow a developing class of hard-crime criminals who become skilled, leading to a gradual – but maybe substantial – reduction in the city’s quality of life in the long run.

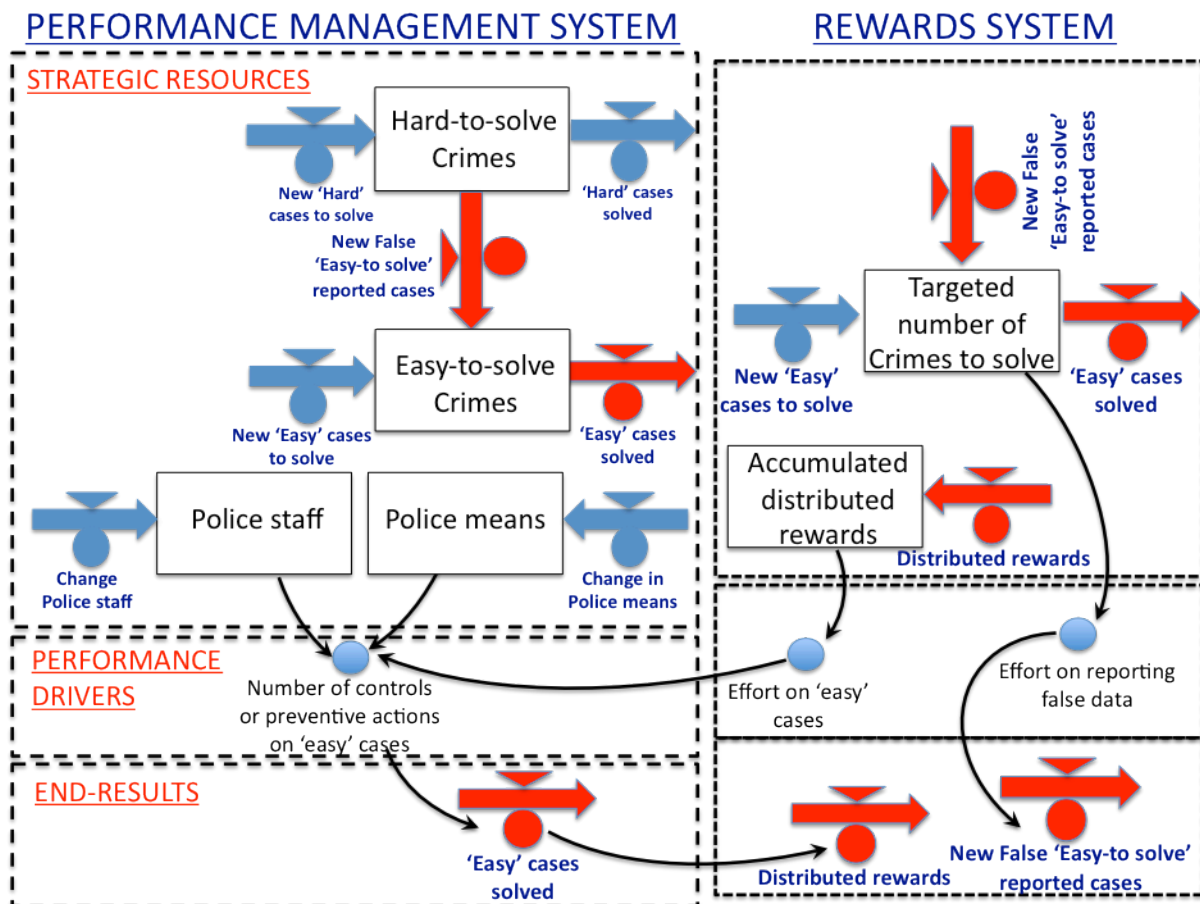


Figure 8: Idiosyncratic behavior of rewards and performance management systems generating “false reporting” on ‘easy-to-solve’ cases.

The insights emerging from the previous examples, about the role of DPM as a method to describe and detect the behavioral dysfunctions associated to the ‘performance paradox’, as in the case of the Compstat program, can support the identification of possible measures to adopt, in order to counteract the analyzed problems.

Figure 9 provides an example on this regard. It shows how improving performance management systems’ quality may lead to higher promptness and selectiveness of reporting (performance driver). This would not only imply investing on performance measurement and information systems, but also on organizational design and human capital. As remarked by Behn (2008, p. 5), one of the main errors committed in implementing Compstat, Citistat and similar programs has been the lack

of dedicated analytic staff to designing proper performance management systems, and to support decision makers in interpreting the results emerging from performance measurement.

For instance, as shown in fig. 9, improving the promptness and selectiveness of reporting may allow decision makers to design policies which better balance capacity allocation in solving ‘hard’ and ‘easy’ cases (end-results). Balancing the mix of achieved end-results (in terms of outputs) may contribute to better pursue a sustainable development of a public sector organization and of the territory where it is rooted. This is shown in fig. 9 through the outcome measure (end-result) named “change in community quality of life”. This variable is, in turn, affected by a performance driver, which is termed as “crime seriousness ratio”, which is affected by the stock of “hard-to-solve” crimes in respect to the total crimes.

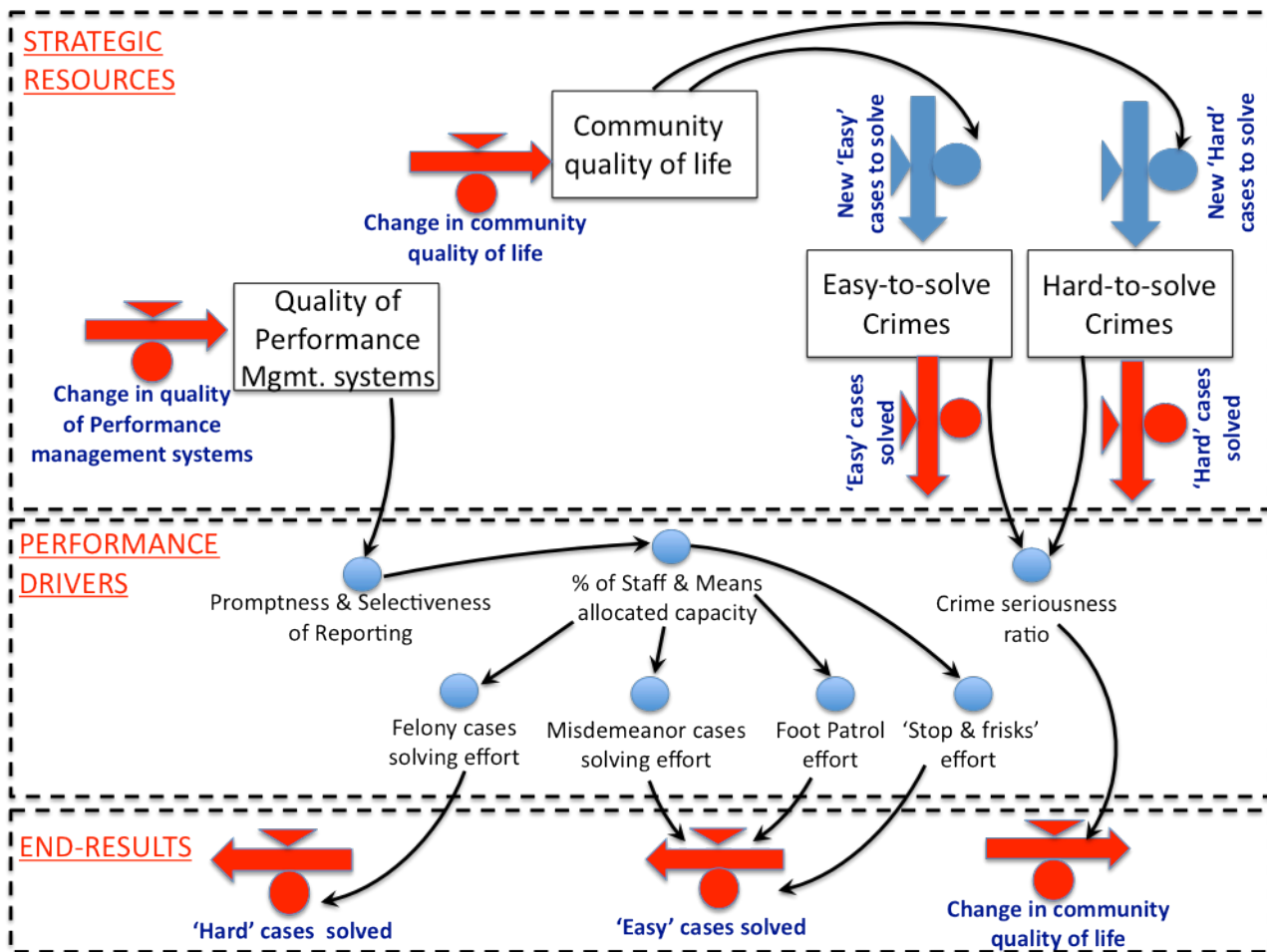


Figure 9: Using Performance Management Systems Quality (promptness & selectiveness) to support policies aimed to properly balance capacity allocation in solving ‘hard’ and ‘easy’ cases.

8. Conclusions.

This paper has shown how ‘dynamic performance management’ can be used as a method to support the analysis and diagnosis of the “performance paradox”, i.e. a weak capability of adopted

performance indicators to measure and affect performance, especially because of behavioral distortions in the use of available data and information.

As described in the paper, although such behavioral distortions are a product of the human nature, particularly when quantitative social indicators are used to affect social decision-making (as suggested by the “Campbell’s law”), they can be amplified by an improper design of performance management systems.

On the one hand, a possible cause of this phenomenon can be identified into the inconsistency between specific adopted performance indicators and the corresponding task environment. In this case, behavioral distortions in using performance measures tend to be rooted at individual – before than group or organizational – level, and are due to the use of performance indicators which do not fit with the characteristics of the activities that are fulfilled by people in an organization.

On the other hand, behavioral distortions in using performance measures can be also primarily related to a bounded and static perspective in the design of overall performance management systems. This phenomenon can imply extensive disconnections between: 1) outputs and outcomes, 2) the political and administrative level, 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 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 the related systems (e.g. rewards and career), which are the primary cause of a flow of organizational actions, which go towards a direction that is not compatible to a sustainable development, in the long run.

This second cause of behavioral distortions in performance measurement has been framed in the paper through a dynamic performance management view, with a specific reference to crime control policies at Municipal level, and particularly in relation to the implementation of 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 to city’s performance measurement programs, we also believe that further research will be needed to better frame, and possibly outline main archetypes related to different typologies of the recurring dysfunctions associated to the analyzed problem contexts.

References

- Argyris K. (1952). The Impact of Budgets on People, Ithaca, New York, *The Controllorship Foundation*.
- Behn R. (2001). The Psychological Barriers to Performance Management: or Why Isn't Everyone Jumping on the Performance-Management Bandwagon?, *Public Performance & Management Review*, 26, 5-25.
- Behn R. (2003). Why Measure Performance? Different Purposes Require Different Measures, *Public Administration Review*, 63, 5, 586-606.
- Behn R. (2008). The Seven Big Errors of PerformanceStat, February, <http://www.hks.harvard.edu/thebehnreport/Behn,%20PerformanceStatErrors.pdf>
- Behn R. (2011). Be Aware (and Beware) of the Campbell's Law, *The Behn Report*, 9, 12, August, <http://www.hks.harvard.edu/thebehnreport/All%20Issues/August2011.pdf>.
- Behn R. – Kant P. (1999). Strategies for Avoiding the Pitfalls of Performance Contracting, *Public Productivity and Management Review*, 22, 470-489.
- Bianchi C. (2010). Improving Performance and Fostering Accountability in the Public Sector through System Dynamics Modelling: From an 'External' to an 'Internal' Perspective, *Systems Research and Behavioral Science*, 27, 361-384.
- Bianchi C. (2012). Enhancing Performance Management and Sustainable Organizational Growth Through System-Dynamics Modelling, in: Grosser S. – Zeier R. (eds.), *Systemic Management for Intelligent Organizations*, Springer, Berlin.
- Bianchi C. – Bivona E. – Cognata A. – Ferrara P. – Landi T. – Ricci P. (2010). Applying System Dynamics to Foster Organizational Change, Accountability and Performance in the Public Sector: A Case-Based Italian Perspective, *Systems Research and Behavioral Science*, 27, 395-420.
- Bianchi C. – Rivenbark W. (2012-a). A Comparative Analysis of Performance Management Systems: The cases of Sicily and North Carolina, *Public Performance & Management Review*, 35, 3, 510-527.
- Bianchi C. – Rivenbark W. (2012-b). Using System Dynamics to Enhance Performance Management in Local Government: An Application to Residential Refuse Collection, *APPAM Fall Research Conference*, November 8-10, Baltimore (USA).
- Bianchi C. – Winch G.- Tomaselli S. (2008). Management Simulation as an Instrument to Aid Turning "Stunted Growth" Round in Family Businesses, *Sinergie*, 75, 109-126.
- Birnberg J. – Turopolec L. – Young S. (1983). The Organizational Context of Accounting, *Accounting, Organisations and Society*, 8, 2/3, 111-129.
- Boba Santos R. (2013). *Crime Analysis with Crime Mapping*, Sage Publications, London.
- Bohte J. – Meier K. (2000). Goal Displacement: Assessing the Motivation for Organizational Cheating, *Public Administration Review*, 60, 2.
- Bouckaert G. – Balk W. (1991). Public Productivity Measurement. Diseases and Cures, *Public Productivity & Management Review*, 15, 229-235.

- Bouckaert G. – Halligan J. (2008). *Managing Performance. International Comparisons*, Routledge, New York.
- Boyle R. (1999). *The Management of Cross-Cutting Issues: Committee for Public Management Research, Discussion paper n. 8.*
- Borgonovi E. (1996). *Principi e sistemi aziendali per le Amministrazioni Pubbliche (Management Principles and systems for Public Administrations)*. Egea, Milano.
- Brownell P. (1982). The Role of Accounting Data in Performance Evaluation, *Journal of Accounting Research*, 20, 1, 12-27.
- Campbell D. (1976). Reforms as Experiment, *American Psychologist*, 24, 4, 409-429.
- Christensen T. – Laegreid P. (2007). *Transcending New Public Management. The Transformation of Public Sector Reforms*, Ashgate, Surrey.
- Cyert R. – March J. (1963). *A Behavioral Theory of the Firm*, Prentice Hall, New Jersey.
- Daividsen, P. (1991): *The Structure-Behavior Graph*. The System Dynamics Group, MIT. Cambridge.
- De Lancer Julnes P. (2001). Promoting the Utilization of Performance Measures in Public Organizations: An Empirical Study of Factors Affecting Adoption and Implementation, 61, 6, 693-708.
- De Lancer Julnes P. (2006). Performance Management. An Effective Tool for Government Accountability? The Debate Goes On, *Evaluation*, 12, 2, 219-235.
- De Lancer Julnes P. (2008). Performance Management. Beyond Instrumental Use, in: Van Dooren W. – Van De Walle, *Performance Information in the Public Sector. How it is Used*, Palgrave MacMillan, Houndmills, UK.
- Dermer J. – Lucas R. (1986). The Illusion of Managerial Control. *Accounting, Organisations and Society*, 11, 6, 471-482.
- Eterno J. – Silverman E. (2010). The NYPD's Compstat: Compare Statistics or Compose Statistics?, *International Journal of Police Science & Management*, 12, 3.
- Eterno J. – Silverman E. (2012). *The Crime Numbers Game. Management by Manipulation*, CRP Press, Taylor & Francis, New York.
- Flamholtz E. (1996). Effective Organizational Control: A Framework, Applications, and Implications, *European Management Journal*, 14, 6, 596-611.
- Forrester J. (1961). *Industrial Dynamics*, MIT Press, Boston, Mass.
- Johnson B. (2005). Strategies for Successful Joined up Government Initiatives, *John Curtin Institute of Public Policy: Institute of Public Administration of Australia*, Department of Premier and Cabinet.
- Hirst M. (1983). Reliance on Accounting Performance Measures. Task Uncertainty, and Dysfunctional Behavior: Some Extensions, *Journal of Accounting Research*, 21, 2, 596-605.
- Hofstede G. (1967). *The Game of Budget Control*, Tavistock, Asen.

- Hofstede G. (1978). The Poverty of Management Control Philosophy, *The Academy of Management Review*, 3, 3, 450-461.
- Hofstede G. (1981). Management Control of Public and not-for-Profit Activities, *Accounting, Organizations, and Society*, 6, 3, 193-211.
- Hopwood A. (1972). An Empirical Study of the Role of Accounting Data in Performance Evaluation, *Journal of Accounting Research*, 10, 156-182.
- Leeuw F. (2003), Evaluation and New Public Management in the Netherlands, in: Wollmann H. *Evaluation in Public Sector Reform. Concepts and practice in International Perspective*, Elgar Publ., Bodmin, Cornwall.
- Maple J. (1999). *The Crime Fighter*, Doubleday, New York.
- March J. (1987). Ambiguity and Accounting: the Elusive Link between Information and Decision Making, *Accounting, Organizations & Society*, 12, 2, 153-168.
- March J. – Simon H. (1958). *Organizations*, Wiley, New York.
- Matheson A. – Scanlan G. – Tanner R. (1997). Strategic Management in Government: Extending the Reform Model in New Zealand, *OECD Report on “Benchmarking, Evaluation and Strategic Management in the Public Sector*, <http://www.oecd.org/newzealand/1902913.pdf>
- Meyer M. – O’Shaughnessy K. (1993), Organizational design and the performance paradox, in: Swedberg R. (Ed.), *Explorations in Economic Sociology* (pp. 249-278). Russell Sage Foundation, New York.
- Meyer, M. W., & Gupta, V. (1994). The Performance Paradox, *Research in Organizational Behavior*, 16, 309-369.
- Merchant K. (1982). The Control Function of Management, *Sloan Management Review*, Summer, 43-55.
- Mintzberg H. (1975). Impediments to the Use of Management Information. Washington, D.C.: *National Association of Accountants*.
- Mintzberg H. (1979). *The Structuring of Organizations*, Prentice Hall, Englewood Cliffs.
- Moynihan D. (2008). *The Dynamics of Performance Management. Constructing Information and Reform*, Georgetown University Press, Washington DC.
- Moynihan D. – Landuyt N. (2009). How Do Public Organizations Learn? Bridging Cultural and Structural Perspectives, *Public Administration Review*, Nov.-Dec.
- Murray M. (2005). Why Arrests Quotas are Wrong, *PBA Magazine*, Spring, <http://www.nycpba.org/publications/mag-05-spring/murray.html>

- Ouchi W. (1979). A Conceptual Framework for the Design of Organizational Control Mechanisms, *Management Science*, 25, 9, 833-848.
- Otley D. (1978). Budget Use and Managerial Performance, *Journal of Accounting Research*, 16, 1, 122-149.
- Perez T. – Rushing R. (2007). The Citistat Model. How Data-Driven Government Can Increase Efficiency & Effectiveness, Center for American Progress, April, http://www.americanprogress.org/wp-content/uploads/issues/2007/04/pdf/citistat_report.pdf
- Pollit C. (2003). Joined-up Government: a Survey, *Political Studies Review*, 1, 1.
- Radnor Z. (2005). Developing a Typology of Organizational Gaming, EGPA Conference Proceedings, Bern http://www.academia.edu/882699/Developing_a_typology_of_organisational_gaming
- Richardson G. – Pugh A. (1981). System Dynamics Modeling with Dynamo, Productivity Press, Portland, Oregon.
- Rouwette E. – Jongebreur W. – van Hoff P. – Heijmen T. – Vennix J. (2007). Modeling Crime in the Netherlands, *International System Dynamics Proceedings*, Boston. <http://www.systemdynamics.org/conferences/2007/proceed/papers/ROUWE343.pdf>.
- Sanger M. (2008). Getting the Roots of Change. Performance Management and Organizational Culture, *Public performance and Management Review*, 31, 4, 620-652.
- Sanger M. (2012). Does Measuring Performance Lead to Better Performance? *Journal of Policy Analysis and Management*, Volume 32, Issue 1, pages 185–203, Winter 2013
- Silverman E. (2001). NYPD Battles Crime: Innovative Strategies in Policing, Northeastern University Press, Boston, MA.
- Simon H. (1955). A Behavioral Model of Rational Choice, *Quarterly Journal of Economics*, 69, 99-118.
- Simons R. (2005). Levers of Organization Design: How Managers Use Accountability Systems for Greater Performance and Commitment, Harvard Business School Press, Boston.
- Smith M. (2013) Records: Texas plant hadn't told feds about explosive fertilizer. Retrieved June 11, 2013, from www.cnn.com/2013/04/25/us/texas-explosion-plant
- Smith P. (1993). Outcome-related Performance Indicators and Organizational Control in the Public Sector, *British Journal of Management*, 4, 135-151.
- Smith P. (1995). On the Unintended Consequences of Publishing Performance Data in the Public Sector, *International Journal of Public Administration*, 18, 2&3, 277-310.
- Sterman J. (2000). Business Dynamics. Systems Thinking and Modeling for a Complex World, Irwin McGraw Hill, Boston.
- Vakkuri J. – Meklin P. (2006). Ambiguity in Performance Measurement: a Theoretical Approach to Organizational Performance Measurement, *Financial Accountability & Management*, August.

Van Dooren W. – Bouckaert G. – Halligan J. (2010). *Performance Management in the Public Sector*, Routledge, New York.

Van Thiel, S. – Leeuw F. (2002). The Performance Paradox in the Public Sector, *Public Performance & Management Review*, 25, 3, 267-281.

Wilson J - Kelling G. (1982). Broken windows: The Police and Neighborhood Safety. *Atlantic Monthly* March, 29-38.

Zink R. (2004). The Trouble with Compstat, *The PBA Magazine*, <http://www.nycpba.org/publications/mag-04-summer/compstat.html>