

When will we ever learn?

Vicki Johansson

Abstract

This paper puts forward the argument that Performance Measurement Systems (PMSs) foster rational, self-interested behaviour and vested values at all levels within organisations, which weakens moral barriers preventing fraud, fabrication of data and bribery. It argues that the longer a PMS is in operation, the greater the probability that rational self-interested behaviour in conflict with fundamental values and goals will be consolidated, aggravated and disseminated within organisations that operate within public welfare policy. If implemented, common incentives aimed at counteracting undesirable behaviour aggravate and speed up this process rather than reversing it. In a worst-case scenario, PMSs are the first step toward corruption, even though PMSs have been implemented with the good intention of improving public policy and strengthening accountability.

Key words:

performance measurement systems, implementation, rational behaviour, intrinsic values

Introduction

A huge and unequivocal amount of research clearly shows that performance measurement systems (PMSs)¹ generate behaviour among actors which counteracts the intention behind the systems: for example, cheating, gaming, tunnel vision, myopia and cherry-picking (e.g. Smith 1995; de Bruijn 2002; Radin 2006; Mannion and Braithwaite 2012). The types of behaviour that promote negative effects are often referred to as dysfunctional, perverse, unethical or unintended (e.g. Ridgway 1956; Perrin 1998; Van Thiel and Leeuw 2002; Pidd 2005; Ordóñez et al. 2009; Fryer et al. 2009), even though they would in fact be regarded as both rational and anticipated if organisational and decision-making theory were applied.

The implementation of reforms inspired by New Public Management is often described in the literature as the main cause behind an increased use of various PMSs (e.g. Pidd 2005; Fryer et al. 2009; Lewis and Triantafillou 2012). The arguments and justifications that preceded the implementation of these reforms (which should today be termed “Normal” Public Administration rather than “New”) were that increased efficiency would be achieved in public policy if management methods and control techniques from the private sector were imported into and applied within the public sector, and if private organisations and operators were involved and engaged in the production of publicly funded services (Osborne and Gaebler 1992; Bouckaert and Halligan 2008). A distinctive feature of NPM-inspired reforms has been that they focus on the output-side of a delivery system and consequently on techniques to steer and control outputs and outcomes (Pierre and Peters 2000). However, behaviour that undermines the intention behind PMSs is not linked to NPM reforms *per se*. In fact, these types of behaviour and their effects are similar, whether PMS research focuses on liberal democracies, authoritarian states (Bevan and Hood 2006; Chan and Gao 2008) or private or public firms (e.g. Berliner 1956; Ridgway 1956; Merchant 1990; Ordóñez et al. 2009).

Undesired behaviour is often presented in taxonomies in the literature, sometimes with different terms that encompass the same behaviour (recently, e.g. Fryer et al. 2009; Mannion and Braithwaite 2012; Hamilton et al. 2013); explanations for the behaviour are discussed and measures are suggested with the aim of counter-acting types of behaviour that are undesirable and inefficient for the model (e.g. de

1 In the literature several comprehensive terms are used in order to cover different characteristics, actors and functions related to Performance Measurement, as for example Performance Regimes, Performance Management Systems, Performance Indicator Systems or Performance Measurement systems. Performance measurement involves, for example: development and assessment of objectives and indicators, actors and behaviours within organisations (e.g. at the management and operational levels), actors and behaviours “outside” a specific organisation (e.g. authorities, competing organisations, media, the public) and so on. Thus, performance measurement activities never take place in a vacuum: characteristics, actors and functions are interwoven and interdependent. Hence, Performance Measurement Systems (PMS) will be used as an umbrella term consistently throughout the text.

Bruijn 2002; Van Thiel and Leeuw 2002; Bevan and Hood 2006). However, PMSs are very seldom challenged or questioned other than superficially. It is rare for the model and its basic, intrinsic, axiomatic assumptions to be questioned; it is their application and how people within PMSs act (and how they could be persuaded to act differently) that are put forward. The merit of PMSs seems to be more or less unquestioned within the research community. It is permissible to conduct critical examinations of PMSs but not to fundamentally question their necessity and utility. For instance, after a critical examination of performance regimes, Moynihan et al. draw the following conclusion:

Thus, our argument here should not be constructed as an attack against performance regimes – measuring performance is almost always preferable to not measuring performance – but a recognition of the complex context in which they operate (2011, i153).

And after a very well-founded and interesting examination of the (alternative) logics of performance management systems, Pollitt draws this conclusion:

We could think of [undesired behaviours] as perennial weeds in the garden – elements in the ecosystem that need to be constantly managed rather than intruders that can somehow magically be eliminated (2013, 360).

One can wonder about the wisdom of vindicating and justifying PMSs when such a huge amount of research unequivocally shows that behaviour undermining the intentions behind PMSs is the rule, not the exception; arguing in favour of PMSs when the cumulative research unquestionably implies that these behaviours are systematic; that PMSs generate highly undesirable effects and consequently outcomes that are indeed the opposite of those desired (e.g. Smith 1995; de Bruijn 2002; Mannion and Braithwaite 2012). The most severe organisational and societal effects highlighted and questioned in the literature include outcomes such as goal displacement, inefficiency and misuse of resources, erosion of democratic values and growing distrust (e.g. Power 1997; Moynihan et al. 2011, Johansson and Lindgren 2013, Johansson and Montin 2014).

Accordingly, this article puts forward the argument that Performance Measurement Systems (PMSs) foster rational, self-interested behaviour at all levels within organisations, which weakens moral barriers preventing fraud, fabrication of data and bribery and thereby creating breeding grounds for corruption. It argues that the longer a PMS is in operation, the greater the probability that rational self-interested behaviour in conflict with fundamental values and goals will be consolidated, aggravated and disseminated within organisations that operate within public welfare policy. If implemented, common incentives aimed at counteracting undesirable behaviour aggravate and speed up this process rather than reversing it. In a worst-case scenario, PMSs are the first step toward corruption, even though PMSs

have been implemented with the good intention of improving public policy and strengthening accountability.

Underpinning and development of the argument

The main purpose of applying PMSs is to improve goal achievement by shaping internal behaviour within organisations and acting as the basis for internal (bureaucratic) and external (democratic or market) accountability (Poulsen 2009; Nutley et al. 2012). PMSs are thus incentives decided by and implemented by the leadership and management of organisations in order to promote outcomes that correspond with the overall values inherent in the goals maintained by the principals of organisations.

PMSs have been used for centuries as management and control tools by both governmental authorities and private companies. Despite the considerable timespan, the taxonomies encompassing behaviours triggered by and within PMSs are approximately the same today as those of the past, even though a number of behaviours/effects have been added recently. The bulk of the behaviours/effects listed in contemporary taxonomies were identified in investigations conducted by researchers in the first half of the 20th century who mainly, but not only, studied PMSs within private and state-owned enterprises in liberal democracies and the Soviet Union (e.g. Berliner 1956; Ridgway 1956; see also Bevan and Hood 2006). Although the enterprises they studied were productive, many of the behaviours they identified are the same as those that contemporary researchers have found in private and public organisations which deliver services. At first glance, this similarity might seem somewhat surprising; however, major dissimilarities would actually be much more astonishing. It is, of course, true that private and public organisations operate under different conditions; while private enterprises that produce goods act on a market, public organisations that deliver services act within a political-administrative system. However, the purpose of PMSs is ultimately the same: to improve effectiveness by shaping internal behaviour within organisations.

PMSs are applied with the ultimate purpose of achieving values inherent in goals maintained by the principals of organisations, not an end in themselves. They are expected to foster behavioural changes among employees so that organisations' efficiency and outcomes are enhanced, whether it be goods or services that are being produced. For corporations active in a market and organisations delivering services in publicly funded free choice systems inspired by NPM reforms, organisations' outputs are also intended to serve as a basis for, respectively, the purchase of goods and the selection of publicly funded services. Consequently, PMSs very often have internal as well as external functions, regardless of the type of organisation.

PMSs are implemented in a real-life context, not in a clean and sterile experimental laboratory free of germs that could interfere with the process. The implica-

tion of this is that the political-administrative context of societies and organisations does in fact affect the intended functions. Inherent and theoretically derived logics within PMSs will be disturbed and altered in relation to the divergent logics in these real life contexts. In the social sciences it is not possible to carry out experimental studies. The empirical data from which we can draw conclusions is consequently self-generated in contemporary societies and organisations. In order to be able to conduct a theoretically informed discussion about the evolutionary process of PMSs, it is essential to comprehend how contexts are composed and how they influence empirically observed behaviour among those involved in applying and maintaining PMSs, as well as behaviour among those who are subject to the control system.

Theoretical and empirical studies of PMSs within different political administrative systems have increased exponentially since the 1980s. It was around this time that Western liberal democracies started to implement PMSs as a result of NPM reforms (e.g. Power 1997; Van Dooren et al. 2010) and authoritarian states started to implement so-called target-based responsibility systems in order to come to terms with economic development and societal problems, such as social instability (Chan and Gao 2008). These contexts – market/quasi-market (NPM), planned economy, liberal democracies and authoritarian states – diverge empirically in essential respects. There is thus no lack of empirical research concerning conditions that have proved to infect the intended functions of PMSs. Many and various comparisons and analyses of intended functions, effects and contextual factors of importance have been made (e.g. Arthur 1994; Baker 2002; Pidd 2005; Bevan and Hood 2006; Moynihan et al. 2011; Lewis and Triantafyllou 2012; Hamilton et al. 2013; Pollitt 2013 and so on).

The argument put forward in this article is therefore based on an analysis of secondary data gathered from research presentations in articles and books where functions, behaviours, effects and contexts of performance measurement systems have been indirectly or directly addressed. Findings emanating from research conducted about PMSs in private and public organisations in liberal democracies and authoritarian states will be used as a basis for the identification of general and specific evolutionary pathways for PMSs within real organisations. The argument will be underpinned and developed in four interrelated steps.

First, the fundamental theoretical premises implicit in the basic performance measurement model will be outlined, and the rational decision-making theory constituted by these premises will be discussed in relation to empirically observed features of real organisations within which PMSs have actually been implemented.

Second, the effect of applying internal and external performance measurement incentives, i.e. rewards and penalties designed to strengthen the efficiency of PMSs, are discussed in relation to four characteristics of public policy that have been empirically proven to infect the causal relationship that is assumed between incentives

and responses, namely service production, multiplexity, eligibility, decision-making principles and intrinsic values.

Third, mechanisms will be identified and outlined that reinforce self-interest and replace altruistic behaviour and values at the individual, unit and organisational levels due to the application of internal and external incentives.

Finally, reasons that lie behind why it is so difficult to prevent and stop undesired processes being set in motion in real-life contexts by PMSs, despite our knowledge of anticipated effects, are discussed, or, in other words, the question is addressed of why we are unable to learn from our knowledge.

The implicit theoretical premises in the basic performance measurement model

The main aim of a PMS is often two-fold, i.e. to promote the achievement of goals through enhanced efficiency and/or opportunities for accountability. However, at an organisational level the purpose of applying performance measurement models is slightly different: to improve goal achievement by shaping internal behaviour within organisations.

PM models usually contain five comprehensive premises that are logically linked to each other (see, for example, Smith 1995):

1. The organisation's goals and desired outcomes are clear, uniform in terms of value and consistent.
2. Objectives/targets are operationalised in performance indicators.
3. Data on performance indicators is collected.
4. The results are compiled and compared with objectives and desired outcomes.
5. Measures are taken in order to rectify deviant results and improve outcomes.

At the core of these premises is the conception of rational decision-making. Theoretically, rational decision-making requires goals to be uniform in terms of value, unambiguous and possible to rank, that casual relationships and means are known, and that decision-makers have all the information necessary to make decisions (Simon 1993). These preconditions are very rarely met outside experimental laboratories (Zsombok 1997; Carroll et al. 2006; Klein 2008). Human beings tend to act rationally but base their actions on a range of ambiguous goals and incomplete information; they act in relation to contexts and other actors, and they can be rational in different aspects; they can base their actions on vested and/or altruistic values as well as on individualistic, organisational, sub-organisational or collective interests (Jones 2003; Simon 2000; Lipshitz et al. 2001; March and Olsen 2004).

However, all five premises take for granted that everyone involved, whether they be, for instance, managers, specialists, technicians or street-level bureaucrats, will act in accordance with the logic that informs the model; that for all of them rationality is equivalent to the behavioural assumptions inherent in the model. In reality, employees have different tasks, functions and positions that influence their goals, interests and behaviours. Goals maintained by an organisation's principal are not necessarily shared by its members, nor is the principal's perception of which goals it is appropriate to achieve in situations where goals are shared. Consequently, when the model is applied in real organisations, the premises behind the model turn out to be invalid.

First, a common feature is that the management and implementation of PM models is carried out by and within bureaucracies. Bureaucracies are not democracies, although the application of PMSs in democratic societies can be made with the intention of strengthening democracy by improving political effectiveness (goal achievement) and legitimacy (rule of law, due process) and/or improving and strengthening not only opportunities for internal bureaucratic accountability but also external democratic accountability (de Bruijn 2002; Poulsen 2009; Chan and Gao 2009; Lewis and Triantafillou 2012). However, bureaucracies are not designed to be democracies. Bureaucracies in the private sector are supposed to implement the decisions of company owners and managers, while public bureaucracies implement decisions made by democratically elected bodies. Members of those bureaucracies have no statutory right to make decisions regarding the organisation's goals, rules or activities; in this respect there are no differences between public and private enterprises.² If a person is employed by a bureaucracy, that person is employed to execute the decisions made by the leadership of that organisation. In theory, an employee is free to leave a position if she/he disagrees with the goal of an organisation, how the goal is interpreted or how it is implemented. In the real world most people do not have that option; we all live in a wage labour society and have to support ourselves, so we cannot leave if no alternative livelihood is available. Accordingly, for both private and public employees who find themselves in such situations, strategies to mitigate and counteract management and control systems are both essential and rational. If they have any sort of discretion, employees within and managed by bureaucracies will not function robotically regardless of return, whether represented by salary, self-esteem or mental/physical health. As an abundance of research findings illustrate, they will defend themselves. They will ultimately develop behavioural strategies such as those described by, for example, street-level bureaucracy researchers, performance measurement scholars and economists who have focused on undesired and so-called dysfunctional or perverse effects developed in response to management and control techniques in private and public organisations (e.g. Lipsky 1980; Mannion and Braithwaite 2012).

2 Even though public employees who also have voting rights in general elections can vote for the party/person they want to elect as governor for a delimited time.

Second, an organisation's goals are very seldom unequivocal; goals within welfare-state policy in particular are multidimensional and interpretable. The decision-makers, i.e. the elected representatives, usually have divergent opinions about both preferable outcomes and effective means to achieve them. Further, the means to reach preferable outcomes are often not known or highly dependent on context (e.g. Van Thiel and Leeuw 2002; Ordóñez et al. 2009; Maynard 2012; Frey et al. 2013). These divergent conceptions among elected representatives are usually also present among employees; for example managers, technicians, street-level bureaucrats. Not only deriving primarily from individual political preferences, but from educational background, job assignments, work situation and position and level within an organisation.

Third, vague, unclear and contested goals along with unknown or complicated causal links are not an optimal starting point for operationalising goals into targets and Performance Indicators (PI) (Moynihan et al. 2011). As a result, PIs might turn out to be too narrow, as only that which can be agreed on is covered by the indicators (Van Thiel and Leeuw 2002; Ordóñez et al. 2009), or too wide, when different interpretations are encompassed in order to reach a consensus or manage uncertainty about causal conjunctions (Pollitt et al. 2010). Too few indicators have the tendency to foster goal displacement through knowledge about how judgments are made (Pidd 2005; Frey et al. 2013), while too many indicators tend to be too challenging and therefore tend to make the model inefficient and undermine ambitions (de Bruijn 2002; Frey et al. 2013).

Fourth, outcomes and objectives are usually formulated in overall qualitative terms, while in practice PIs optimally deal with outputs (Baker 2002; Smith 1995; Gao 2010). One of the most common and reiterated conclusions in PM research is that PIs usually focus on what can be measured in practice, what can be quantified given the bureaucratic and technical systems available (e.g. Perrin 1998; Wilson et al. 2006), which implies that indicators used for goal fulfilment often are somewhat superficial and occasionally peripheral.

Fifth, results emanating from indicator systems are usually compared with the indicators themselves, not with the goals they are supposedly implementing, with means transformed into objectives in the process. This has been encapsulated in what have become classic quotes from PM research: "What gets measured, gets done", "teach to the test" (Behn 2003; Phelps 2011). Measures taken in order to come to terms with undesired results consequently tend to focus on why organisations or units within them have not met the indicators and what measures ought to be taken in order to improve the organisation's or unit's indicator scores, i.e. the focus is on PI instead of outcomes, which often results in an escalation of performance measurement *per se* and performance measurement systems in particular (Pollitt et al. 2010; Pollitt 2013).

To sum up, the main reasons behind undesired behaviours and, by extension, undesired effects, are that the basic performance measurement model assumes that everyone who participates shares and defines the main goal in the same manner and that everyone also agrees on (and knows) how input through process generates outputs that together frame outcome. Furthermore, the objectives and targets of organisations in liberal and authoritarian states are imposed upon their members, who have no statutory right or profound or real possibility of defining or changing the organisation's basic purpose or content, even though some opportunities might exist to express opinions and to influence objectives and targets.

Performance measurement incentives and characteristics of public policy

The basic design of PMSs takes for granted that everyone involved, whether managers or employees, endeavours to achieve the same goals, interprets those goals consistently, shares the same perception of the road toward goal fulfilment and collectively ranks values and goals cherished by the principal of the organisation higher than all other possible individual and organisational goals that she/he might have. Accordingly, the basic design of PMS models does not include any specific internal and/or external performance measurement incentives, even though the need for some sort of feedback loop is often emphasised in the literature (e.g. Smith 1995; Fryer et al. 2009). However, when PMS models are applied in real organisations, whether these organisations are situated in liberal democracies or authoritarian states, and regardless of whether they are public, quasi-private or private, or whether they are operating in the early 1900s or 2000s, some sort of internal and/or external incentives are usually incorporated (Berliner 1956; Arthur 1994; Smith 1995; Van Thiel and Leeuw 2002; Maynard 2012; Chan and Gao 2009 and so on).

Internal and external performance measurement incentives are used in order to alter the basis for action of participants in PMSs, so whatever complementing or competing objectives they might have are suppressed to the benefit of organisational objectives. The general aim is to influence and amend participants' individual or collective behaviours so that the internal logic of PMSs is reinforced.

There are two main types of incentives: rewards and penalties, i.e. sanctions. Rewards and penalties can be of different kinds – financial, legal and/or status/reputational – and they can be applied and linked to individuals, units within organisations or organisations as a whole (e.g. Van Thiel and Leeuw 2002; de Bruijn 2002; Gao 2010; Lingard and Sellar 2013). Internal as well as external performance measurement incentives are used in order to change behavioural patterns within organisations; however, whereas internal incentives are applied from within, external incentives are applied from without.

Extrinsic incentives negatively strengthen and create gaming and cheating behaviours in all kinds of organisations (e.g. Berliner 1956; de Bruijn 2002; Pollitt et al. 2010; Gao 2010; Moynihan et al. 2011; Maynard 2012; Lingard and Sellar 2013; Frey et al. 2013), but they seem to provoke behaviours that generate more severe outcomes in publicly funded organisations than in private goods-producing organisations (e.g. Van Thiel and Leeuw 2002; Fryer et al. 2009; Frey et al. 2013). The question is why? Why would the effects be more severe in the public sector? And is it really the case? Which are the most important characteristics that are put forward as causal explanations for temporal effects of internal and external performance measurement incentives?

Service production: The first characteristic is related to what an organisation is supposed to produce and deliver; goods vs. social services. Incentives have been shown to increase output when applied in organisations that produce goods whether the organisation is privately or publicly owned, in both liberal democracies and authoritarian states (Gao 2009, etc.). Many products have what is sometimes referred to as a single value: in organisations where the aim is to produce a certain kind of product, let's say cars or rubbish collection, targets are comparatively easy to define and set, and they are easy to quantify (e.g. de Bruijn 2002; Pidd 2005; Frey et al. 2013). In organisations where the aim is to produce welfare services, such as care for the elderly, education or social stability, the link between goals and targets is much more complicated, as they are usually multi-valued (Smith 1995; de Bruijn 2002; Moynihan et al. 2011). Further, the output, and the quality of that output, is often a synthesis resulting from co-production with other organisations and interaction with recipients of the service (e.g. Smith 1995; Moynihan et al. 2011). Multi-valued goals, or alternatively multi-dimensional goals, are very difficult to operationalise into measurable targets. The difficulties stem from three forms of multiplexity.

Multiplexity: Typically, participants in PMSs operating within organisations which provide welfare have divergent opinions on how the organisation's goals should be realised, for example how quality should be defined (Baker 2002). Divergent comprehension of policy goals among participants in PMSs can often be explained by their formal function and role within public organisations (e.g. Otley 1999; Evans and Harris 2004; May and Winter 2007; Noordegraaf 2007). Formally elected officials represent the people, and their function is to govern the administration. Managers guide the implementation of public policy, and their function is to apply management tools that govern resource allocation and behavioural patterns within the organisation, and finally, professionals implement policy, and their function is to use their professional knowledge in decision-making. Accordingly, due to their formal functions and roles, participants within PMSs will embrace various ideas about how goals ought to be formulated and put into practice, and how targets and performance indicators are preferably set.

Further, welfare goals are always evolving and, especially if they are embedded with quality, the target is continuously in motion. When certain aspects of quality are achieved, new ones can be added and old ones reconsidered due to changes in society and a higher level of expectations. It is not an easy task to achieve a goal that is evolving, in constant movement and comprised of different definitions. Despite the fact that there will always be divergent opinions about whether goals have been accomplished or not, knowledge about the road towards the goal is very often fragmentary, incomplete or even absent altogether (e.g. Frey et al. 2013), which makes identifying reasonable targets and performance indicators challenging. Knowledge of causal relationships within social sciences is highly contextual and therefore temporary.

Contexts shape and impact goals, behaviours, attitudes and values that members maintain within a specific context, for example within welfare states, risk regimes, political systems or educational systems. Consequently, the context that PMSs constitute for the participants within it also shapes values, rationalities and behaviours. The ability of PMSs to initiate behavioural change and alter organisational and societal relationships increases when targets and indicators are known, since knowledge about causal relationships affects human behaviour. More than one expert on performance measurement has quoted Goodhart and his law from 1975, which states that when people become aware of causal relationships, they change their behaviour and the relationship collapses (e.g. Pidd 2005; Bevan and Hood 2006). This dilemma is also the main theme in *Foundation*, a series of science-fiction stories by Asimov published in 1942–1950, where the First Foundation's knowledge of the Second Foundation is a constant threat to the Second Foundation's ability to shorten the timespan of destruction in the Galactic Empire. The Second Foundation uses probability models to forecast behavioural patterns and implements actions to ensure that the most favourable development occurs. However, the instant the First Foundation realises that the Second Foundation exists, the probability of a long period of destruction increases since it influences the behaviour of the inhabitants of the First Foundation. The recurring problem in the Foundation series is actually how to deal with Goodhart's law.

Decision-making principle for eligibility: The characteristics of the public welfare sector and the recipients of its services deviate from those presupposed in a market. First the distribution of the welfare state's core services is mainly based on eligibility conditions and needs (education, healthcare, child care, elderly care). Second the quality and content of publicly funded services determined by elected officials in national, regional and/or local parliaments is often statutory. For these reasons, the mechanisms supposed to balance supply and demand in a market – purchasing power and preferences – do not work. Welfare in modern welfare states is a political responsibility. Regulations, funding and control are almost always in the hands of the elected (indirectly the electorate) and the political-administrative system, regardless of whether private or public organisations produce and deliver

welfare services. Further, political-administrative systems are often multileveled in different respects and to different degrees. There are unitary states, federal states, states with a high degree of local self-government and so on. In other words, in modern welfare states there is no principal who can interpret and determine the content of welfare goals on her or his own authority. Interpretation and content of goals, as well as applications of means to achieve goals, derive from discussions, negotiations and compromises between actors within a political administrative system, for example political parties, management within organisations, employers organisations, trade unions, professionals, interest organisations, consultants, media etc. (Fryer et al. 2009; Johansson 2012, Johansson 2015). Welfare goals are hard to define as such, and modern democratic decision-making processes that are inclusive and enable the participation of a large number of actors make it even harder to reach a unanimous and unambiguous interpretation of goals, targets, indicators and means to achieve goals, and the question is whether full consensus is indeed desirable from a democratic as well as a development perspective.

Intrinsic values: Research findings indicate that people who are involved in organisations that implement public policy within core welfare state services such as education, social services, elderly care and healthcare have a greater commitment to the main goal of the organisation (Baker 2002; Maynard 2012; Frey et al. 2013). Views are divided on whether this commitment is due to personal characteristics or is a result of socialisation processes or a combination of both, but it seems to be clear that the commitment is greater among employees within welfare state bodies (e.g. Bevan and Hood 2006; Dur and Zoutenbier 2014). Public organisations that deliver welfare seem to attract employees (or socialise employees) who are governed by intrinsic rather than extrinsic values to a higher degree than employees in private organisations (Arthur 1994; Baker 2002; Bevan and Hood 2006; Frey et al. 2013).

To sum up, welfare service production, multiplexity, decision-making principles for eligibility and intrinsic values are four characteristics that have the potential to infect presumed causal relationships between incentives and anticipated responses within PMSs.

Mechanisms set in motion by incentives at the level of individual, unit and organisation

So, what processes are set in motion when incentives are applied in organisations characterised by production of welfare services, multiplexity, decision-making principles for eligibility and intrinsic values? When are incentives applied in organisations that deal with multidimensional, value-laden, negotiated and complex goals? When is a service a right that does not depend on purchasing power but on need (i.e. healthcare, elderly care) or obligation (compulsory education)? What happens

in these kinds of organisations when desired outputs are rewarded and undesired outputs are punished?

If there were no incentives to change behaviour, people would probably continue to act as they have always done (Schneider and Ingram 1990; Salomon 2002), whether behaviour can be understood as derived from the logic of consequences or the logic of appropriateness (March and Olsen 2004). Behavioural change is triggered by incentives at the individual, organisational or societal level, and they can be intrinsic and/or extrinsic in relation to the actor (e.g. Arthur 1994; Baker 2002; Frey et al. 2013).

Incentives are applied in order to steer behaviour so that the probability decreases of employees allowing themselves to be governed by rationalities other than those inherent in, for example, PMSs (Argyris 1990). The application of incentives is based on the assumption that human beings act rationally and according to self-interest. People who are subjected to financial or legal incentives are thus presumed to change their behaviour if such a change delivers higher economic dividends or entails the risk of legal sanctions. Further, participants in PMSs who are exposed to financial or legal incentives are expected to change their attention towards performance measurement indicators, scores or targets defined within PMSs that give the individual, the unit or the organisation of which they are a member something positive in return, or to avoid something negative. However, it has long been known that direct and indirect financial incentives produce rather than prevent gaming behaviour in all kinds of organisations where such incentives are implemented (Van Thiel and Leeuw 2002; Ordóñez et al. 2009; Maynard 2012). So, what mechanisms are actually instigated within organisations at the individual, unit and organisational levels when incentives are applied or strengthened?

Individual level

Financial incentives presuppose that the ultimate goal of individuals and organisations is to secure or increase individual or collective financial resources, that human beings are maximisers of economic utility. However, professionals implementing public welfare policy are also motivated, or indeed motivated more, by incentives other than economic benefit. They are motivated to a higher degree than other employees by intrinsic rather than extrinsic values, i.e. these kinds of employees feel satisfaction without external rewards (e.g. Ordóñez et al. 2009). The motivation can be enjoyment-based, i.e. fulfilling a flow of tasks satisfactorily, or obligation-based, i.e. doing a good job including following norms and promoting the welfare of others (Frey et al. 2013), and is therefore not, or to a very minor degree, triggered by external incentives.

Employees within organisations are employed to perform certain tasks, and in return they receive a salary. When rewards are related to the amount produced by individual employees, either through piece-rates, bonus systems or future wage ne-

gotiations, it is not only salary but also work conditions that are affected. Employees in such situations, whether they produce products or services, usually take actions to protect themselves, their finances, their health and their alternative objectives, provided that they have some kind of discretion.

Blue-collar workers employed within piece-rate systems develop strategies to cope with time and motion studies that monitor their work processes in order to set rates. For example, during observations they perform their tasks more slowly and thoroughly to avoid piece-rates being lowered and their workload increasing. Street-level bureaucrats who work under comparable conditions but produce welfare services use similar coping strategies, for example they prioritise simple cases above complex ones, more profitable cases above others and so on (see, for example, Lipsky 1980; Meyers and Vorsanger 2007). In addition, street-level bureaucrats, i.e. professionals who perform welfare services, are governed to a higher degree than others by intrinsic incentives and values, a condition that can promote coping behaviours aimed at defending professional and administrative ethics (see, for example, Baker 2002; Frey et al. 2013). Many of the behaviours defined as gaming in the performance literature are designated coping strategies in the street-level literature. The concept, gaming or coping that an author uses to interpret the same behaviour provides a clue as to how the phenomenon is apprehended and the perspective applied. The gaming concept is usually linked to a top-down perspective, where the steering capacity of the model is regarded as the main problem, while the coping concept is more often linked to a bottom-up perspective where negative consequences for employees and recipients of services are in focus.

Further, studies of private sector organisations emphasise two main systems that can be used in order to enhance productivity: control and commitment systems, where commitment systems seem to improve goal achievement to a greater extent than control systems (e.g. Arthur 1994). Both systems are used in order to shape behaviour, but while control systems use measures such as monitoring and inspection to promote behavioural change, commitment systems use trust building through, for example, formal participation as a means to develop “committed employees who can be trusted to use their discretion to carry out job tasks in ways that are consistent with organisational goals” (Arthur 1994, 672). Accordingly, commitment systems are implemented with the aim of replacing extrinsic values with intrinsic ones so that as a consequence goal fulfilment increases. The consequence of applying financial incentives in public welfare organisations is the reverse; a commitment system is replaced by a control system, and there is much to suggest that intrinsic values will start to erode (Ordóñez et al. 2009; Maynard 2012). Direct and indirect financial incentives seem to change employee’s attention toward external confirmation, and the mechanism that triggers this process is the seemingly inevitable response to indirect and direct financial incentives: gaming/coping behaviours.

Gaming/coping behaviours at the individual level can thus be attributed to self-defence and vested as well as professional and/or altruistic values, which from this perspective are highly rational responses to incentives even though such behaviours are irrational from a PMS perspective.

Unit level

Employees usually perform their tasks within an organisational unit. Units within organisations have different missions, internal organisational structures and resources. Of course, the formation and functions of units varies between different types of organisations, but organisations without units are rare. Units tend to act as unified actors when rewards and penalties are based on results at the unit level. Units develop gaming and coping behaviours similar to those that individual actors develop, and for the same reasons and the same motives. Resources are usually limited and finite within an organisation's performance measurement cycle, which implies that if one unit is rewarded, others have to take a back seat. Types and forms of gaming/coping behaviour vary depending on whether bad or good scores are linked to additional resources, but they do occur (e.g. Gravelle et al. 2010). As with individuals, unified actors that produce publicly funded welfare services can justify their behaviours through commitment to public and professional values, which they feel are threatened by targets and indicators inherent in PMS. However, performance incentives applied within organisations in relation to units operate more like external than internal incentives, in particular when performance measurement scores are made public within organisations and rewards and penalties are allocated according to performance scores at the unit level. Like corresponding behaviour at the individual level, gaming behaviour at the unit level can be attributed to self-defence and vested, but also professional and/or altruistic, values. However, due to the growing importance of incentives for the unified actor as a collective, vested interests increase at the expense of professional and intrinsic values, and from this perspective it is logical that the breeding ground for rational gaming behaviour improves.

Organisational level

An organisation is usually only one of several organisations that produce similar products or services. Services can be produced by organisations located at different levels within a political administrative system, for example counties, municipalities, schools, hospitals and care centres for the elderly. At these levels, financial incentives are commonly closely linked either to the performance of organisations in operation, i.e. pay for performance and performance contract systems (Gravelle et al. 2010; Maynard 2012), or to direct or indirect financial rewards and penalties based on comparisons of performance scores between organisations (Gao 2009; Pollitt 2013).

The purpose of comparing performance measures between organisations is that efficacy and quality development is assumed to be enhanced and opportunities for accountability improved, mainly because performance measurements are made public. Transparency gives units and organisations, as well as service recipients, a basis for comparisons, and consequently service producers start to compete in order to win the race, particularly if good results are rewarded and bad ones punished.

Distribution of rewards and penalties can be made within systems entirely composed of public organisations or systems where private organisations also produce welfare services, i.e. in NPM-inspired systems with quasi-markets. In these contexts a bad reputation (Lingard and Sellar 2013) and loss of anticipated rewards can be apprehended as indirect financial penalties (Maynard 2012). These kinds of incentives are external in relation to single organisations, since rewards and penalties are to a great extent based on transparent accounts, comparisons, competition and/or ranking activities (Pollitt et al. 2010; Gao 2010; Chan and Gao 2008).

Pay for performance, performance contracts linked to financial rewards as well as rewards and penalties allocated by results is a very fertile ground for *gaming and cheating behaviour*:

Measuring and rewarding products may be a major incentive for game-playing (de Bruijn 2002, 581).

The logic of summative approaches is for incentives and/or sanctions to become associated with performance and for various forms of 'gaming' to arise in response (Pollitt et al. 2010, 21).

The rise of non-mission based targets also induces widespread gaming behaviours among Chinese local officials (Gao 2010, 70S).

for *intrinsic value deterioration*:

Intrinsic motivation is distorted (Frey et al. 2013).

and the *erosion of motivation*:

The potential for ... P4P incentives to erode motivation is considerable (Maynard 2012, 8).

Motivation decreases when individuals perceive the rewards they receive as controlling, contingent on a particular performance (Kallio and Kallio 2012, 11).

Or in other words there will be:

a rise in unethical behavior, ... corrosion of organizational culture, and intrinsic motivation (Ordóñez et al. 2009, 6).

Gaming at an organisational level is to a very high degree attributed to self-defence and vested extrinsic values, rather than professional and intrinsic values, which from an organisational perspective is highly rational, even though such behaviours are irrational from both a PMS perspective and the values inherent in goals maintained by the principals of organisations.

Why we never learn...

In modern societies, internal and external performance measurement incentives applied within PMSs are closely linked to each other in a web of dependencies. Patterns of behaviours at an individual, unit or organisational level emanating from rational responses to internal and external, direct or indirect extrinsic incentives are reinforced in a spiralling manner. An escalation process is set in motion (Bevan and Hood 2006; Pollitt et al. 2010; Pollitt 2013; Hamilton et al. 2013) where measures based on financial incentives and comparisons aimed at counteracting undesired behaviours and effects are doomed to instead exacerbate and reinforce them.

What is actually most interesting is that the problem of gaming does not seem to emanate from incorrect assumptions about what stimulates self-interested rational behaviour, rather the problem seems to be that internal and external extrinsic incentives stimulate precisely that.

First, incentives, particularly financial incentives aimed at altering individuals' bases for rational action, have proven to be very successful; people do change their behaviour. Gaming (coping) is a highly rational response to performance-based incentives at both an individual and a collective level. Incentives trigger and shape self-interested rational behaviour among individuals, units and organisations. When the basis for rational behaviour is altered, it produces a concomitant shift in values. Extrinsic values start to emerge at the expense of intrinsic values and gaming/coping behaviour is aggravated. When additional corrective incentives are applied, opportunities for extrinsic values to emerge are expanded even more. If intrinsic values are suppressed to the degree that extrinsic values start to become prevalent, motivation and ambition among employees decreases. In this situation it is a small step from gaming to cheating behaviour. In situations where intrinsic values loaded with obligation and professionally and democratically based values are repressed or replaced, the moral barrier preventing fraud, fabrication of data, bribery and violence is very fragile.

Second, when levels and incentives within PMSs are linked to each other and performance measurement scores determine the distribution of penalties and rewards through a web of dependencies at the individual, unit and organisation levels, the probability that intrinsic values will be suppressed or replaced increases exponentially. In PMSs dealing with the fulfilment of welfare policy management, professions and auditing units with obligations other than welfare professional core

activities usually grow and gain influence and power (e.g. Power 2003, Lapsley 2008; Evetts 2009; Noordegraaf and Schinkel 2011). These new management professions tend to treasure and promote the value of efficiency to a higher degree than traditional professions and public bureaucrats. Professional, ethical and democratic values, as well as the value of legal security, seem to be more strongly connected to the functions and tasks carried out by traditional professions and public officials. The increased power of management professions and auditing units to exert influence within organisations and over the design of performance measurement systems promotes the spread of management values within organisations which deliver public welfare.

Third, when undesired behaviours are aggravated, becoming marginally corrupt, citizens and politicians alike are on very good grounds in demanding additional measures to deal with these behaviours. However, if these measures are applied in order to handle a performance measurement system with the characteristics typical of organisations providing welfare services in situations where intrinsic values have been crowded out, the odds of success are far from favourable.

Performance Measurement systems dealing with welfare services might have the potential to be effective and appropriate within small self-governed local organisations. However, those local systems most probably need to be more or less cleansed of financial incentives, rewards and penalties and not linked to a web of other organisations and structures. In other (existing) contexts there is much to suggest that PMSs promote self-interested rational individuals, units and organisations with vested interests and are thereby, in practice, the first step towards corruption.

It is not possible to instigate a future of commitment, motivation, intrinsic values and trust through the application of external and internal incentives that promote rational self-interested behaviour and vested interests and institutionalise distrust.

Funding

This research was supported by a grant from RJ, The Swedish Foundation for Humanities and Social Sciences (grant number P12-0317:1).

References

- Argyris, C. 1990. "The Dilemma of Implementing Controls: The Case of Managerial Accounting." *Accounting Organizations and Society* 15(6), 503–511.
- Arthur, J. B. 1994. "Effects of Human Resource Systems on Manufacturing Performance and Turnover." *Academy of Management Journal* 3(3), 670–687.

- Baker, G. 2002. "Distortion and Risk in Optimal Incentive Contracts." *The Journal of Human Resources* 37(4), 728–751.
- Behn, R. D. 2003. "Why Measure Performance? Different Purposes Require Different Measures." *Public Administration Review* 63(5), 586–606.
- Berliner, J. S. 1956. "A Problem in Soviet Business Administration." *Administrative Science Quarterly* 1(1), 86–101.
- Bevan, G. and C. Hood. 2006. "What's Measured is what Matters: Targets and Gaming in the English Public Health System." *Public Administration* 84(3), 517–538.
- Bouckaert, G. and J. Halligan. 2008. *Managing Performance: International Comparisons*. New York: Routledge.
- Caroll, J. S. et al. 2006. "Naturalistic Decision Making and Organizational Learning in Nuclear Power Plants: Negotiating Meaning between Managers and Problem Investigation Teams." *Organisation Studies* 27, 1037–1057.
- Chan, H. S. and J. Gao. 2009. "Putting the Cart before the Horse: Accountability or Performance." *The Australian Journal of Public Administration* 68(S1), S51–S61.
- Chan, H. S. and J. Gao. 2008. "Performance Measurement in Chinese Local Government." *Chinese Law and Government* 41(2–3), 4–9.
- de Bruijn, H. 2002. "Performance Measurement in the Public Sector: Strategies to Cope with the Risk of Performance Measurement." *The International Journal of Public Sector Management* 15(7), 578–594.
- Dur, R. and R. Zoutenbier. 2014. "Working for a Good Cause." *Public Administration Review* 74(2), 144–155.
- Evans, T. and J. Harris. 2004. "Street-Level Bureaucracy, Social Work and the (Exaggerated) Death of Discretion." *British Journal of Social Work* 34, 871–895.
- Evetts, J. 2009. "New Professionalism and New Public Management: Changes, Continuities and Consequences." *Comparative Sociology* 8, 247–266.
- Frey, B. S. et al. 2013. "Organizational Control Systems and Pay-for-Performance in the Public Sector." *Organizational Studies* 34(7), 949–972.
- Fryer, B. S. et al. 2009. "Performance Management in the Public Sector." *International Journal of Public Sector Management* 22(6), 478–498.
- Gao, J. 2010. "Hitting the Target but Missing the Point: The Rise of Non-missioned-Based Targets in Performance Measurement of Chinese Local Governments." *Administration & Society*, 42(IS), 565–765.
- Gao, J. 2009. "Governing by Goals and Numbers: A Case Study in the Use of Performance Measurement to Build State Capacity in China." *Public Administration and Development* 29, 21–31.

- Gravelle, H. et al. 2010. "Doctor Behavior under Pay for Performance Contract: Treating, Cheating and Case Finding?" *The Economic Journal* 120 (February), F129–F136.
- Hamilton, L. et al. 2013. "Improving Accountability through Expanded Measures of Performance." *Journal of Educational Administration* 51(4), 453–475.
- Johansson, V. 2015. "Policy Networks: A Threat to Procedural and Expert-Based Decision Making and the Quality of Public Risk Decisions?" *International Journal of Critical Infrastructure Protection* 9, 3–12.
- Johansson, V. 2012. "Negotiating Bureaucrats." *Public Administration* 90(4), 1032–1046.
- Johansson, V. and L. Lindgren (eds). 2013. *Uppdrag offentlig granskning* [Mission public scrutiny]. Lund: Studentlitteratur.
- Johansson, V. and S. Montin. 2014. "What if Performance Accountability Engenders Distrust?" *Urban Research & Practice* 7(2), 213–227.
- Jones, B. D. 2003. "Bounded Rationality and Political Science: Lessons from Public Administration and Public Policy." *Journal of Public Administration Research and Theory* 13(4), 395–412.
- Kallio, K.-M. and T. J. Kallio. 2014. "Management-by-Results and Performance Measurement in Universities: Implications for Work Motivation." *Studies in Higher Education* 39(4), 574–589.
- Klein, G. 2008. "Naturalistic Decision Making." *Human Factors: The Journal of the Human Factors and Ergonomics Society* 50(3), 456–460.
- Lapsley, I. 2008. "The NPM Agenda: Back to the Future." *Finical Accountability & Management* 24(1), 77–96.
- Lewis, M. J. and P. Triantafillou. 2012. "From Performance Measurement to Learning: A New Source of Governmental Overload?" *International Review of Administrative Sciences* 78(4), 597–614.
- Lingard, B. and S. Sellar. 2013. "'Catalyst Data': Perverse Systemic Effects of Audit and Accountability in Australian Schooling." *Journal of Education Policy* 28(5), 634–656.
- Lipshitz, R. et al. 2001. "Focus Article: Taking Stock of Naturalistic Decision Making." *Organizational Behavior and Human Decision Process* 14, 331–352.
- Lipsky, M. 1980. *Street-Level Bureaucracy: Dilemmas of the Individual in Public Services*. New York: Russell Sage Foundation.
- Mannion, R. and J. Braithwaite. 2012. "Unintended Consequences of Performance in Health Care: 20 Salutary Lessons from the English National Healthcare." *Internal Medicine Journal* 42:5, 569–574.

- March, J. and J. P. Olsen. 2004. "The Logic of Appropriateness." Working Paper 04/08. Oslo: Center for European studies, University of Oslo.
- May, P. J. and S. Winter. 2007. "Politicians, Managers, and Street-Level Bureaucrats: Influence on Policy Implementation." *Journal of Public Administration Research and Theory* 19, 453–476.
- Maynard, A. 2012. "The Power and Pitfalls of Payment for Performance." *Health Economics* 21, 3–12.
- Merchant, K. A. 1990. "The Effects of Financial Controls on Data Manipulation and Management Myopia." *Accounting, Organizations and Society* 14(4), 297–313.
- Meyers, M. and S. Vorsanger. 2007. "Street-Level Bureaucrats and the Implementation of Public Policy." In G. Peters and J. Pierre (eds). *The Handbook of Public Administration*. London: Sage, 153–165.
- Moynihan, D. P. et al. 2011. "Performance Regimes Amidst Governance Complexity." *Journal of Public Administration Research and Theory* 21, i141–i155.
- Noordegraaf, M. 2007. "From 'Pure' to 'Hybrid' Professionalism: Present-Day Professionalism in Ambiguous Public Domains." *Administration & Society* 39(6), 761–785.
- Noordegraaf, M. and W. Schinkel. 2011. "Professional Capital Contested: A Bourdieusian Analysis of Conflicts between Professionals and Managers." *Comparative Sociology* 10, 97–125.
- Nutley, S. et al. 2012. "Scrutinizing Performance: How Assessors Reach Judgments about Public Services." *Public Administration* 12(90), 869–885.
- Ordóñez, L. D. et al. 2009. "Goals Gone Wild: The Systematic Side Effects of Overprescribing Goal Setting." *Academy of Management Perception* 23(1), 6–16.
- Osborne, D. and T. Gaebler. 1992. *Reinventing Government: How the Entrepreneurial Spirit is Transforming the Public Sector*. New York: Plume.
- Otley, D. 1999. "Performance Measurement: A Framework for Management Control System Research." *Management Accounting Research* 10, 363–382.
- Perrin, B. 1998. "Effective Use and Misuse of Performance Measurement." *American Journal of Evaluation* 19(3), 367–379.
- Phelps, R. P. 2011. "Teach to the Test? Most of the Problems with Testing have one Surprising Source: Cheating by School Administrators and Teachers." *Wilson Quarterly* 35(4), 38–42.
- Pidd, M. 2005. "Perversity in Public Service Performance Measurement." *International Journal of Productivity* 54(5/6), 482–493.
- Pierre, J. and G. Peters. 2000. *Governance, Politics and the State*. London: MacMillan Press.

- Pollitt, C. 2013. "The Logics of Performance Management." *Evaluation* 19(4), 346–363.
- Pollitt, C. et al. 2010. "Performance Regimes in Health Care: Institutions, Critical Junctures and the Logic of Escalation in England and the Netherlands." *Evaluation* 16(1), 13–29.
- Poulsen, B. 2009. "Competing Traditions of Governance and Dilemmas of Administrative Accountability: The Case of Denmark." *Public Administration* 87(1), 117–131.
- Power, M. 1997. *The Audit Society: Rituals of Verification*. Oxford: Oxford University Press.
- Power, M. 2003. "Evaluating the Audit Explosion." *Law and Policy*, 25(3), 185–202.
- Radin, B. 2006. *Challenging the Performance Movement: Accountability, Complexity, and Democratic Values*. Washington D.C.: Georgetown University Press.
- Ridgway, V. F. 1956. "Dysfunctional Consequences of Performance Measurements." *Administrative Science Quarterly* 1(2), 240–247.
- Salomon, L. (ed.). 2002. *The Tools of Government: A Guide to the New Governance*. New York: Oxford University Press.
- Schneider, A. and H. Ingram. 1990. "Behavior Assumptions of Policy Tools." *The Journal of Politics* 52(2), 520–529.
- Simon, H. 2000. "Bounded Rationality in Social Sciences: Today and Tomorrow." *Mind and Society* 1, 25–39.
- Simon, H. 1993. "Decision Making: Rational, Nonrational, and Irrational." *Educational Administration Quarterly* 29(3), 392–411.
- Smith, P. 1995. "On the Unintended Consequences of Publishing Performance Data." *Public Sector* 18(2), 277–310.
- Van Dooren, W. et al. 2010. *Performance Management in the Public Sector*. New York: Routledge.
- Van Thiel, S. and F. L. Leeuw. 2002. "The Performance Paradox in the Public Sector." *Public Performance and Management Review* 25(3), 267–281.
- Wilson, D. et al. 2006. "'What Gets Measured Gets Done': Head Teachers' Responses to the English Secondary School Performance Management." *Policy Studies* 27(2), 153–171.
- Zsombok, C. and G. Klein. 1997. *Naturalistic Decision Making*. New York: Lawrence Erlbaum.