A Good Servant But a Poor Master: The Side Effects of Numbers and Metrics

Alexandre Asselineau, Gilles Grolleau, Naoufel Mzoughi

A Good Servant but a Poor Master: The Side Effects of Numbers and Metrics

Alexandre Asselineau

Burgundy School of Business, Université Bourgogne Franche-Comté, Dijon, France

alexandre.asselineau@bsb-education.com

Gilles Grolleau

CEE-M, Montpellier SupAgro, INRAE, CNRS, Université de Montpellier, Montpellier, France

grolda@gmail.com

Naoufel Mzoughi

INRAE, ECODEVELOPPEMENT, 84000, Avignon, France

naoufel.mzoughi@inrae.fr

Abstract: A common practice in managerial and public service contexts is to quantity, calculate and use numbers and metrics which provide a presumption of scientificity, a sense of measurability, objectivity, reliability and precision upon which smarter decisions can be made. Besides providing a theoretical background, we analyze counter-productive effects of over-relying on numbers and metrics, notably in public administration. We discuss the following traps: preferring what is measurable over what is important, replacing the strategy by a measure and dehumanizing the decision making. We suggest some practical ways to facilitate a more parsimonious, smarter and adequate use of numbers.

Keywords: numbers; measurement; metrics; dehumanization; surrogation bias.

A Good Servant but a Poor Master: The Side Effects of Numbers and Metrics

Numbers, metrics and related practices are everywhere in our professional and personal lives. Since childhood, most individuals are conditioned to measure, quantify, compare and behave according to various numbers and metrics such as height, weight, price, student scores and rankings. According to Hummel (2006), the dominance of numbers is such that nothing is considered to be real unless it is measured. Some professions seem even number-based, such as the accounting profession that is frequently perceived in a simplistic fashion as revolving around numbers and number crunching, relegating its human dimension to the background. Numbers constitute "social resources" (Vollmer, 2007) that deeply influence the way we perceive the world around us and play an ordering and orientating function. For instance, public communications regarding the management of the Covid-19 crisis is frequently based on and justified by numbers such as daily new cases, daily deaths, number of patients in ventilation beds. The digital revolution also led to a quantification of ever more areas of human life and society, especially domains that were previously untouched by the measurement fever such as friendship or love or environmental issues. Some neologisms such as 'omnimetrics' and 'numerocracy' (i.e., governance by numbers) are used to describe the hold of measurement and numbers in everyday life (e.g., number of retweets, number of 'friends' on Facebook, number of views or visitors on a public service, or on a YouTube channel).

In the public administration as in the business and private world, numbers and metrics are everywhere. (Public) management by numbers and metrics is booming in recent years with performance-enhancing promises (Hood, 2007; Hood, 2012), leading some authors to label it as a 'revolution' (James et al., 2020). In the New Public Management, numbers serve as governance tools that are not without negative side effects (Siltala, 2013). In the public

administration context, metrics are sometimes considered as the counterpart of prices on markets (Muller, 2018). Hood (2007) described three types of systems of performance measurement expressed in the form of intelligence systems (background information on performance but without fixed data interpretation), ranking systems (comparing the performance of units against one another) or target systems (the measured performance is compared against one or several standards). Analysts and deciders are frequently overwhelmed by the amount of analytic data proposed by technology companies such as Facebook or Twitter. More and more training courses, diplomas or jobs are oriented towards data analysis and public administration is no exception.

Number-based practices can be conducive to a lot of benefits such as communicating a sense of measurability, transparency, objectivity and reliability. In the managerial world, a well-known mantra states that '*What gets measured gets managed*' or the variant *'What gets measured gets done'*. Numbers, metrics and related practices facilitate many processes and operations within and between organizations. By providing numerical targets, even when they are arbitrarily fixed, accountants, public and private managers or leaders can be convinced that they communicate objectively, precisely and clearly. These targets provide a benchmark upon which performance or results can be quickly compared and evaluated, allowing to determine easily whether objectives are met or not. For instance, Google popularized the use of Objectives and Key Results (OKR) with a well-known formula of a former Google's Vice President, Marissa Meyer who stated: '*If it does not have a number, it is not a Key Result*.' Numbers constitute key inputs in the decision-making process. They are believed to elicit more rational thinking and pull away the risk of a too emotionally loaded decision making. Analytics are more and more deployed in order to improve organizational performances (Parmenter, 2010). Interestingly, when citizens are asked to evaluate the performance of a

public organization, they overwhelmingly choose statistical information over episodic information (Olsen, 2017).[1]

Taken together, these benefits can explain the pervasiveness of numbers and metrics in many managerial and accounting activities. Sometimes, managers at all layers can become obsessed, willing to quantify and put numbers and metrics on everything, especially in evaluation activities. Interestingly, it is usual to say that numbers 'speak for themselves'.

Numbers and metrics thus promise a more efficient and productive system. But there is a flip side on the coin: they frequently occult a kind of incompleteness, subjectivity and the intents of number-makers or users. Simply deciding whether or not to measure a representation of reality, and how to measure it, may however reflect policy choices or, more deeply, a conception of the Society. For instance, it is well-known that the United States and France are 'multi' cultural/racial/ethnical societies. Nevertheless, while collecting statistics referring to 'racial or ethnic origin' is forbidden by law in France, notably to support the egalitarian view of the French Nation, these statistics are on the contrary common in the United States. Moreover, several dimensions such as gender equity cannot be reduced to numbers and statistics (Frey, 2019) and even if it is the case, there is frequently a substantial information loss. This information loss can seem necessary to make life easier. Obviously, number-based heuristics are useful because they reduce mental efforts, simplify complex situations, help solving problems and allow to reach quick and accurate decisions in many circumstances.

Nevertheless, even with the best intentions in the world, numbers and metrics can lead to perverse and counterproductive effects. Therefore, we caution public managers to not over-

[1] To make justice to this issue, Olsen (2017) found that even if citizens have stated preferences in favor of numerical performance information, the reality is different. When they evaluate public organizations, the information that actually affects their evaluation and recall of public sector performance is geared towards episodic information (e.g., personal experiences, media case stories) rather than statistical one.

rely on them (Siltala, 2013). Numbers and metrics cannot account the full complexity of the real world, do not adequately communicate sense making and frequently occult important dimensions, notably qualitative ones such as beauty and aesthetics, friendship or love (Frey, 2019; Siltala, 2013). If music is grounded on very mathematical laws, by evidence, music cannot be reduced to mathematics: music evokes emotions which are (at best) only partially measurable, and probably do not have to be. Concretely, an isolated numerical sale target oversimplifies the reality, does not deliver sense making to collaborators and can push them to game the system (e.g., sandbagging) and neglect other high-order dimensions, such as delighting citizens or consumers or improving the underlying system. Giving objectives to a salesperson or to a frontline public servant is not enough to give sense to his/her work. However, sense giving is the essential (non-measurable) dimension to keep motivation, involvement and interest in his/her work over the long term. Everything is not measurable and even if it is, it does not imply that measurement and putting numbers on any possible dimension are desirable. Interestingly, it has been argued that the rise of measurement in many aspects of life will increase the desire and willingness to pay of people for unmeasured aspects (Frey, 2019). The increasing tendency to measure everything and everywhere is frequently intrusive and viewed by individuals as a threat that can undermine and crowd-out their 'intrinsic preferences'. Again, some intrinsic preferences such as love, friendship, the yearning of identity or the joy of jogging cannot be reduced to measures.

A natural domain where numbers play a major role in business is accounting. The latter explicitly exploits several advantages of numbers, although ill-intentioned individuals can pervert the system by manipulating metrics and figures. Surprisingly, one of the very first accountants in United States, James Anyon, urged his successors in 1912 to use figures as little as they can to not be mesmerized by figures and trust them, since clients do not want figures but rather brains (Anyon, 1925). Misappropriations include 'cosmetic accounting',

'creative accounting' and 'cooking the books'. In some cases, numbers can be just misunderstood and misinterpreted or rearranged to orient the reasoning or may be even outrightly falsified. The intents behind numbers matter. Beyond factual or technical factors, the development of international accounting standards (IFRS) also illustrates political issues ('soft power') or even philosophical perspectives (Ball, 2006; Raffournier, 2007). Smith (1992) documents that much of the apparent growth in profits of corporate enterprises in the United Kingdom in the 1980s was as a result of accounting sleight of hand rather than genuine economic growth. Behind rosy numbers, many of the involved firms were providing a false picture of the company's health and subsequently went bankrupt.

As a 'right medication', numbers have some side effects. A better understanding of these side effects of numbers and metrics can help to avoid some undesirable consequences and to develop more comprehensive processes. Using a historian perspective, Muller (2018) argues that metric fixation contributes to the emergence of three misleading promises, precisely (*i*) metrics are truth and transparency conveyers and as a consequence, judgement based on personal experience and talent needs to conform to numerical indicators (*ii*) public metrics facilitate accountability by allowing to check whether institutions are actually carrying out their purposes, (*iii*) metrics help to motivate individuals within organizations by attaching rewards or penalties to their measured performance. We are not the first to sound the alarm bell (e.g., Hood, 2012; James et al., 2020), but we go deeper by examining three mechanisms by which metrics and measurement obsessions can deteriorate rather than improve decision-making and success. An important originality of our contribution is the development of practical solutions to encourage a wise use of numbers that reduces their detrimental effects. We posit that a relevant use of numbers can provide an unexpected advantage to public organizations to reach their long-term goals. We illustrate our arguments by relevant real-world examples borrowed from the public and private organizations, and, all

along the paper, we draw some managerial implications. Indeed, we believe that public administrations can benefit from lessons learnt in the private and for-profit sector and vice versa.

AN OBSESSION TO NUMBERS AND METRICS: A THEORETICAL BACKGROUND

A logical departure point is the rational choice theory. This theory assumes that agents (e.g., managers) make decisions based on relevant information in a logical, timely and optimized fashion. These agents with a well-defined set of preferences have all the relevant information, evaluate options by calculating their respective resulting utility and select the option that maximizes their utility (Becker, 1976). This selection process frequently assumes that numerical values or utilities are associated with each option, making obvious the choice in favor of the alternative that delivers the highest utility. Interestingly, this theoretical framework implicitly gives a natural role to numbers that allow to synthetize information, perform tradeoffs and compare alternatives in a logical manner. But a real-world application of this model is very demanding and costly.

Over time, this view of human decision and behavior has been heavily questioned, especially since the seminal contribution of Herbert Simon on bounded rationality. Indeed, there is a substantial literature emphasizing that managers and public servants do not follow the prescriptions of rational choice theory and frequently use heuristics to make decisions. Exploring how humans really decide and behave (when conditions assumed by rational choice theory are not met) allowed the emergence of the heuristics and biases-based programs (Kahneman, 2003; 2011; Gigerenzer and Brighton, 2009; see also Olsen, 2015 and James et al., 2020 regarding the application of the behavioral science revolution in the context of public performance metrics). These heuristics constitute simple and efficient rules of thumbs

that individuals and especially managers often use to make decisions (Luan et al., 2019). These mental shortcuts usually imply focusing on one (or few) aspect(s) of a complex problem and neglecting other dimensions. They can even deliver decisions that are better than the theoretically optimal procedure, especially in complex and uncertain environments. Metrics and numbers can conceptually serve as heuristics, leading to analyze which numbers and metrics will yield smart judgments, and which ones will not. So, numbers are not bad in themselves, they are even necessary, but they can be ill-used and lead to disastrous consequences.

Numbers can even play a performative role in reconfiguring the reality. For instance, H.H. Asquith, who served as Prime Minister of the United Kingdom, stated that the War Office kept three sets of figures: '*one to mislead the public, another to mislead the Cabinet, and the third to mislead itself*' (Horne, 1962). Using the considerable psychological literature on how numbers influence human attitudes and behaviors, Olsen (2015) proposed to consider their application to numerical performance information. He notably developed four aspects that illustrate previously ignored possible effects of performance information, namely (1) human attention to round numbers and the leftmost-digit bias, (2) numerical equivalence framing, (3) numerical precision and confidence, and (4) comparison of numbers in space and time. This interest in applying behavioral insights to public administration has led to the emergence of a relatively new field: behavioral public administration, even if this nascent literature focused more on documenting the existence of cognitive biases in public administration contexts (Grimmelikhuijsen et al., 2017; Battaglio et al., 2018) and less on proposing or designing behaviorally informed solutions (Bhanot and Linos, 2020).

Interestingly, a recent approach labelled 'behavioral public performance' proposes to connect the performance metrics revolution and the behavioral science revolution in order to improve the design and use of performance metrics in public management and democratic

accountability (James et al., 2020). On the basis of this research program, we posit that numbers are used not only for their objective content but also for their ability to activate automatic processes in human beings. This dual dimension can make the effects of numbers and metrics very different from what was initially expected or predicted by a rational perspective. Let us now consider some inappropriate uses of numbers and metrics.

THE WRONG TARGET: PREFERRING WHAT IS MEASURABLE OVER WHAT IS IMPORTANT

"Not everything that counts can be counted, and not everything that can be counted counts" (Albert Einstein)

A collateral effect of the 'measurement bias' is to select dimensions that can be measured rather than important dimensions. In many contexts, what really matters[2] (e.g., helping others, passion for work, making the world a better place, identity) is not amenable to measurement. The willingness-to-put a number on everything can switch the attention from important but difficult-to-measure matters to less important but easy-to-measure dimensions. Individuals are sometimes formatted to see value in what they can measure rather to measure what they really value. For instance, because it is measurable, the time spent at the workplace, is more used, documented and monitored (e.g., time clocks), compared to the difficult-to-measure motivation and involvement at work, even if they are much more likely to influence the overall performance of an organization. In short, rather than *measuring what really matters*, the limitations of measurement tools push leaders and managers to prefer *what can be measured* and to label it afterwards as what really matters.

[2] What really matters can be defined by the strategy of the organization or by individuals inside the organization.

The numbers and metrics-based managerial practices in many organizations focus the attention on some (potentially wrong) dimensions and may lead inadvertently to neglect more crucial dimensions, which are not fully reflected in accounting data or captured by current numbers and metrics used in the company. For example, how to really measure a manager's (lack of) leadership or skills, or workplace atmosphere? Even if scholars have developed some measurement scales regarding some difficult-to-measure dimensions (e.g., Kanungo, 1982; Gagné et al., 2010), they are mainly designed to address research questions rather than guide managerial actions. These multidimensional 'hidden dimensions', which have frequently a qualitative nature, escape the measurement tools used in most organizations and are consequently disregarded. Nevertheless, we argue that they can constitute a goldmine as a source of performance and sustainable competitive advantage for organizations. This analysis echoes a nascent literature labelled the 'socio-economic approach to management' (Conbere et al., 2016) arguing that, as people interact with organizational structures, there are gaps in performance which can create a very costly drag on the economic performance of an organization. Moreover, those costs may remain hidden because the cause is difficult to isolate. As such, even the best financial reporting is often too aggregated for business leaders to see the true reasons why a given organization is not performing as expected.

The growing impact investing industry offers another convincing example. Indeed, by definition, impact investments are those made in order to achieve positive, *measurable* social and environmental impact alongside a financial return (GIIN, 2012). Beyond financial returns which seem a priori easy-to-measure with well-standardized and consensual procedures and relevant benchmarks (e.g., risk-adjusted market returns), measuring multidimensional and non-financial outcomes is more challenging. This condition can push managers to select investments that are amenable to measurement rather than those that really matter to make the world a better place. Public managers can fall into the same trap by rewarding on the basis of

metrics that are easy-to-measure without necessarily reflecting the pursued objective. These measures also facilitate the automation of some functions or decisions, by providing numbers that feed algorithms.

To avoid the 'wrong target' trap, an important and first step is to delineate what really matters for the individual, team, public administration, or company. What gives significance and meaningfulness to the performed tasks? What are the public administration priorities? An important issue that follows the previous ones is related to checking whether the numbers and metrics currently used really help users to focus on what really matter. Then, it makes sense to verify if a wrong approach has been followed by leaving metrics and numbers define the core missions and determine what matters. In this case, it can be beneficial to make a step back and rethink what really matters, what gives meaning and significance to the public administration or company, in alignment with the mission of the organization. A concrete example is provided in a study conducted by Grant (2008a, 2008b), who found that collaborators often fail to live up to their potential due to the fact they have lost track of the significance and meaningfulness of their jobs. More precisely after fundraising callers serving a public university met a fellowship student who benefited from the funds raised by the organization, they significantly improved their performances (Grant, 2008b). By reminding them very concretely how their jobs benefit others or why their jobs are important (e.g., thanks to citizen or customer stories or testimonials) and not just the numerical target(s) they are asked to hit, employees might become more motivated and productive. Even accountants can become more productive and satisfied by complementing usual operations with focus on the impact they generate for their clients' businesses and lives. To do this, we must perhaps start from a basic principle: having your nose glued to the financial or other indicators is probably not the best way to improve these financial indicators.

‘THE TREE THAT HIDES THE FOREST’ OR THE SURROGATION BIAS

The well-known Goodhart’s law goes as follows: *When a measure becomes a target, it ceases to be a good measure.* In a similar vein, the surrogation bias describes the tendency through which the measure of a strategy evolves to replace the strategy itself. This confusion can be very dangerous, especially when the metrics do not match well with the strategy (Harris and Tayler, 2019). Choi et al. (2012) defines surrogation as the tendency for managers to lose sight of the strategic construct(s) the performance measures are intended to represent, and subsequently act as though the measures are the constructs of interest. Such a bias is likely to be exacerbated by perverse incentive schemes that promote hitting the numerical targets in the short term rather than achieving the real and long-term strategy of an organization. In other words, a risk associated to metrics is their ability to divert attention from what was really looked for thanks to the measure itself. An interesting example is provided by the public scheme designed to reduce the snake overpopulation in Abruzzi, Italy. The town officials offered a monetary reward for dead vipers on a per head basis, but the snake population continued to increase. This pay-for-performance metrics backfired. Indeed, inhabitants began breeding snakes in order to kill them and make extra money (Rhoads, 1985).

Intangible or abstract deserving goals (e.g., producing ‘good research’) are sometimes denatured by metrics that can be misleading (number of publications, citation counts or h-index) and even sometimes by subsequent malicious compliance. Too frequently, several (flawed) metrics are available, leading people to just select (or even create) the measure just to look good under the used measure (Harris and Tayler, 2019). For example, in 2008, the French educational institutions were not well-ranked in the Shanghai Academic Ranking of World Universities according to a formula that takes into account the number of Nobel laureates and Fields Medals, highly-cited researchers, articles published in top reviews, some

citation indexes and the per capita academic performance. A rival ranking, based not on academic performance, but on the number of CEOs from a given university in world class firms as defined by the Fortune Global 500, was created. As a result, the French educational institutions got much better rankings (Galochkin et al., 2012). In the same vein, Espeland and Sauder (2007) show the considerable effects of rankings applied to law schools in the USA. They found that rankings generate reactivity, in the sense that people change their behavior in reaction to being evaluated, observed and measured. This reactive conformance is notably induced by two mechanisms, precisely commensuration (transformation of qualities in quantities that share a metric) and self-fulfilling prophecies.

Numerical targets encourage narrow focusing and make individuals neglecting the pursued goals and even engage in unethical behaviors such as over-reporting (e.g., salami publication). Let us consider another realistic situation where the organizational goal is citizen or consumer satisfaction and delight. This goal is measured thanks to citizen or consumer satisfaction surveys. Rather than attempting to satisfy and delight citizens, frontline public servants can engage more in pestering and even bribing citizens to get good scores rather than improving the citizens' experience. For instance, several Amazon sellers have been accused to offer 'bribes' to push consumers to leave great (fake) reviews for their products and public servants can be tempted to adopt similar strategies. In a similar vein, many managers can attempt to inflate service delivery or sales in order to make-up the performance of the public administration or company.

In order to hit the numerical targets rather than the real underlying objectives, many examples evidence that individuals engage in unethical behavior (Ordóñez et al., 2009). A well-publicized example is the Wells Fargo scandal that was partly caused by the pressure to meet unattainable sales goals sell that pushed employees to open fraudulently customer accounts. In 2016, the Wells Fargo company fired more than 5,000 employees, paid a $185

million fine and reimbursed $5 millions of wrongly-charged fees to customers (Zoltner et al., 2006). A Wells Fargo mantra was '*Eight is Great*', with the meaning that the goal was to put eight Wells Fargo products into the hands of each customer. To address this issue, the Wells Fargo CEO, John Stumpf, stated: '*We are eliminating product sales goals because we want to make certain our customers have full confidence that our retail bankers are always focused on the best interests of customers*.' In the context of public administration, Speklé and Verbeeten (2014) found empirical evidence that using the performance measurement system for incentive goals may generate dysfunctional consequences and negatively influences organizational performance, although this effect is less pronounced when contractibility is high.

In many managerial contexts, managers use numerical targets. They can be considered according to two main characteristics (Table 1). The first one corresponds to the degree to which the target overlaps or is aligned with the real goal. Sometimes there is a discrepancy between the real pursued goal and the numerical target. This gap can be due to the inability of measurement tools to measure the progress towards the real goal. These tools can push deciders to use a more or less close substitute or indicator that is more amenable to measurement. This gap can be reinforced by strategic considerations that make people gaming the system by reaching the target without delivering the real goal (see Hood, 2006 for an interesting application in public administration). For instance, in some hospitals, staff was rewarded according to wait times, given that waiting less is considered as better for patients' health. Many individuals cheated the system to meet the numerical targets on the paper without delivering adequate services to patients. Efforts were oriented to meet the target, by making patients waiting outside, in order to delay their registered arrival time, but not to enhance their health. The numerical targets diverted attention from the important dimensions, eroded morale and created perverse incentives that were not in the interests of the patients'

health but of the staff's vested interests. The second characteristic corresponds to whether the numerical target is accompanied with other relevant information. This extra information can mix qualitative and quantitative elements and makes the best of two worlds. It gives a more complete picture, allows to better capture what is really looked for, even if it is multidimensional and avoids over focusing on a single aspect.

Please, insert Table 1 around here

Although these two characteristics are obviously continuums, we focus our discussion on the emphasized polar situations. In situation A, everything goes wrong and a detrimental outcome is very likely. An interesting example occurred at Lincoln Electric where the pay of secretaries was tied to their productivity, with counters installed on typewriters to measure how much was typed. The secretaries responded by spending their lunch hours typing useless pages of manuscript in order to increase their pay. The incentive was quickly abandoned (McKenzie and Lee, 2006). A similar outcome occurred when lines of code or bug fixed were used to evaluate the productivity of software developers who produce more lines of code or breed bugs, without necessarily improving productivity.

In situation B, the overlap between the pursued goal and the numerical target is weak, but the extra information is relevant and can partially compensate and lead to a not too bad outcome. For instance, churn rate is ideally considered as an indicator of customer satisfaction and loyalty. Some companies define churn as customers who have not bought anything for 30 days since their last purchase. If you sell airline tickets, this definition can fit well a high-volume corporate customer and indicate a serious concern, but the same measure can be misleading for customers who fly only two or three times a year. A good knowledge of customers can prevent naïve interpretation of numbers (Wijono, 2020).

The situation C corresponds to a very frequent one, where a dimension is over-emphasized while surrounding and useful extra-information is not available or neglected. This 'silo' effect prevents individuals from seeing the global picture and leads to suboptimal decisions and behaviors. For instance, given that police officers are evaluated on how low the crime rate in their precinct is, they may prevent or discourage people from filing reports. Apparent crime rates will remain low because many crimes are not reported (Wijono, 2020).

The situation D is the ideal situation where the alignment between the pursued goal and the numerical target is good and relevant extra information is available and wisely used. Unfortunately, this situation is scarce in many organizations, because it is resource-demanding and not necessarily rewarding in the short run.

To avoid or reduce the surrogation bias, a crucial operation is to check honestly whether the metrics used are really and not only superficially aligned with the strategy, especially vision and core mission of the company. Empowering wisely top and middle managers who will implement the strategy can help. Given that managing by numbers encourage number manipulation, it is important to evaluate the possible misalignment or discrepancy between the used metrics and the pursued strategy and to reduce them accordingly. Loosening the relationship between metrics and financial incentives can also help by reducing the gamification incentive. Rather than using a one-dimensional metric, multiple metrics, including qualitative data such as complementary narrative reporting can allow to cross-check and reduce the surrogation tendency (Bentley, 2019). Speklé and Verbeeten (2014) encourage politicians and policymakers to consider a repertoire broadening beyond conventional New Public Management thinking to enhance the public administration effectiveness. An example that is frequently reported concerns the management of a restaurant chain seeking to improve customer satisfaction (Warren et al., 2019). They used online customer reviews as a measure of this strategic objective. Few after, they noticed one

particular location that was enjoying dramatic increase of the online customer review rating and paid out the increased bonus. Upon further investigation, management found that the manager of this location was offering 15% discounts on all purchases to anyone who gave the restaurant a 5-star review. Investigating the story behind the numbers was key to understand what was really happening.

Rather than just shooting for the numbers, public managers have to retrieve the underpinning aspirations that inspired the goals in the first place. When setting a goal, it is important to document why this goal is important and contributes to the strategy and how to achieve it and then consider appropriate metrics. Metrics must remain an instrument or a tool, not an end in itself. The problem is often not the metrics and numbers *per se* but the leaders and managers who are not accountable for the weak or flawed metrics they select, use and do not act to eliminate any detrimental misalignment (Schrage, 2019). Sometimes, it makes sense to let professionals (not managers) design metrics that fit their professional ethos and really help them in achieving their higher-level goals without being guided by monetary incentives or penalties (Muller, 2018). To make justice to this issue, it must be acknowledged that a lot of metrics are not decided by the managers but are often imposed to them, such as reporting obligations, addressing the requests of the hierarchy, other administrations or departments, meeting demands of auditing bodies, complying with regulations and so on. Sometimes, even when a good metric is used, it is beneficial to let scientific evidence and practice inform the practical implementation. To give a concrete example, research has demonstrated that people enjoy goals that are both attainable and challenging. By making a small change from single number goals to high-low goals or a range (i.e., goals that have a high-low range that averages the same, e.g. getting 8 – 12 new client contracts as opposed to 10), leaders and managers can increase performance and satisfaction (see Scott and Nowlis, 2013). Another practical way

can be to proactively encourage counterfactual thinking and the constructive consideration of disconfirming evidence (Kray and Galinsky, 2003).

THE DEHUMANIZING POWER OF NUMBERS AND METRICS

'*A single death is a tragedy; a million is a statistic'*, said Joseph Stalin. Numbers and metrics can make managers forgetting the 'human reality'. The accounting profession is frequently described as number-centric-rather than people-centric. Numbers and statistics are frequently used to elicit a more rational and less emotional thinking. It is sometimes argued that this dehumanization through numbers and statistics will lead to improved problem solving, but this view is not supported by the evidence (Christoff, 2014; see also Siltala, 2013 in the context of public administration). Managing by numbers is likely to transform the nature of some important activities such as health care or education by running them as for-profit businesses. Thinking only with numbers can create blind spots. For instance, following the insightful distinction introduced by Frank Knight (1921), a number-driven manager can take into account risk (the outcomes are unknown but governed by probability distributions that are known at the outset) but neglect uncertainty (in which both the outcomes and the probability models governing them remain unknown) and the occurrence of 'black swans'.

It is well documented that when the numbers of victims go up, the sympathy onlookers feel goes down according to the human tendency to turn away from mass suffering. Statistics allow to avoid the identifiable victim effect that could temper some potentially unethical managerial decisions such as firing unfairly employees or cheating at a large scale (Yam and Reynolds, 2016). Similarly, experimental and archival studies have showed that the perceived harm of an identical crime or wrongdoing decreases when the number of victims is greater and has labelled this effect as the scope-severity paradox (Nordgren and McDonnell, 2011). The desensitizing or numbing power of numbers, or blindness to scale, is powerful and can be

used to make the unacceptable more acceptable (Slovic, 2007), especially in a big data environment. Recently, this effect was remarkable in some messages related to the Covid-19 crisis. Choosing to downplay the numbers (e.g., number of deaths, number of new cases, total recovered and so forth) is likely to lead to a lack of empathy (Slovic, 2007).

Interestingly, just putting a face or a narrative regarding a specific individual rather than just getting an anonymous ID number can transform a task. Giving to employees the possibility to enjoy the face of a delighted citizen or consumer can transform their work. This effect is exemplified in a well-known statement of Mother Theresa: 'If I look at the mass I will never act. If I look at the one, I will.' (Slovic, 2007). For instance, when a photograph of a patient was included on a CT or X-Ray, doctors recommended more caring and attentive treatment (Wendling, 2009). In a public administration context, Olsen (2017) recommended that policy makers and managers pay closer attention to how performance data can provide citizens with a more vivid and emotional account of public services to complement the often pallid statistical performance data. Provoking regular meetings where users of statistics (e.g., senior management in a hospital) are put in customer- or employee-facing positions, experience close contact with well-identified victims (e.g., complaining patients, frustrated medical staff) can recreate connections with the ground and avoid a too dehumanized perspective. In a similar vein, recent research points that donors can be nudged to be more generous when they are asked to give a hypothetical amount to one person before deciding how much they will actually donate to a group of needy people (*Hsee* et al., 2013).

IMPLICATIONS AND CONCLUSION

In addition to the implications mentioned all along the paper, we would like to emphasize four major insights. First, numbers and measures have their place in public and private organizations, but believing that only what can be measured can be managed is wrong.

Using judgment to put measurement back into its proper place (Hummel, 2006) and getting the numbers right will help but will not replace true leadership skills that allow to manage intangible dimensions such as commitment to work and trust, to fully address the uncertainty and/or the complexity of many situations, but also to select relevant numbers and metrics in an accountable fashion (George, 2003; Luthans and Avolio, 2003; Schrage, 2019).

Second, even if it is commonly assumed that the numbers speak for themselves, there is an urgent necessity to avoid number obsession, metrics fixation and question them. Auditing used metrics to check whether they really serve the public administration goals is crucial. Numbers are a good servant but a poor master. This quest could lead to a better understanding of their (ir)relevance according to circumstances, the intentions of number-makers and number-users, but also the limitations of numbers and the need to supplement them with unmeasured aspects. Numbers and most quantitative analysis help to address the 'how many', 'how often' issues but not the 'why' and 'how' ones (the story behind the numbers) that also deserve attention, perhaps more than ever before, to succeed in a sustainable way (see Eisenhardt and Graebner, 2007). There is power in numbers, but this power needs to be understood and channeled in the right direction.

Third, in the 'number game', less can be more. In many situations, less but *better* numbers can be what is really needed, with more accompanying materials such as the human content, the narrative or story behind the numbers. This balanced perspective echoes the recommendation of a business management expert, Tom Peters, in favor of 'managing by walking' or 'management by wandering' where situations are not only understood from a world that is theoretically or conceptually contemplated with an overwhelm of numbers but from a world that is practically experienced (Zundel, 2012). For instance, this experience can imply to spend time with frontline public service employees to better understand their emotional labor and even with beneficiaries.

Fourth, identical numbers do not say the same things to various audiences. Numbers interact in predictable but not necessarily rational ways with people who are exposed to them. The needs of various audiences are not the same (see Walker et al., 2018). Understanding how numbers replaced in their context could be purposefully or not used and even manipulated to activate System 1 or System 2 thinking mode is crucial to promote a wiser use. Rather than just believing the numbers speak for themselves, we advocate to adopt a broader approach that addresses crucial contextual issues such as who makes numbers speaking, to whom, why, how and when.

We live more and more in a data driven world where numbers and metrics are everywhere. There is no such thing as neutral numbers and metrics. The latter are double-edged swords. They can cover up some realities of organizations while contributing at the same time to enhance the understanding of other dimensions. They can help analysis and decision making but they can also distort reality and occult important dimensions. They can involuntarily impose a tyranny and narrow our vision of the surrounding individuals and world. We discussed three mechanisms by which numbers, metrics and related activities can lead to counterproductive effects, but the list is not exhaustive. For instance, individuals can be (strategically or not) overwhelmed with numbers, leading them to miss the point and to poor decision making. The number frame and presentation can be manipulated through various parameters such as selecting between row numbers, percentages, ratios, proportions or the covered period and so forth (James et al., 2020). Individuals can fail to reach targets to avoid ratchet effects and so on. Managing by numbers can also undermine intrinsic motivations of individuals because they are not well reflected in the used metrics. Metrics can generate conflicts by facilitating ‘silo’ thinking.

Rather than adhering blindly or suppressing numbers such as numerical targets or quantitative goals or simply eliminating them, we advocate in favor of a wise and equilibrated

use. Numbers and metrics are highly context and perspective-dependent. Providing context with numbers, such as narratives, can help reducing their side effects. Sometimes, deciding to not measure and put a number on some dimensions makes sense. Considering explicitly the limits of numbers and their side effects can help public leaders and managers to take advantage and make the best of them.

REFERENCES

Anyon, J. (1925). *Recollections of the early days of American accountancy*, 1883-1893. New York: Reprint Zeff, 1988.

Ball, R. (2006). International Financial Reporting Standards (IFRS): pros and cons for investors. *Accounting and Business Research*, 36, 5–27. https://doi.org/10.1080/00014788.2006.9730040

Battaglio Jr, R. P., Belardinelli, P., Bellé, N., & Cantarelli, P. (2019). Behavioral public administration ad fontes: A synthesis of research on bounded rationality, cognitive biases, and nudging in public organizations. *Public Administration Review*, 79(3), 304-320.

Becker, G. (1976). *The economic approach to human behavior*. University of Chicago Press, Chicago.

Bentley, J. W. (2019). Decreasing operational distortion and surrogation through narrative reporting. *The Accounting Review*, 94(3), 27–55. https://doi.org/10.2308/accr-52277

Bhanot, S.P. and Linos, E. (2020). Behavioral Public Administration: Past, Present, and Future. *Public Administration Review*, 80: 168-171. https://doi.org/10.1111/puar.13129

Choi, W., Hecht, G., & Tayler, W. B. (2012). Strategy selection, surrogation, and strategic performance measurement systems. *Journal of Accounting Research*, 51(1), 105–133. https://doi.org/10.1111/j.1475-679X.2012.00465.x

Christoff, K. (2014). Dehumanization in organizational settings: Some scientific and ethical considerations. *Frontiers in Human Neuroscience*, 8, 748. 10.3389/fnhum.2014.00748

Conbere, J., Savall, H., & Heorhiadi, A. (2016). *Decoding the socio-economic approach to management. Results of the second SEAM conference in the United States*. Information Age Publishing.

Eisenhardt, K. M., & Graebner, M. E. (2007). Theory building from cases: Opportunities and challenges. *Academy of Management Journal*, 50(1), 25–32. https://doi.org/10.5465/amj.2007.24160888

Espeland, W. N., & Sauder, M. (2007). Rankings and Reactivity: How Public Measures Recreate Social Worlds. American Journal of Sociology 133 (1): 1– 40.

Frey, B. S. (2019). Awards in the digital world. *International Review of Economics*, 66(1), 29–36. https://doi.org/10.1007/s12232-018-0291-1

Gagné, M., Forest, J., Gilbert, M. H., Aubé, C., Morin, E., & Malorni, A. (2010). The motivation at work scale: Validation evidence in two languages. *Educational and psychological measurement*, 70(4), 628–646. https://doi.org/10.1177/0013164409355698

Galochkin, I., Grolleau, G., & Sutan, A. (2012). Escaping the zero-sum game of positional races. *Kyklos*, 65(4), 464–479. https://doi.org/10.1111/kykl.12002

George, B. (2003). *Authentic leadership: Rediscovering the secrets to creating lasting value*. San Francisco: Jossey Bass.

Gigerenzer, G., & Brighton, H. (2009). Homo heuristicus: Why biased minds make better inferences. *Topics in Cognitive Science*, 1(1), 107–143. 10.1111/j.1756-8765.2008.01006.x

Global Impact Investing Network (GIIN). (2012). What you need to know about impact investing. https://thegiin.org/impact-investing/need-to-know/#what-is-impact-investing

Grant, A. M. (2008a). The significance of task significance: Job performance effects, relational mechanisms, and boundary conditions. *Journal of Applied Psychology* 2008, 93(1), 108–124. 10.1037/0021-9010.93.1.108

Grant, A. M. (2008b). Employees without a cause: The motivational effects of prosocial impact in public service. *International Public Management Journal,* 11(1), 48–66. https://doi.org/10.1080/10967490801887905

Grimmelikhuijsen, S., Jilke, S., Olsen, A. L., & Tummers, L. (2017). Behavioral public administration: Combining insights from public administration and psychology. *Public Administration Review*,7 7(1), 45-56.

Harris, M., & Tayler, B. (2019). Don't let metrics undermine your business. *Harvard Business Review*, September–October, 63–69.

Hood, C. (2012). Public management by numbers as a performance-enhancing drug: Two hypotheses. *Public Administration Review*, 72(s1), 85–92. https://doi.org/10.1111/j.1540-6210.2012.02634.x

Hood, C. (2006). Gaming in target world: The targets approach to managing British public services. *Public Administration Review*, 66(4), 515–520. https://www.jstor.org/stable/3843937

Hood, C. (2007). Public service management by numbers: Why does it vary? Where has it come from? What are the gaps and the puzzles? *Public Money and Management*, 27(2), 95-102.

Hsee, C. K., Zhang, J., Lu, Z. Y., & Xu, F. (2013). Unit asking: A method to boost donations and beyond. *Psychological Science*, 24(9), 1801–1808. 10.1177/0956797613482947

Hummel, R. P. (2006). The triumph of numbers. Knowledges and the mismeasure of management. *Administration and Society*, 38(1), 58–78. https://doi.org/10.1177/0095399705284202

James, O., Moynihan, D. P., Olsen, A. L., & Van Ryzin, G. G. (2020). *Behavioral public performance: How people make sense of government metrics* (Series: Elements in Public and Nonprofit Administration). Cambridge University Press.

Kahneman, D. (2003). Perspective on judgment and choice: Mapping bounded rationality. *American Psychologist*, 58(9), 697–720. 10.1037/0003-066X.58.9.697

Kahneman, D. (2011). *Thinking, fast and slow*. New York, NY: Farrar, Straus and Giroux.

Knight, F. H. (1921). *Risk, uncertainty and profit*. Boston: Houghton Mifflin.

Kanungo, R. N. (1982). Measurement of job and work involvement. *Journal of applied psychology*, 67(3), 341–349. https://doi.org/10.1037/0021-9010.67.3.341

Kray, LJ., Galinsky, AD. (2003). The debiasing effect of counterfactual mind-sets: Increasing the search for disconfirmatory information in group decisions. *Organizational Behavior and Human Decision Processes*, 91(1): 69-81.

Luan, S., Reb, J., & Gigerenzer, G. (2019). Ecological rationality: Fast-and-frugal heuristics for managerial decision making under uncertainty. *Academy of Management Journal*, 62(6), 1735–1759. https://doi.org/10.5465/amj.2018.0172

Luthans, F., & Avolio, B. J. (2003). Authentic leadership: A positive developmental approach. In K. S. Cameron, J. E. Dutton, & R. E. Quinn (Eds.). *Positive organizational scholarship*. San Francisco: Barrett-Koehler, 241–261.

McKenzie, R. B., & Lee, D. R. (2006). *Microeconomics for MBAs: The economic way of thinking for managers*. Cambridge University Press: Cambridge.

Muller, J. Z. (2018). *The tyranny of metrics*. Princeton University Press

Nordgren, L. F., & McDonnell, M.-H.M. (2011). The scope-severity paradox: Why doing more harm is judged to be less harmful. *Social Psychological and Personality Science*, 2(1), 97–102. https://doi.org/10.1177/1948550610382308

Olsen, A. L. (2017). Human interest or hard numbers? Experiments on citizens' selection, exposure, and recall of performance information. *Public Administration Review*, 77(3), 408–420. https://doi.org/10.1111/puar.12638

Olsen, A. L. (2015). The numerical psychology of performance information: Implications for citizens, managers, and policymakers. *Public Performance & Management Review*, 39(1), 100-115.

Ordóñez, L., Schweitzer, M. Galinsky, A., & Bazerman, M. (2009). Goals gone wild: The systematic side effects of overprescribing goal setting. *Academy of Management Perspectives*, 23(1), 82–87. https://doi.org/10.5465/amp.2009.37007999

Parmenter, D. (2010). *Key performance indicators. Developing, implementing, and using winning KPIs*. New Jersey: John Wiley & Sons.

Raffournier, B. (2007). Les oppositions françaises à l'adoption des IFRS : examen critique et tentative d'explication. *Comptabilité-Contrôle-Audit*, 3(3), 21–41. https://doi.org/10.3917/cca.133.0021

Rhoads, S. E. (1985). *The economist's view of the world*. Cambridge University Press, Cambridge, UK.

Schrage, M. (2019). Don't let metrics critics undermine your business. *MIT Sloan Review*, October 23.

Scott, M. L., & Nowlis, S. M. (2013). The effect of goal specificity on goal reengagement. *Journal of Consumer Research*, 40(3), 444–459. https://doi.org/10.1086/670766

Siltala, J. (2013). New public management: The evidence-based worst practice? *Administration and Society*, 45(4), 468–493. https://doi.org/10.1177/0095399713483385

Slovic, P. (2007). "If I look at the mass I will never act": Psychic numbing and genocide. *Judgment and Decision Making*, 2(2), 79–95.

Smith, T. (1992). *Accounting for growth*, London: Random House.

Spekle, R. F., & Verbeeten, F. H. (2014). The use of performance measurement systems in the public sector: Effects on performance. Management Accounting Research, 25(2), 131-146.

Vollmer, H. (2007). How to do more with numbers: Elementary stakes, framing, keying, and the three-dimensional character of numerical signs. *Accounting, Organizations and Society*, 32(6), 577–600. https://doi.org/10.1016/j.aos.2006.10.001

Walker, R. M., Lee, M. J., James, O., & Ho, S. M. (2018). Analyzing the complexity of performance information use: Experiments with stakeholders to disaggregate dimensions of performance, data sources, and data types. *Public Administration Review*, 78(6), 852-863.

Warren, C., Jones, J. P., & Tayler, W. B. (2019). *Financial & managerial accounting* (15th Edition). Cengage Learning, Inc, Mason, OH, United States.

Wendling, P. (2009). Can a photo enhance a radiologist's report? *Clinical Endocrinology News*, 4, 6–9.

Wijono, W. (2020). Musings on metric, manipulation, and mechanism. *The Startup*, February 18.

Yam, K. C., & Reynolds, S. J. (2016). The effects of victim anonymity on unethical behavior. *Journal of Business Ethics*, 136, 13–22. https://doi.org/10.1007/s10551-014-2367-5

Zoltners, A. A., Sinha, P. K., & Lorimer, S. E. (2016). Wells Fargo and the slippery slope of sales incentives. *Harvard Business Review*, September 20.

Zundel, M. (2012). Walking to learn: Rethinking reflection for management learning. *Management Learning*, 44(2), 109–126. https://doi.org/10.1177/1350507612440231

Table 1: A simple characterization of numerical targets

	Numerical target used without other information	**Numerical target used with other information**
Target unaligned with the real pursued goal	***Situation A*** Worst situation	***Situation B*** Uninformative measure possibly counterbalanced with other information
Target aligned with the real pursued goal	***Situation C*** Risk of overemphasis on one dimension	***Situation D*** Ideal situation for good analysis