1 Introduction: Connecting Two Revolutions

This Element responds to developments in two coinciding revolutions. The first is a revolution in governance that has made quantification of performance a central strategy in managing public goods and services. The second is a revolution in social science research that employs new behavioral theories and experimental tools of inquiry to better understand how we think and behave. These two revolutions arose separately but in parallel over the last few decades, the first primarily in the halls of government, the second largely in the ivory towers of academia – although they have important reciprocal implications. The goal of this Element, therefore, is to explain how the performance revolution in governance can be informed by insights from the behavioral revolution in the social sciences. By doing so, the focus on real-world performance metrics reveals the usefulness of behavioral theories for understanding often messy and complicated questions in politics and administration.

The two revolutions motivating our Element appeared suddenly but had deeper roots, some of which were shared in common. Governments have always measured things, and research interest in human behavior in this area is long-standing. Take for example the career of Nobel Prize-winner Herbert Simon, who, as a graduate student, undertook a benchmark study of public-sector performance measurement. In 1939 he asked a basic set of questions for which we still do not fully have good answers: "What are statistics and how can they help solve administrative problems … what pitfalls are to be avoided?" By the time he finished his dissertation, Simon had migrated to the study of administrative behavior, observing how people used cognitive shortcuts to make imperfect decisions rather than analyze all possible outcomes. He won a Nobel Prize in Economics in 1978 for this work. But his twin interests as a graduate student – how human behavior and performance information jointly influence public-sector outcomes – remained largely unconnected in his lifetime, both by governments and his home field of public administration.

Following Simon's insights, we can see how governments produce more metrics than ever before but often do not fully grasp how best to use them. This mass production of numbers is built on the assumption that at some point in the policy process a group of human beings – politicians, civil servants, private contractors, think tanks or citizens – will sit down and make sense of it all. Performance information use is primarily a cognitive process – it requires people to choose pieces of information, interpret them and apply them in the form of a belief, decision or action. The problem is that we know very little about that process. Governments adopted performance reforms without a solid understanding of the micro-foundations underpinning how people make use of

performance metrics. The assumption was: if we build it, they will come (and use it). In contrast to the abundance of guidance available on the technical aspects of data analytics, relatively little support has been provided about the psychological and contextual aspects of performance information use.

This Element addresses the gap in knowledge using the rapidly developing literature on behavioral public performance. This literature is both theoretically grounded in ideas about human sensemaking and cognitive biases and methodologically rigorous in its empirical approach. The findings improve understanding of information use by citizens, managers and policymakers, supporting novel and important recommendations for the design and implementation of performance systems. These findings and recommendations contribute to broader debates about the potential and limitations of performance information use in public management and democratic accountability.

1.1 The Performance Measurement Revolution

Recent decades have seen a revolution in the measurement and reporting of government activities and their consequences, with burgeoning use of published metrics, report cards, league tables and rankings (Hood et al. 1999; Hatry 2006; Moynihan 2008; Van de Walle and Van Dooren 2008; Moynihan and Beazley 2016; Olsen 2017a). Such performance information is typically quantitative, capturing government inputs, processes, outputs and outcomes across a range of policy areas including the environment, the economy, health, education, public safety, and other public policies and services. The information often resides in analysis and reporting systems that enable comparisons of performance to be made over time or across organizations.

Performance reporting regimes are found in many jurisdictions. Table 1.1 shows some exemplars with their rationales. For example, the US Government Performance and Results Act and its successor created a performance framework intended to hold federal government agencies to account for their goals (Moynihan 2008). The UK government developed elaborate standardized systems for reporting the performance of state schools, hospitals and local government units in the 1990s. These systems were intended not only to improve accountability and inform local voting but also to help facilitate user choice and drive competition in quasi-markets (Hood et al. 1999; Boyne et al. 2009; James 2011a). Similarly, most local governments, cities and states in the USA have some form of metric-based reporting of performance to citizens, reflecting demands for more transparency (Piotrowski and Van Ryzin 2007; Grosso and Van Ryzin 2011).

Table 1.1 Examples of performance reporting systems.

Jurisdiction of systems	Illustrative examples
International metrics	Program for International Student Assessment run by the Organisation for Economic Co-operation and Development (over eighty jurisdictions); World Bank Governance Indicators about rule of law, corruption and government effectiveness (200 countries and territories); metrics for United Nations' Sustainable Development Goals
National government or government-wide	Government Performance and Results Act (GPRA) 1993 and GPRA Modernization Act 2010 in the US federal government; UK government public service agreements (1998–2010) and subsequent public service performance dashboards
City-based reporting	New York City Mayor's Management Report; London Mayor's Annual Report including statutory measures
Local public service provider metrics	No Child Left Behind for schools in the USA; Care Quality Commission and hospital metrics in England; systems of comparative performance measures (supplemented by inspection reports) for schools, hospitals, police and other local services in England

International measures and systems for comparison across countries have blossomed (Hansen, Olsen and Bech 2015; James and Petersen 2018). For the United Nations' Sustainable Development Goals, for example, a range of different indicators are used to track progress against targets reflecting seventeen goals, including an end to poverty and zero hunger. Comparisons help managers and policymakers benchmark performance and look for innovation from high-performing jurisdictions, and they potentially generate public pressure for improvement. If government can show citizens that it is doing a good job, then trust in government and support for public activity may increase, addressing concern about the decline in trust evident across several countries in recent decades (Hetherington 2005; James and Van Ryzin 2017a).

1.2 The Behavioral Science Revolution

Recent decades have witnessed significant advancements in the behavioral sciences. By the time Herbert Simon passed away in 2001, a new generation was building on his work. Insights from cognitive and social psychology, neuroscience and evolutionary psychology have come to show how social identities, motivations and emotions all shape human judgment and decision-making. This development includes the pioneering cognitive psychology of Daniel Kahneman and Amos Tversky, who were themselves deeply influenced by Simon. In their view, two general systems explain much about how heuristics and biases operate in the human mind (Kahneman 2011).

System 1 represents the fast, intuitive and automatic thought processes that we all depend on for much of our daily functioning and survival. Thought processes have evolved in the human brain, such as our natural responses to facial expressions or potential threats, or the deep learning acquired over a lifetime. These are automatic reactions, akin to the responses an experienced tennis player displays on the court in a fast-paced match. System 2 represents the slow, deliberate and effortful thinking that we engage in when we debate complex alternatives, find our way using a map, or attempt to solve a math or analytical problem. Because system 2 takes time and energy, occupying much of our mental capacity, we rely on it only when necessary. Thus, although the heuristics that Simon observed public managers using to deal with complex environments (which he termed satisficing) often serve as useful shortcuts to efficient decisions (Gigerenzer, Todd and ABC Research Group 2000 2001), they can also cause cognitive biases that lead to systematic errors.

Social psychology and neuroscience have also shown how personal identities and social relations affect people's reception and use of information, and how motivations can affect reasoning by promoting reliance on a biased set of

cognitive processes (Kunda 1990; Kruglanski 1996). Indeed, humans are profoundly social beings who evolved in cooperative groups and thus remain highly attuned and sensitive to what others do, think and say (Lieberman 2013). As emotional creatures, a system of constructed feelings often guides our thinking and behavior by predicting and anticipating our reactions to people and our environment (Barrett 2017). Thus, we often make decisions based on gut feelings – and much of the time we are right to do so (Gigerenzer 2007).

The 'behavioral' label as such comes from the application of the ideas and insights from modern psychology to economics, which as a field had traditionally assumed that human beings are rational, self-interested, utility-maximizing agents with strong mental powers of calculation – *Homo economicus*. But the work of Kahneman and Tversky, along with the work of Vernon Smith, Richard Thaler, Robert Schiller, George Akerlof and many others showed how cognitive biases, social identities and emotions influenced economic decisions and behaviors (Thaler 2015). Behavioral science emerged, with increasing influence not only on the field of economics but also finance, decision theory, business, legal studies, political science and in turn public management.

1.3 Connecting the Two Revolutions

Within Simon's home field of public administration, the behavioral revolution has grown in influence. Earlier research about performance measurement in government was dominated by mostly case-based studies of discrete reforms. However, Moynihan and Pandey (2010) argued that the best way to study performance management was to examine the broader question of performance information use as a behavioral phenomenon. Studies began to link performance information to political outcomes, such as its influence on electoral support for incumbent local governments (James and John 2007; Boyne et al. 2009), the relationship between performance reporting and trust in government (Grosso and Van Ryzin 2010), and the effects of published performance measures on citizens' perceptions of service performance and satisfaction (James 2011a).

A new subfield, behavioral public administration (BPA), arose (Olsen 2015a; Grimmelikhuijsen et al. 2017). This subfield draws on psychological and related theories and contributes to the understanding of how performance information is used by citizens and public managers, extending the range of questions addressed and increasing the breadth of theories informing research. BPA has improved the ability to make causal claims by using experimental designs. Experiments incorporate the manipulation of discrete factors, randomly assigned, combined with the measurement of outcomes, to allow stronger

inferences about causal connections than typically found in observational studies.[1] The presentation of performance information is especially suitable for experimentation because it can be relatively straightforwardly varied experimentally (James, Jilke and Van Ryzin 2017).

1.4 Key Questions and Contributions

The behavioral perspective set out in this Element considers how people process different types of performance information, presented in different ways and from different sources. We address important questions such as: Why do people use different kinds of performance information to make decisions? How do they do so? What are the biases they are vulnerable to? What are the broader implications for government and other users of performance information? How can governments improve the presentation and use of performance metrics?

Figure 1.1 summarizes three factors that together shape the use of public performance metrics and provide a roadmap for the Element. The first factor concerns how the nature of performance metrics, including different numerical forms, can influence how people process the information (Section 2). For example, people exhibit clear 'left digit' bias in their evaluation of metrics, such as school test scores. How information is presented or framed can also change how people respond to it (Section 3). Additionally, the presentation of comparisons over time or across government units, and use of different benchmarks, has consequences for the interpretation of performance metrics (Section 4).

Our second factor centers on how performance information use depends on the characteristics of the user, for example, their capacity to handle numeric data, their social and political identities, and their motivations for engaging with information (Section 5). People seek out and select performance information is ways that confirm their prior beliefs. Moreover, politicization of a public service (e.g., health care in the USA currently) makes this confirmation or disconfirmation bias worse. People exhibit motivated reasoning, interpreting metrics to fit with their political beliefs or party political identities. The source of information, such as government agencies or more independent sources, also influences its reception by people with different beliefs or identities (Section 6).

[1] Correlational studies have generated strong insights, especially for factors difficult to study using experimental designs. Kroll (2015) reviews performance information use to identify six "important drivers of use" from survey-based research designs. These are (a) measurement system maturity (a well-developed performance system with available data), (b) stakeholder involvement, (c) leadership support, (d) support capacity (investment of resources into the performance system), (e) innovative organizational culture and (f) goal clarity.

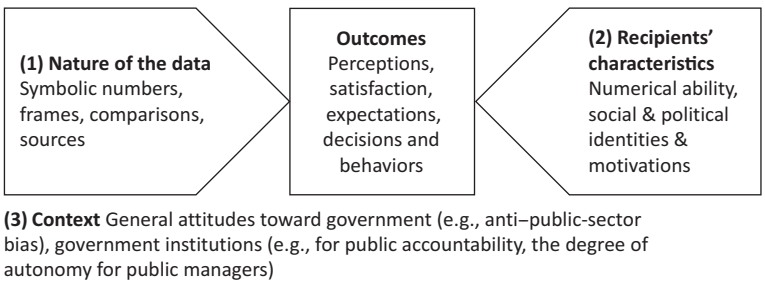

Figure 1.1 Behavioral processing of performance metrics.

The third factor highlights how social and political contexts shape the interpretation of performance information. Published metrics are part of a broader effort to improve the 'transparency' of government.[2] Performance information circulates among politicians, public managers, users of public services and the broader citizenry, and is reported by different media. General attitudes toward government in society can create a more or less receptive environment for the use of performance information. In some jurisdictions, anti-public sector bias (Section 7) makes it difficult to use reporting to boost the legitimacy of public activity. Institutions that structure government activity, especially for managerial and public accountability, also constitute important contextual factors. The extent of managerial autonomy, and the potential for deliberation about information, affects how managers interact with performance information (as discussed in Section 8). We conclude by setting out the implications of the evidence about behavioral public performance for the design and use of performance reporting and suggest an agenda for future research (Section 9).

2 What's in a (Performance) Number?

Performance measurement attempts to assign meaning to government efforts and accomplishments using numbers. This is not as straightforward as it might first appear. Numbers ultimately have to be interpreted and used by human beings. People are not intuitive statisticians, preferring instead to process numbers through their psychological prisms. This section examines this process by addressing the following questions: How are quantitative metrics viewed differently from qualitative forms of performance information? How do managers, politicians, policymakers and citizens make sense of various types of numerical performance information, and what is the role of nonnumerical cues in this process? How do individuals draw different inferences from abstract

[2] For a review of transparency research, see Cucciniello, Porumbescu and Grimmelikhuijsen (2017); we focus specifically on numerical aspects of reporting.

numbers as opposed to concrete cases or episodes, which tend to be more vivid and thus more emotionally engaging? How do symbolic numbers, in particular round numbers, serve as cues about performance and thus become readily available benchmarks that citizens rely on when judging metrics?

2.1 Statistical or Episodic Information: How Compelling Are Performance Metrics?

Interest in how people make use of performance metrics is long-standing, dating from at least the turn of the twentieth century (Woolpert 1940; Roher 1941; Merwin 1942). The hopes of US progressive reformers for how performance information could engage citizens were tempered with a concern about the limit of people's natural tolerance for numbers. Upson (1915, pp. 65–66) called for a "new method of making municipal government a concrete reality to the men on the street" (p. 65), but cautioned that "cumbersome reports filled with unintelligible and inaccurate statistics not only fail to arouse the citizen, but destroy potential interest." In the 1930s, Herbert Simon and Clarence Ridley advised cities that metrics needed to be attractive to citizens, noting: "Attractiveness refers to those qualities which have the power to draw attention and sustain interest for more than just a moment. In attracting attention, reports appeal to the eye and to the mind" (Ridley 1939, p. 48).

Governments turn to quantification as a means to communicate their commitment and efforts to produce tangible results, hoping that accountability engenders trust. Quantitative data implicitly convey a scientific cue to the receiver, whether citizen or policymaker, that is independent of the underlying validity and reliability of the actual performance data. As stressed by Jackson (2011, p. 24), "Numbers generated by sophisticated statistical techniques can give a false sense of objectivity and reliability." But the credibility of numbers does not necessarily make them attractive.

What sort of information is compelling to people? Across a number of fields, the answer is 'soft' episodic information that focuses on a single case or episode and portrays a topic with 'human interest'. The power of such information has been well established in communication research (Daschmann 2000; Zillmann 2006), political psychology (Gross 2008; Aarøe 2011) and social psychology (Pettus and Diener 1977; Herr, Kardes and Kim 1991; Jenni and Loewenstein 1997; Kogut and Ritov 2005). For many decisions, studies have found that 'image-provoking' episodic information exerts more persuasive effect than statistical information (Pettus and Diener 1977; Herr, Kardes and Kim 1991; Jenni and Loewenstein 1997; Kogut and Ritov 2005). The vividness of information focuses human attention and increases the weight of sensory information (Nisbett and Ross 1980). Vividness of performance metrics can be

enhanced by including specifics about actors, actions and context. For example, statistics about school graduation rates can be made more vivid by including a story of a student who, with help from her caring teachers, overcomes personal hardships and earns a diploma.

"People are more interested in people than in anything else. They prefer a rendition of the melody to a copy of the music score" (Freeman 1954, p. 124). To extend the analogy, a statistical table reporting performance can be seen as the music score, while a vivid media report about a citizen's experience with government is the more engaging, emotional melody. This tendency appears to be true even among professional managers who rely a great deal on verbal, ad hoc or qualitative feedback (Mintzberg 1973; Moynihan 2008; Kroll 2013). If even public managers who are legally tasked with paying attention to performance metrics focus on other types of information, it should come as little surprise that citizens do the same. For US citizens, one survey suggests that the three most frequent sources of performance information were local news, friends or neighbors, and public meetings or hearings (Grosso and Van Ryzin 2011), sources where we would expect episodic information to be more likely than statistical data (see also Slattery and Hakanen 1994).

In contrast to episodic cases, such as stories or anecdotes about public-service successes or failures, numerical performance information is perceived as more objective and authoritative (Herbst 1993). Citizens seem to understand this, with a stated preference for numbers but a revealed preference for other sorts of information. Figure 2.1 shows the results from an experiment where Danish citizens were given a choice between episodic and statistical performance information to help inform their decisions about a hypothetical stay at a hospital (Olsen 2017a). Respondents were split into three groups and asked to choose either a personal story or client satisfaction statistics. Regardless of the particular case, respondents overwhelmingly said they preferred the statistical material.

The subjects were next presented with five different vignettes about poor treatment at a hospital. Three of them presented cases of episodic performance at the individual level, either 'A patient', '51-year-old Birgitte' or '31-year-old Erik', who were in great pain after a hospital surgery. Two other groups were provided with vignettes featuring simple statistical performance: In the first, they were told that 1 in 100 patients were in great pain after their surgery. In the second, the rate of pain due to poor treatment was 10 out of 100 patients, or 10 percent. Subjects were then asked to rate the quality of the hospital (see Figure 2.2). All three episodic reports of performance caused significantly lower evaluations than the statistical description of a 1 percent error rate. Informational vividness causes us to give greater weight to case-based descriptions than statistical evidence when evaluating a public organization. Only

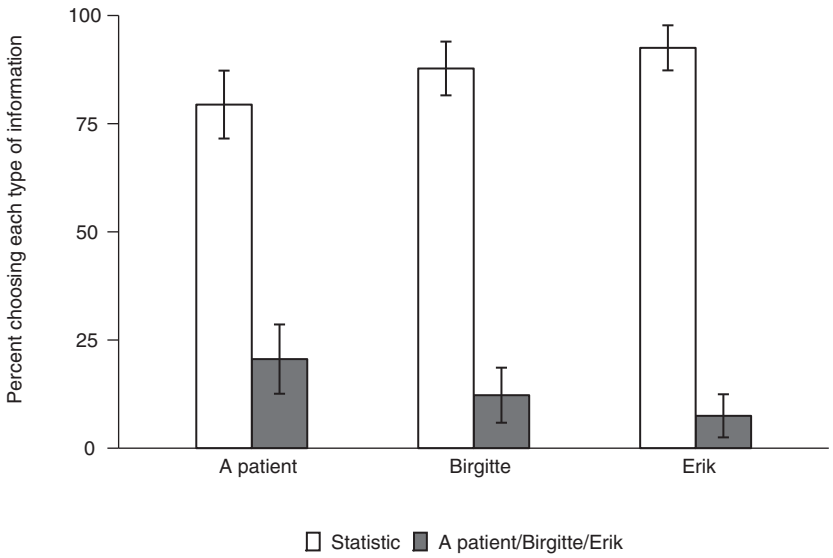

Figure 2.1 Citizens say they prefer statistical over episodic information.

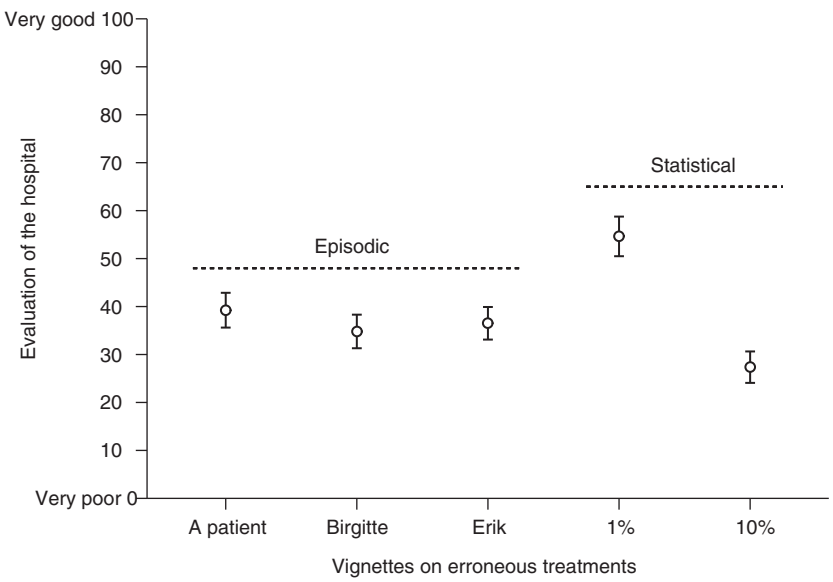

Figure 2.2 Episodic information is more influential than equivalent statistical evidence.

when the statistical information describes very poor performance (a 10 percent error rate) does the statistical information have a greater influence than the anecdotes.