Algorithms and vulnerable citizens: The cost of Australia's experiment with automation in the governance of its social welfare system

Jasper Brown
MA Public Policy and Management
Class of 2020
The University of Melbourne

Photo by Mazhar Zandsalimi on Unsplash
Photo cropped and modified to black and white
INTRODUCTION

Australia is often cited as one of the most advanced nations in terms of digitisation and the use of innovations in information and communication technology (Djeffal, 2018). However, after employing its Online Compliance Intervention (OCI) system in July 2016, The Department of Human Services (DHS) found itself at the centre of a debate about the problematic application of automated decision-making systems. Colloquially described as the ‘Centrelink robo-debt’ program, the OCI system used an automated algorithm to investigate and initiate the department’s debt-recovery procedures (Carney, 2018). Executing a Coalition reform commitment titled, “Better Management of the Social Welfare System” (Senate Community Affairs References Committee, 2017, p.8), the measure aimed to repair the Budget by recovering more than $2 billion of estimated welfare overpayments from current and former Centrelink recipients for a period that stretched back over several years (Dunleavy & Evans, 2019, p.16).

The OCI was implemented to improve Australia’s social welfare system, premised on criticisms that existing governance arrangements were “extremely complex, inflexible, costly to maintain and difficult to ensure compliance” (cited in Arthur, 2015). These governance arrangements were the result of a decades-long legislative agenda, which enabled the department to monitor and regulate user compliance to prevent social security fraud and prescribed the necessary requirements for pursuing recoverable debts. For approximately twenty years, the DHS used its information gathering powers under the Social Security Act 1991 to conduct data-matching activities, which would identify discrepancies in the welfare recipient’s reported income to the Australian Tax Office (ATO) and their reported income to Centrelink (Glenn, 2017). Tasked to determine whether there had been an overpayment, DHS compliance officers were
allowed to use their statutory powers under the Social Security (Administration) Act 1999 to investigate these discrepancies by compelling banks and employers to supply wage records (ibid.). While under the OCI system, the data-matching activities ultimately stayed the same. Rather than triggering enquiries, the OCI’s algorithm extrapolated data discrepancies to calculate an overpayment value and subsequently issued debt notices to alleged debtors (Carney, 2018).

By no longer exercising the investigative powers enabled under the social security legislation, the DHS significantly accelerated the process of identifying welfare overpayments and therefore increased the number of overpayments that were identified each year (Zalnieriute, Bennett Moses, Boughey, Burton Crawford & Bhathela, 2019). According to DHS data, the scale of its regulatory activities increased from 20,000 compliance interventions per year to over 780,000 interventions in the 2016-17 financial year (Glenn, 2017).

While the OCI system dramatically increased the scale of compliance activities, it came at a significant cost. Improvements to the cost-efficiency of the department’s compliance activities were quickly eroded by reports that exposed the disproportionate impact the OCI system’s outputs had on Australia’s most vulnerable citizens. The DHS was confronted with senate inquiries, a Commonwealth Ombudsman investigation and a review by the Australian National Audit Office, all calling into question the integrity of the automated debt-recovery scheme. In addition, NGOs and legal academics also extensively criticised the significant design flaws in the OCI’s governing architecture (Dunleavy & Evans, 2019, p.16). The first inquiry by Commonwealth Parliament Senate Community Affairs References Committee concluded that the OCI “lacked procedural fairness at every stage and had put thousands of innocent Australians through the trauma of having to prove they
do not owe the money the welfare agency demanded” (ibid., p.16; Zalnieriute et al., 2019).

Across the board, the DHS was critically for launching an algorithmic system with inaccuracies in its formula. There was a lack of transparency in debt correspondence, punitive debt collection strategies and debt notices being issued where no debt existed (ibid.; Zalnieriute et al., 2019, p.3). The error-rate of the OCI system was calculated to be approximately 20 percent, with 445,000 debts being raised through the system between July 2016 and December 2018, 77,500 of which were either reduced, waived or written off altogether (Zalnieriute et al., 2019, p.29). It was on these grounds that the Senate Committee inquiry recommended that the system be put on hold (Senate Community Affairs References Committee, 2017, p.108). The OCI system “reversed the onus of proof” as it placed the responsibility to obtain such information to disprove the existence of debt onto the welfare recipient (ibid., p.9; Carney, 2018, p.41). These findings have also formed the basis of several legal challenges, including one by Victoria Legal Aid and a class action that will be launched against the Government by Gordon Legal and backed by the Labor opposition.

Despite attempts by DHS to reform certain elements of the OCI’s design, the system remains firmly governed by the same logic that triggered the initial integrity crisis. A second parliamentary investigation into the department’s methods for identifying and recovering alleged debts using income-matching software is currently underway (Henriques-Gomes, 2019). Efforts to reform the department’s compliance activities must unpack how a system initially designed as a measure to strengthen the integrity of the welfare payments system ultimately weakened it. The case reveals accountability and legitimacy challenges that stem from a regulatory system that effectively removes voice in favour of compliance. Disentangling the unique ways in which the OCI system fundamentally altered the bureaucratic
governance environment will reveal the governance costs for achieving augmentations in efficiency through algorithmic and automated means. This will unpack the cost-benefit calculus upon which the decision to deploy algorithmic decision-making systems are made. Through excavating the literature on governance and public administration, this paper will draw a roadmap for navigating the challenges posed by the application of an automated algorithm to identify, investigate and initiate the department’s debt-recovery procedures in the social security system.

OVERVIEW

Advocates for the rapid adoption of algorithmic forms of governance in public administration present it as a remedy for a range of public policy problems (O’Reilly, 2013; Djeffal, 2018). However, there is a growing body of governance literature critically examining algorithmic power and the deployment of automated data-driven systems to inform public administration and decision making (Yeung, 2018; Hughes, 2018; Gillespie, 2014). While algorithmic governing presents some new challenges, improving the machinery of government has always been a pursuit of academics and policymakers. The extent to which new models of governance have fundamentally altered the practices, processes and procedures of governments over the past five decades have been thoroughly explored (see: Erkkilaä, 2007; Aucoin, 1990; Conside & Lewis, 2003). As liberal democratic nations continuously embrace innovations in technology, “all seeking to harness their tremendous potential to enhance the quality and efficiency of many activities” (Yeung, 2018, p.505), it is useful to develop an understanding of the historical underpinnings for pursuing new forms of governance and its impact on traditional and well established tenants of good governance.
This paper will disentangle the unique properties of algorithmic governance models from their historical origins to identify the destabilising consequences of governing or regulating by an algorithm. To isolate the ways in which algorithmic governance changes the quality of bureaucracy, the analysis will consider the administrative and policy contexts in which algorithms exist. This includes the people who design, implement and interact with algorithmic governance, meaning “the data and users upon which they act and the institutions that provide these services” (Gillespie, 2014). Empirical evidence suggests that the deployment of algorithmic governance mechanisms can have disastrous consequences for the legitimacy of public services and can compromise existing control mechanisms that regulate the relationship between the citizen and the state (Galloway, 2007, p.89). Finally, this paper will identify a repertoire of strategies for controlling these challenges when improvements to governance performance are pursued through algorithmic means.

**ANALYSIS OF INSIGHTS AND LESSONS**

The impact of administrative reform extends beyond changes to institutional design and has the potential to have a substantial and ongoing impact on administrative culture (Erkkilää, 2007; Stark, 2002). The administrative culture that gave rise to algorithmic forms of governance and administration is firmly rooted in the performance-based approach that focuses on outputs and results, rather than on “governmental processes and activities” (Yeung, 2018, p.510). Performance-based ideologies propose that removing the procedural constraints on productive management will enable improvement in public sector administration (Aucoin, 1990). This reveals that the challenges posed by the practicalities of governing by an algorithm are not new. They are deeply rooted within the ideological turf war competing for a dominant conceptualisation of good governance.
The Australian Government’s drive to achieve an increase in performance and outputs, while simultaneously reducing the size of public administration, can be traced to the enactment of New Public Management (NPM). The NPM movement was a wave of reforms that aimed to decrease the size of governments and fundamentally changed the Australian governance landscape in the late 1980s (Mulgan & Ure, 2001; Mulgan, 2000; Aucoin, 1990, p.122). These reforms included efforts to privatise government initiatives, systematically de-regulate government agencies and contract out various public services to private sector providers (Aucoin, 1990). Driven by a commitment to market principles, these reforms strongly encouraged performance measurement and empowered managers as agents of efficiency and accountability (Stark, 2002; Considine & Lewis, 2003). Generally, the new administrative culture favoured replacing centralised and hierarchical control mechanisms as a method of promoting greater efficiency and achieving improved productivity and effectiveness within governments (Aucoin, 1990, p.132; Erkkilaä, 2007). However, as scholars have noted, when established models of government are replaced with new institutions and methodologies, traditional mechanisms for accountability are put at risk (Erkkilaä, 2007; Deleon, 1998).

The emergence of algorithmic governance systems to inform public administration, decision making and regulation can be understood as a market governance solution to the problem of bureaucracy. At its core, the adoption of algorithmic techniques in public administration rests on the superiority of outcomes-focused governance. This suggests that when public management is freed from the time consuming and resource-intensive practices of governing by processes, substantive results in efficiency, economy and effectiveness will be achieved. The move away from complex governance practices is distinctly related to the market governance mentalities that drove the NPM reforms. This latest episode in the struggle against the “bureaucratic order
of things” (Considine & Lewis, 2003, p.131) reveals important challenges in how governments are dealing with advancements in technology and how transplantable they are within public sector contexts. To contextualise the unique impact of automated governance systems, an investigation is required to go beyond the normative assumptions that underpin performance-based governance practices in public administration. The cost-benefit calculus upon which decisions to automate governance activities are made must also be considered. Specifically, what kinds of public services would benefit from cost-efficiency augmentations and how the automation of the processes and procedures of governance in each context fundamentally alters the bureaucratic environment in which they operate.

Deploying algorithmic systems in the context of vulnerable populations, as in the Australian robo-debt case, has particular social, political and legal ramifications. As Coglianese and Lehr (2016) highlight in their analysis of automated decision making systems in the judiciary, regulatory, social welfare systems, figuring out how “regulating by a robot or adjudicating by an algorithm” (p.85) can be situated within the established legal and bureaucratic practices and processes is yet to be solved. The empirical evidence demonstrates that the failure to do so has had devastating consequences on the populations that these systems are designed to serve (Desai & Kroll, 2017; Coglianese & Lehr, 2019). When they are applied from within as social welfare, healthcare, job services, employment and human resourcing, sentencing and national security systems, there is an increased likelihood they will interact with more vulnerable populations. In these contexts, fidelity of implementation becomes even more critical, as when algorithmic governance systems are poorly implemented, they can negatively impact human rights. Most notably, they impact procedural rights, privacy and the right to be free from discrimination (Desai & Krill 2017; Zalnieriute et. al. 2019; Coglianese & Lehr, 2016). Many pioneering
legal scholars have demonstrated that the implementation of algorithmic and automated systems of decision-making often compromise the procedural fairness rights of individuals and undermine the capabilities of citizens to “influence or challenge a decision affecting them” (Zalnieriute et al., 2019, p.31; Carney, 2018; Galloway, 2017). Craig (2012) notes procedural fairness is “intended to ensure that those affected by the decisions of public officials should be given an opportunity to participate in decisions that affect them, and to ensure that such decisions are made in an unbiased manner” (cited in Yeung, 2018, p.515).

Procedural fairness is an ensemble of bureaucratic procedures that enables citizens to exercise various voice mechanisms that safeguard the fair and just determination of decisions about matters affecting them (Devarajan & Reinikka, 2004). From a governance standpoint, it acts as a critical control mechanism that regulates the relationship between citizens and the state. This is important as citizens engaging in various forms of social security services often have no alternative if they are dissatisfied (Paul, 1992). As governance scholars note, when there are limited or no avenues for citizens to exercise choice, structured opportunities to exercise voice become a crucial part of the integrity architecture (Paul, 1992; Craig, 2012; Geoghegan, Lever & McGimpsey, 2004). The consequences of either poorly automating procedural fairness measures, or bypassing them altogether, is particularly destabilising for traditional standards of good governance. Consequently, it is always the most vulnerable citizens who are disproportionately impacted by the reduction or circumvention of procedural fairness processes (Zalnieriute, 2019; Carney, 2018; Coglianese & Lehr, 2017).

According to emerging literature, altering the bureaucratic safeguards that regulate the power relationship between citizens and the public service has important implications for the transparency of governance activities (Galloway, 2017). The systemic challenge to transparency relates to the risk of
algorithms obfuscating hidden biases or errors in decision-making processes (Yeung, 2018, p.515). It is widely understood that the automation of government decision-making through the adoption of rules-based systems, such as an algorithm, poses “potential benefits in enhancing transparency and accountability” (Zalnieriute et al., 2019, p.6). In theory, an algorithmic decision can be traced back to its derivatives (Gillespie, 2014). In other words, the outputs can be reverse-engineered as an aggregate of the inputs. However, governance literature is yet to point to an example of an algorithmic governance system that maintains sufficient transparency to accurately describe it as an example of good governance. The Office of the Victorian Information Commissioner (OVIC) (2019) released a report into the technical, social and legal aspects of artificial intelligence (AI) in August 2019 that echoed these wider concerns, warning that;

“[h]umans can justify and explain how they reached a decision. AI systems, on the other hand, make decisions based on statistical predictions from algorithms that are usually complex to the point of being incomprehensible to humans. This is a critical issue for government use of AI — government is expected to be transparent, and if the rationale behind an AI decision cannot be explained, that decision can hardly be described as transparent.” (p.58)

In effect, algorithmic governance systems often become black boxes when the data inputted is known and the decisions derived from the data are visible, but the ways in which data is used to inform decisions is not transparent (OVIC, 2019, p.3). As Yeung (2018) describes, when algorithmic governance systems are designed to expedite procedural fairness activities in social security contexts, they essentially operate as a “one-way mirror” (p.517) that enables governments to better track, monitor and regulate its citizens while disabling the public’s capacity to return a critical gaze (Yeung, 2018, p.517). To this effect, algorithmic governance risks transforming the decision-making apparatus in
public administration into algorithmic black boxes that sever the democratic relationship between the bureaucracy and the public.

Thus, where decisions are fully or partially automated in social protection systems, the extent to which its outputs comply with accountability and transparency standards relies on the people designing the system. They must adequately incorporate the existing control mechanisms that safeguard an open and transparent evaluation of the inputs (Zalnieriute et al., 2019, p.23). This is not to argue that the principles of transparency and the practice of bureaucratic control mechanisms should not eradicate the pursuit of cost-benefit and efficiency values. Rather, as Yeung (2018) argues, “these value trade-offs should be openly debated rather than simply resolved in favour of efficiency by technological fiat” (p.517). To better understand how the trade-offs between performance and efficiency, and accountability and transparency, can be negotiated, it is important to interrogate citizen perceptions of the governance activities that algorithmic systems are designed to improve. Investigating the literature on democratic legitimacy, and more specifically, how democratic legitimacy can be achieved, will help define a set of criteria upon which algorithmic governance systems can be designed and assessed.

In the early 2000s, governance scholars began to critically evaluate the democratic legitimacy of governance activities (Boedeltje and Cornips, 2004). As Mena and Palazzo (2012) argue, the democratic legitimacy of a governance system, regulatory body or policy initiative can be understood as the “socially shared belief” (p.528) that the public administration holds the necessary capabilities and the authority to “impose rules on a community of citizens” (p.528). The legitimacy of governance activity can be measured by two key indicators. Firstly, input legitimacy is measured by the extent to which governance activities are perceived as justified or credible (Boedeltje & Cornips, 2004). Secondly, output legitimacy is measured by the
extent to which the governance activity effectively solves the issues that they target (Mena & Palazzo, 2012). In the regulation of welfare services, governance scholars point to a delicate social exchange between citizens or clients and the service providers that ultimately generates both the output and input legitimacy of regulatory activities (Alford, 2002).

For governance systems concerned with regulating citizens, the quality of the input legitimacy is evaluated by the design of political processes and how the rules citizens are supposed to comply with were developed (Mena & Palazzo, 2012, p.528). Therefore, the output legitimacy of regulation and compliance activities is evaluated on the grounds of its effectiveness, or how effectively regulations solve any particular problem with fiscal due diligence and user compliance (Mena & Palazzo, 2012; Boedeltje & Cornips, 2004). For departments and agencies to receive sufficient output legitimacy for their regulatory policies and procedures, they need to be seen to apply their coercive powers as far as possible. Importantly, for input legitimacy to be generated at an equal level, regulatory activities need to be regarded as “fair and just” (Alford, 2002, p.343). The challenges to the output legitimacy of algorithmic forms of governance are not so much the act of monitoring citizen behaviour in social security settings, but the methodology employed to undertake the particular governance activity (ibid.). The precise way in which regulatory activities are undertaken to achieve outputs, or as Alford (ibid.) states, in the case of welfare regulation, to “elicit contributions from clients such as cooperation or compliance” (p.343), can significantly augment or diminish the legitimacy and integrity of government.
TOWARDS A STRATEGY FOR INTEGRITY IN ALGORITHMIC GOVERNANCE

The current focus on outputs and performance, rather than on governance processes for achieving the desired results, reveals that algorithmic governance activities will need to reconcile their advancements in performance capabilities with their impact on the established bureaucracy. Failing to do so will perpetuate deficiencies in both input and output legitimacy and will ultimately undermine its long-term democratic integrity. Bringing citizen perceptions into the analysis will help offset the central tenants of performance and efficiency from dominating the cost-benefit calculus (Hughes, 2018).

Algorithmic governance systems cannot fundamentally alter the status quo in the social exchange, which ultimately defines the terms and conditions for a functioning and transparent relationship between the state and service provider, and the citizen and user. When efficiency is pursued at the expense of fairness, input and output legitimacy is compromised (Alford, 2002). While not necessarily guaranteeing user compliance, algorithmic governance that successfully integrates criteria for strengthening both the input and output legitimacy may strengthen the overall integrity of the service.

The ways in which market governance mentalities and reforms fundamentally altered the nature of public administration can provide insight into changes related to algorithmic governance. Like the market governance project before it, a practical solution to the emerging challenges posed by automating governance can be solved through closer regulation and oversight. As Olsen (2006) argues, “the claim there is too little bureaucracy is as relevant as the criticism of excessive bureaucracy” (p.12). It stands to reason that perhaps governments should direct their attention towards the capacity of existing and traditional mechanisms of accountability, voice and control and how they
can be introduced within new models of governance (Erkkilaä, 2007). This could involve finding ways to incorporate procedural fairness into the automated system to ensure that there is a “systematic and easily-accessible method for citizens to exercise such rights as and when required” (Zalnieriute et al., 2019, p.9). This could involve finding ways to incorporate procedural fairness into the automated system to ensure that citizens can exercise their right to “influence or challenge a decision affecting them” (Zalnieriute et al. 2019, p.31; Carney, 2018; Galloway, 2017). Building bureaucracy back into algorithmic systems may well act as the necessary protective safety-net between pre-algorithmic practices that pre-date or stand in opposition to the primary functions and purposes of algorithmic processes (Gillespie, 2014). Traditional bureaucratic architecture could reinstall accountability processes and procedural fairness for decisions made by such automated systems. In practice, such measures could include a process which involves a human evaluation of algorithmic outputs, clear guidelines on the potential for error and the circumstances in which error can arise, and a transparent and “sufficiently resourced process for appeals” (Zalnieriute et al., 2019, p.8).

Reconstituting bureaucracy into algorithmic systems through human oversight does not guarantee procedural fairness and transparency. Human oversight might successfully generate some levels of output legitimacy by reinserting a hierarchical control system of accountability. However, simply inserting a human into the procedure does not necessarily resolve the normative issues associated with regulating by a machine. Yeung (2018) notes that decisions made by algorithmic systems constitute a “powerful form of choice architecture”, moulding user behaviours and perspectives (p.516). Relatedly, Carr (2015) highlights the danger of “automation bias”, referring to the propensity of people to defer to computational judgments even when they can recognize that the situation calls for another choice. Building greater levels of
bureaucracy around the algorithmic governance system is not enough to tame the transformative power algorithmic decision-making processes can yield. However, constructing mechanisms to restore democratic legitimacy might control it. Sufficient efforts must be made to transform the algorithmic black box into a transparent system open to systematic scrutiny, evaluation and optimisation.

One of the recommendations from the Senate Community Affairs References Committee report into the OCI system (2017) has perhaps been overlooked in developing a solution to the challenges posed by the deployment of an algorithmic system in Australia’s welfare system. The inquiry found that in the drafting of the policy there was a clear “lack of consultation with key stakeholders who could give feedback on the potential impact to vulnerable Australians” (Senate Inquiry, 2017). Like bureaucracies, these algorithms need not just human oversight, but robust democratic control procedures and citizen engagement to ensure its fair and just application (Hughes, 2018).

As algorithmic governance techniques are being implemented, finding ways to improve democratic oversight through citizen participation could help solve the integrity challenge posed by its application within the public sector (Hughes, 2018). Enlisting citizen participation in the consultation, design, testing and deployment stages of an algorithmic governance project may restore the power imbalances. This could help to transform the relationship between the citizen and the algorithmic governance system from one where citizens are mere data points to be controlled, to active agents participating in the reform of government (Hildebrandt, 2016). This will require government departments and agencies to take a “proactive approach” (OVIC, 2019 p.34) to ensure affected communities are engaged in the design process (ibid., p.31). This could deepen the system designers’ understanding of the problem that the particular
algorithmic system is aiming to solve, the data on which the algorithm governs and the social impact of its decisions (ibid.). Considering the heavy cost to both the input and output legitimacy of pursuing efficiency and performance in isolation, initiatives to improve the overall integrity of algorithmic governance should engage their target citizens in the design and ongoing optimisation processes. This may transform the algorithmic governance system from one of a black box into a two-way relationship.
REFERENCES


