

The Challenge of AI Governance for Public Organizations

Justin B. Bullock
Associate Professor
Independent Researcher

Hsini Huang
Associate Professor
Graduate Institute of Public Affairs and Department of Political Science
National Taiwan University

Kyoung-Cheol Kim
PhD Student
Department of Public Administration and Policy
School of Public and International Affairs
University of Georgia

Matthew M. Young
Assistant Professor of Public Administration
Faculty of Governance and Global Affairs
Leiden University

Abstract

In this chapter we examine the use of artificial intelligence (AI) systems by public organizations and the resulting critical challenges of dehumanization and loss of human control. We identify these core challenges to modern public organizations and modern governance through the lens of Max Weber's *ideal type bureaucracy* and Herbert Simon's *administrative behavior*. Furthermore, our analysis reveals how understanding and adjusting to the interrelated dynamics of bounded rationality, hierarchy, facts, specialization, and human judgment can help public organizations maintain AI while preserving and reasserting human dignity and control.

Keywords: Artificial Intelligence, Bureaucracy, Organizations, Governance, Dehumanization, Control, Machine

Introduction

Government and public organizations have already begun to rely extensively on artificial intelligence (AI) systems to provide public services (Young et al. 2019), though adoption has been uneven (Bullock et al. 2020). Meanwhile, though many public service providers have coevolved along with these assistive technologies (Bovens & Zouridis 2002), their evolution has varied across both agency and type of task (Busch & Henriksen 2018). In the midst of this real-time process, which is birthing new forms of public organizations, we argue that it is useful to revisit their classic, complementary formulations, to examine how this co-evolution with technology challenges and otherwise affects questions of modern governance generally, and AI governance in particular.

To this end, we return to Max Weber's concept of *ideal type bureaucracy* (Weber 2019) and Herbert Simon's description of *administrative behavior* (Simon 2013). Given the increasing capability of AI to act as agents within public organizations (Bullock 2019; Bullock & Kim 2020), we maintain that Weber's characterization introduces three challenges, including: 1) a decreased scope of tasks for human bureaucrats, 2) a loss in human managerial control, and 3) an evolving bureaucracy structure. The framework provided by Simon's version highlights several problems as well, including: 1) a reshaping of the bounded rationality of agents, (2) value alignment of public organizations, and (3) a changing balance of formal and informal communication.

Finally, we argue that considering AI challenges through these classic views reveals two very troubling trends for public organizations and their behavior: increases in dehumanization and losses in human control. At the same time, reflecting upon Weber and Simon can help guide the field through these two dangers, which were core concerns to both writers. Bounded rationality of agents, hierarchy, delineation of facts and values, specialization, and human judgment are offered as tools by which modern public organizations can reassert human dignity and human control.

This chapter will first introduce the nature of AI-related risks in the public sector. From there, we review Weber and Simon's theoretical approaches to understanding administrative functioning, structure, and behaviors in relation to the use of AI. Finally, in response to the pervasive spread of AI across the decision making functions of many organizations and governing systems, we propose strategies to address these threats to human influence and control.

Artificial Intelligence and Public Organizations

Machine learning-based AI technology has rapidly advanced over the past two decades. As its ability to successfully optimize decision functions under complex scenarios increases, some forecast that AI credibly threatens to displace “white collar” labor, including professional roles found in public organizations. In the U.S. context, early signalers of intended AI adoption include the U.S. General Services Administration (GSA), which plans to use AI for rule setting at the institutional level (Farber et al. 2021). The U.S. Department of Defense is developing a similar rule-setting application, the Joint All-Domain Command and Control (JADC2), and is also designing and conducting trials for autonomous or “self-flying/navigating” airplanes and ships (Hambling 2021; Dempsey 2021). In most cases, the stated motivation behind the push for adopting AI is to increase the organization’s efficiency.

These systems, however, are not planned, adopted, or implemented in a vacuum, but, rather, exist within larger organizational contexts and frameworks. Public organizations operate under particular pressures to optimize along multiple value dimensions, some of which, such as transparency, accountability, and equity, often require tradeoffs against efficiency. As a technical artifact, AI’s behavior is conditioned by its inner environment and interfaces (Simon 1997), leading to poorly understood risks to the entire decision-making structure. In most organizations, and virtually all public ones, this structure is commonly understood to be bureaucratic.

Bureaucracies structure authority to act across clear vertical and horizontal lines of command, and employ specialists whose jobs consist of purpose-clustered tasks and responsibilities (Weber 1922; Waters & Waters 2015; Simon 1946, 1997). Rational legal authority is thus derived from hierarchical relations between agents, with individual agents responsible for explicitly defined, specialized sets of tasks. This structure has its limits and tradeoffs. On one hand, the professionalization of bureaucrats—those dedicated specialists protected from political threats that Weber called *Beamte*—demands standardization, rationality, control, and administrative behavior that reflects the hierarchical order and will of the authority. On the other hand, formal rules, laws, and delegation structures are designed in a way that requires organizational agents to practice some level of discretion (Rourke 1972).

Indeed, the scope and complexity of tasks necessary to solve collective-action problems requires delegation of work to agents, who in turn must exercise discretion in deciding how to perform them (Arrow 1984; Bertelli 2012; Huber & Shipan 2002). Prior work on public sector innovation via information and communication technology (ICT) has produced three related theories tracking ICT’s effect on administrative discretion: systems-level bureaucracy (Bovens & Zouridis 2002), digital discretion (Busch & Henriksen 2018; Busch & Eikebrokk 2019), and artificial discretion (Young, Bullock, & Lecy 2019). In this framework, artificial discretion extends the general concept of digital discretion to describe machine learning-based AI.

AI agents are designed to practice discretion in decision making, either alone (i.e., the task is fully automated using AI) or as a supporting player (i.e., as a second set of eyes, akin to a human peer or collaborator). Take, for example, the use of AI for facial recognition. The desired outcome of this discrete task is a determination of whether a given facial image can be reliably paired to a separate image with associated metadata about the individual (e.g., name, age, known residence, etc.). The AI-as-agent ultimately produces a decision: it reports a credible match (or a set of possible matches) or reports that it has failed to do so. On a sub-task level, this process involves comparing n pre-known images of faces against the new image; at an even lower level, the AI agent is performing k subroutines, evaluating and comparing the geometries, shade gradients, and other constituent elements of the images.

This process, however, is not “rule bound.” The AI agent’s final decision cannot be deterministically reverse engineered from the set of input factors: many of the most ‘sophisticated’ and popular machine-learning (ML) architectures (e.g., artificial neural networks) involve stochastic processes, just as with a human agent. While it is possible to articulate and enumerate certain steps, subject to the limitations of time and physics, any given agent’s final decision output is probabilistic, not deterministic. As such, from an organizational perspective, AI agents are afforded discretion for this task as any human agent: they can be trained, and held to performance standards, but the outcome of any discrete task is not perfectly knowable *a priori*. In this way, these AI systems are acting as artificial bureaucrats empowered with artificial discretion (Young et al. 2021; Bullock & Kim 2020).

As with human bureaucrats, this power carries an implicit risk of error. AI has been shown to fail to generate precise outcomes, especially in novel circumstances (Bellotti 2021). Other work has found that AI cannot perfectly secure and process data and information while simultaneously developing and adjusting its protocol, resulting in poor decision making (Bullock and Kim 2020). Autonomous vehicles, reliable in repetitive, predictable situations, may judge and react sub-optimally when faced with unexpected road, weather, or pedestrian conditions. Minimizing this risk can occur at the individual agent level (upgrading one ML-based system to a higher performer), or at the institutional level (structurally changing whether and how to integrate the AI into the organization’s decision-making and task-executing processes).

These two distinct opportunities for managing AI error (individual vs. institution) introduce a tradeoff to scholars selecting their unit of analysis for studying the benefits and risks of AI in public organizations. Our goal for this chapter is to offer the reader a framework for leveraging two classic perspectives on administration that emphasize each analytic unit in their own way. In the following section, we synthesize the work of Weber and Simon to provide a construct for considering both the systemic and discrete risks and opportunities of AI for public organizations.

Theoretical Lenses: Weber & Simon

Max Weber and Herbert Simon's influential descriptions of bureaucratic functioning and administrative behavior can lay the groundwork for understanding AI's challenges to government. Importantly, we do not see these as competing descriptions of organizational behavior, but as distinct characterizations driven by differing goals and emphases, producing different frames, language, and concepts by which to evaluate their use of AI.

Weber sought to describe the design of a formal, rational structure to produce a controlled, standardized public organization, reliant on professional agents to adapt the delivery of public services to the needs of any particular situation. He also speculated on the troubling parallels between his description of bureaucratic structure and that of machine efficiency. Weber (2019) identified six guiding features of the ideal type bureaucracy:

1. Well-defined and organized competencies
2. A hierarchical organizational structure, with codified channels for making and transmitting information and decisions
3. Modern administration is based on a) documents preserved as original copies or concepts, and b) a staff of subordinated *Beamte* and writers of all kinds. *Beamte* working in the office, together with relevant resources including material goods and documents, constitute a bureau.
4. The work of the *Beamte* usually requires significant training for the specific tasks to be accomplished.
5. A full-fledged position occupies all the professional energy of the *Beamte* to process its tasks, regardless of limits to his mandatory working hours.
6. The duties of the position undertaken by the *Beamte* are based on general learnable rules and regulations, which are more or less firm and more or less comprehensible. The knowledge of these rules and regulations thus constitutes a special kind of 'applied science,' which the *Beamte* possesses.

Simon, on the other hand, started from the perspective of the individual decision, individual decision maker, and the general challenges of organizing information, communication, and general problem solving for effective and efficient organizational behavior. Simon focused on information flows across an organization, and how to arrange these flows so that the decision maker with the best information, most representative of the values of the organization, addresses a given problem (Simon 2013). Simon also believed that this design could benefit types of artificial systems beyond organizations. His conditions for a curriculum of social design, that is, an identification of the core features that influence social behavior (in large part through organizations), rely on six guiding ideas:

1. *Bounded rationality*. The meaning of rationality in situations where the complexity of the environment is immensely greater than the computational powers of the adaptive system.
2. *Data for planning*. Methods of forecasting the use of prediction and feedback in control.
3. *Identifying the client*. Professional-client relations, society as the client, the client as player in a game.
4. *Organizations in social design*. Not only is social design carried out mainly by people working in organizations, but an important goal of the design is to fashion and change social organizations in particular.
5. *Time and space horizons*. The discounting of time, defining progress, managing attention.
6. *Designing without final goals*. Designing for future flexibility, design activity as goal, designing an evolving system. (Simon 1996, 166)

It should further be noted that both Weber and Simon were concerned with the roles that both machines and information communication technology use within human organizations, particularly those charged with providing the public good. For example, Weber (1978, 979) argues:

The idea that there can be a law without loopholes is generally strongly contested. Also, the idea that the modern magistrate is nothing more than a ‘judging machine’ is rejected with disgust. In such a judging machine, files and costs would be thrown into the top of the machine in order that it would spit out the verdict along with a mechanical reasoning at the bottom . . . In fact, though, as within the domain of the findings of justice, there are areas where the bureaucratic judge is instructed by the legislature to use ‘individual’ paths to find justice. Moreover, in the area of the actual administrative functions, including all government activities that do not belong to the area of creating or implementing the law, or the finding of justice, the freedom and domination (Herrschaft) of individual approaches to tasks are simply taken for granted. In contrast to such emphases on individualized approaches stand the invariable norms that play a negative role because they restrict the never-to-be-regulated ‘creative’ work of the Beamte, which is normally seen as a positive thing.

In the fourth edition of his classic work *Administrative Behavior* (1997, 22), Simon addresses the role of information communication technologies for administrative behavior, declaring that:

We have learned by now that the computer, too, is something far different from an oversized adding machine, and far more significant for our society. But its significance is only just beginning to emerge, the appearance of the personal computer about a decade ago perhaps being a decisive turning point. One way to conjecture what important novel tasks computers may take on is to review the many metaphors that have been applied to them. **First, the computer is an incredibly powerful number cruncher.** We have already proceeded, especially in engineering and science, a considerable way toward

discovering what can be done by number crunching, but we will find new uses as computer power continues to increase. **Second, the computer is a large memory**, and we are just beginning to explore (for example, on the World Wide Web) how large data bases must be organized so that they can be accessed selectively and cheaply in order to extract information they contain that is relevant to our specific tasks. **Third, the computer is an expert, capable of matching human professional-level performance** in some areas of medical diagnosis, of engineering design, of chess playing, of legal search, and increasing numbers of others. **Fourth, the computer is the core of a new worldwide network of communications**, an ‘information superhighway.’ Everyone can now communicate with ‘everyone,’ almost instantaneously. **Fifth, the computer is a ‘giant brain,’ capable of thinking, problem solving, and yes, making decisions.** We are continually finding new areas of decision – evaluating credit risks, investing funds, scheduling factories, diagnosing corporate financial problems – where computers can play an important role or sometimes do the whole task. . . . **The central lesson the computer should teach is that information is no longer scarce** or in dire need of enhanced distribution. In contrast with past ages, we now live in an information rich world.

This focus on information management informs Simon’s recommendations on how to match information systems to appropriate technology:

The key to successful design of information systems lies in matching the technology to the limits of the attentional resources. From this general principle, we can derive several rules of thumb to guide us when we are considering adding a component to an existing information system. In general, an additional component (man or machine) for an information-processing system will improve the system’s performance only if:

1. Its output is small in comparison with its input, so that it conserves attention instead of making additional demands on attention;
2. It incorporates effective indexes of both passive and active kinds (indexes are processes that automatically select and filter information for subsequent transmission); and
3. It incorporates analytic and synthetic models that are capable not merely of storing and retrieving information, but of solving problems, evaluating solutions, and making decisions. (Simon 1997, 248)

The Bureaucracy of Max Weber

The implementation of AI into Weberian bureaucracy presents challenges for the status quo of these organizations, including 1) a decreased scope of tasks for human bureaucrats, 2) the loss of human managerial control, and 3) an evolving bureaucracy.

Decreased scope of tasks for human bureaucrats

The number and extent of AI applications being deployed across public services is slowly, but steadily, increasing. Objective decision making requires rules and rationality, but how might these guidelines, couched in these classic principles of bureaucracy, be reshaped via human-AI collaboration systems?

One major change will be the attenuation of specialization, a crucial element of efficiency, as AI takes over low-end, labor-intensive, low discretionary, and protocol-defined tasks. For example, Agarwal (2018) found machine learning and computer vision technologies to be promising methods for understanding and interpreting images. While scholars disagree on whether disruptive AI innovations will enhance, worsen, or take over the distributed capital to labor (Korinek and Stiglitz 2017), we predict that human bureaucrats' tasks will gradually focus more on cross-checks and coordination of cross-boundary problems. This result will challenge the specified division of labor scheme proposed by Weber. This uncertainty will soon affect the interdependence of the task process and eventually create inefficiency in public organizations.

Moreover, Weber's ideal type bureaucracy relies on the concept of technocracy, which emphasizes rationality and logic, leading to an interesting juxtaposition of human versus AI technocracy (Sætra 2020). Before AI's emergence, the terms 'technocracy' and 'technocrats' indicated rule by experts and their proponents (Putnam 1977). However, the entrance of AI complicates the concept within the Weberian organization. Are humans and AI complementary or supplementary in bureaucracy? How does the objective of rationality factor into the process of AI reform? Using Weber's theoretical lenses, we discuss three potential risks below.

Loss of human managerial control

Weber's view of managerial control of the bureaucracy defines the modern bureaucratic organization as a structure of hierarchy, specialization, and formalization. Hierarchy creates positions and status, and defines the relationships between supervisors and subordinates. Specialization occurs within this hierarchical structure, as tasks are divided into simple, functional, and routine sub-components. Under this power structure, supervisors can manage both performance and their employees with ease and efficiency. Meanwhile, formalization, critical in large public agencies, supports organizational control by standardizing and

documenting the processes by which a task is to be completed. Similarly, for the purpose of control, whether autocratic or monopolistic, Weber advocated minimizing free and arbitrary “bureaucratic judgment,” declaring that “for all state activities that fall outside the field of law creation and court procedure, one has become accustomed to claims for the freedom and the paramountcy of individual circumstances” (979).

The development of the internet and ICT improved information flows and communications across both organizations and the globe, but it also radically changed the public management regime (Dunleavy 2005). Organizations face a growing need for coordination and monitoring to deal with complex, cross-boundary public affairs. Digital governance has spawned new public issues and problems that agencies look to solve using new AI tools and machine algorithms. For example, a national cybersecurity information-sharing platform must develop automated and collaborative data-sharing and reporting protocols to ensure information immediacy and comparability across public and private sectors (Pala & Zhuang 2019). Japan’s Regional Economy and Society Analyzing System (RESAS) uses big data from multiple sources and AI analytics to visualize the performance of the local economy. In this data-driven era of public administration, such processes may generate new, specialized jobs or tasks to coordinate such functions.

This introduction of AI and algorithms challenges the structures, functions, and controls within Weber’s ideal bureaucracy by affecting both political legitimacy and managerial capacity (Young Bullock & Lecy 2019; Kim & Bullock 2020). Organizations must grow and innovate to deal with large and complex public problems, complicating the management of bureaucracy from the perspectives of network (O’Toole 1997) and complexity theory (Miller & Page 2007). To prevent the loss of human managerial control, we must cultivate a better understanding of the role of adaptive and self-steering AI agents within the bureaucratic labor force.

Evolving bureaucracy

The implementation of AI is evolving bureaucracy, particularly administrative discretion, presenting another challenge to Weber’s bureaucracy. As powerful algorithms affect organizational routines and structures, specialized experts within the ideal bureaucratic organizations may act as information stewards to address growing criticism from the outside (Huang et al. 2021). Bovens and Zouridis (2002) argue that under the diffusion of ICTs, the roles and functions of public servants have shifted from street- to screen- and system-level bureaucrats, suggesting a more technical administrative shift. They note that new technologies remove or limit the discretion of public servants’ professional discretion across multiple tasks, such as communicating with citizens, processing data and information, and making a final judgement.

AI's powerful predictive analysis can deliver a probabilistic-oriented, predictive discretion approach within an organization, directly affecting the knowledge-focus and rule-based practice of administrative personnel. Bullock et al. (2020) present case analyses of policing and public health insurance administration and find the relationship between AI and discretionary power depends on the organizational contexts and task characteristics: system-level changes may not follow in a linear fashion. In policing, for example, tasks often require higher discretion by humans, and citizens and organizations may resist system-level AI transitions that limit that discretion.

The Administrative Behavior of Herbert Simon

Simon, like Weber, was interested in the question of organizational control, but focused on mechanisms beyond government-controlled organizations that provide public services. Simon (1997) proposed the theory of bounded rationality, based on 1) the human tendency to choose among preferred alternatives, 2) human cognitive limitations, and 3) time-constrained decision making. He considered organizations to be artificial things that systematically process information in order to comprehend and influence an increasingly complex modern environment.

Human cognitive processing, which Simon explained as having three levels, can be simulated and modeled by tracking information-processing and decision-making tasks. His bounded rationality theory challenged classic rationality thinking and emphasized the limited ability of human beings to adapt, delineate, and act optimally in a complex environment. Simon's insight informs the conclusions of this chapter, that organizational complexity and the radical improvement in artificial algorithms heighten problems of incomplete information about alternatives. Simon's higher mental level requires the most scholarly attention: it involves organizational goals such as problem-solving, concept-attainment, rule induction, and communication processing tasks (Simon 1979).

Below, we identify three challenges AI presents to organizations based on Simon's conceptualizations of administrative behavior and managing information complexity. We argue that the introduction of AI (1) reshapes the contours of the bounded rationality of agents, (2) complicates the value alignment of public organizations and their mission, and (3) changes the balance of formal and informal communication.

Reshaping the Contours of the Bounded Rationality of Agents

Under Simon's information-processing approach to understanding complexity, organizations are the vehicles through which humans can coordinate towards common goals that require multi-agent efforts. Decision-making agents are placed at decision and action points along the organizational information-processing pathway. Simon (1997) recognized the important role of

ICT tools in reshaping the capabilities of individual decision-making agents and the limits placed upon their rationality. Imagine organizations before the advent of personal computers, emails, search engines, and real-time web maps. These new digital communication and processing technologies greatly enhance the decision-making capacities both of modern institutions and the boundedly-rational humans within them.

AI tools have begun to resemble learning and acting agents deployed into organizational decision making, representing a qualitative shift in the nature of the technological tools themselves and their various roles and responsibilities (decision making, communication, and action). This shift cannot help but impact how the organization gathers, analyzes, organizes, and communicates information: human counterparts will require different approaches to management, control, information flow, decision making, and action execution.

Sunstein and Kahneman 2021 articulate this difference in *Noise: A Flaw in Human Judgement*. This work considers the assertion of philosopher Dan Dennett (and others) that deep meaning and understanding, which constitute the ability to morally reflect upon one's actions, are absent from machine reasoning. This points to the important distinctions between knowledge and understanding, and the relative importance of each as applied to decision-making and action-execution tasks.

Furthermore, the boundedness of agent rationality may present a challenge to decision making systems built to leave humans predominantly “out of the loop,” or in which multiple specialized AI agents engage in solving joint decision-making tasks. AI processes information, makes decisions, and acts in ways that do not perfectly overlap with their human equivalents: since human observers may struggle to comprehend and control these systems, cooperation may be difficult.

Value Misalignment of Public Organizations

Simon's organizations are information-processing entities, and decisions, based upon assumed or implicit facts and values, are the basic building blocks of this information flow. The co-mingling of facts and values give rise to decisions, and thus to administrative behavior. As such, an organization should be very concerned with *which facts* and *which values* are brought to bear upon any particular course of action.

Simon's framework highlights the difficulty of aligning and controlling human behavior when organizational goals do not perfectly overlap with those of their human agents, creating problems of authority, motivation, and value alignment. While both Simon and Weber write directly to this challenge, Simon drives home the point that values, as basic input premises, infiltrate at individual decision points throughout organizational flows of action and information processing.

In other words, in the core message of organizational value misalignment and AI alignment safety research, the dangers of AI grow mostly from their disruptions to individuals' decision making and value systems rather than from the technology itself (Cho & Cain 2020). These arguments assume that there is some quality of *human decision making* and *human values* that does not naturally arise in the modern machine learning form of AI. Put bluntly, human value systems are contextual and complex, and rooted in subjective human experiences, and thus are unlikely to come forth organically from the manipulation of digital machine processing and code. So, as AI acts as *decision maker* in the administrative behavior process, these decisions *inherently* follow some value system that is unlikely to match those of the organization or its individual human agents.

The value misalignment of a human-AI collaboration is likely to be a two-way interaction, as either humans input biased training data to the machine, or the algorithms feed biased data to humans. Surprisingly, in the presence of AI-supporting tools, the interaction of human bias and algorithm bias is likely to worsen the problem (Sun, Nasraoui, & Shafto 2020). Recent empirical studies have found an increases in racial and social discrimination in web-searching algorithms (Noble 2018). Janssen and Kuk (2016) revealed that human faults appear in the workings of algorithms, as, for example, discrimination and inequality.

Additionally, Simon's view on the stimulus and response mechanism of an individual's decision-making behavior explains the uncertainty in the exercise of administrative discretion, which is likely to bias the perceptions of public organizations and their people on emerging technologies, or, even worse, lead them to make the wrong choices. Cognitive biases may confirm their existing perceptions of AI and its value (Huang et al. 2021), similar to what Simon called the mechanisms of behavior persistence, creating internal stimuli for further actions. Furthermore, users may practice automation bias, accepting faulty recommendations based on type II errors (false positives) made by AI and algorithms (Young et al., forthcoming).

Biases may decrease the effectiveness of public performance and increase value misalignment in service provision. Therefore, the introduction of AI agents into organizational information processing and decision-making flow presents unique and important challenges for value alignment throughout a public organization.

The Balance of Formal and Informal Communication & Authority

Simon reminds us that it is key to use communications to encourage the flow, absorption, and storage of knowledge within organizations. Organizational communications tie closely with organizational levels; formal organizational charts and rules determine the dispersion of hierarchical authority. Formal and informal expressions along the rank of authority play an equally important role in the dispersion of meaningful communication throughout an

administrative organization. While formal communications may be decisive, informal authority based around social relationships, organizational power, and other intraorganizational networks enduringly influence how power is conferred and understood and grease the organizational wheels. All formalized systems of authority operating in an environment this complex are, by definition, incomplete tools for dispersing authority required across all decision sets. Similarly, formal information channels play the dominant or decisive role in communication, but lack the capacity to manage all needed information flows, motivate personnel, and strengthen team commitment.

The promotion of human-AI collaboration could change the balance of formal and informal communication within public organizations. Considering that the purpose of communication is to reduce information asymmetries, emerging AI tools raise questions about interactions between communicators and the resulting transmission of information for decision making. Here, we propose two possible scenarios. First, the use of a social AI tool, with its personalized predictions, may improve intra-organizational communications and truthful information transmission by spotting individual problems and dissatisfactions (Schenk 2021). However, this social algorithm could alternately amplify the curation of certain information and affect trust and communications within organizations (Lazer 2015). This triad of storage mechanisms (human brains, computer records, and machine learning algorithmic systems), operating in the presence of an artificial intelligent tool with the ability to learn and adjust its solutions based on prior experience and mistakes, broadens the scope of communication problems.

In summary, both Weber and Simon reflected on the control of administrative behavior, considering structure and human psychology as opportunities and challenges for effective, controlled coordination. However, they addressed these questions through differing lenses. Weber emphasized the structures and functions of bureaucratic behavior, whereas Simon dealt with the psychology of information processing and administrative decisions. While Weber used this structured power and control across organized human behavior as a starting point to illustrate the ideal bureaucracy, Simon's logical framework built from the organizational requirement for individual information processing and decision making. Meanwhile, the comparison of these two theoretical views suggests overlapping strategies for managing AI's risks to organizational decision making and effective human control. We explore these strategies below.

Discussion: Dehumanization and Loss of Control

Weber and Simon highlight distinct and shared challenges to public organizations. In an evolutionary process influenced by rules, humans, and technology, AI is leading the reshaping of resources and communication, and thus decisions and actions. In this trajectory, human judgment and discretion are constricted in many domains. Recent related work suggests that AI may engage in actions that are systematically unaligned with the interests of individual humans or of

humanity more generally. Beyond value misalignment, dehumanizing impacts, amplified by AI, are already at play in modern organizations, systematically elevating dangers of administrative evil in public organizations that wield power and direct resources (Young et al 2021).

While these elements of dehumanization and loss of control in a changing world paint a disturbing picture, the classic theories of Weber and Simon can help us navigate it. The first step is to apply Weber's concept of a *Beamte* to the development of bounded, controlled machine with certain levels of discretion, similar to the explicit and tacit rules that enable and constrain civil servants. Next, we recommend using Weber's concepts of hierarchy, communication, and specialization as guides for structuring the working tasks of AI or machine agents, and his repulsion for a judging machine to help determine the types of decisions AI are authorized to make.

Simon also gives us tools and guidance for managing administrative organizations as they integrate decision making and action with AI. For Simon, the careful identification of facts and values that guide administrative decision making is a reminder to apply AI with care, and with the knowledge that AI may substitute its own implicit value at any decision point. Simon's notion of complexity and specialization also highlights the types of tasks that AI may usefully address, and paths for controlling its influence. He also highlights the challenges of coworking arrangements and the need for effective communication and decision making. This analysis suggests that creating effective human and AI teams will be challenging and critical.

It may further be the case that the integration of AI into large public organizations forces us to move beyond the insights of Weber and Simon and reconceptualize larger modern public organizations as their own emergent phenomenon. Bureaucracy scholars have already noted the significant degree of evolution that has occurred among certain types of public organizations, from street-level bureaucracies to systems-level bureaucracies, which predated the fuller integration of AI and machine agents into decision making. Brockman (2020) has described these new large public organizations as "hybrid superintelligences" of human and machine. This combination allows the entity to surpass the intelligence and decision making capability of any single group of humans.

However they might be known, modern public organizations are clearly evolving with their environment and the tools at their disposal. As AI is dispersed throughout the decision-making functions of these institutions, they may feature greater dehumanization and lower levels of human control. This is, to steal an overused term, unprecedented in the history of public organizations.

Conclusion

Max Weber and Herbert Simon each provided enduring pictures of how organizations decide, act, behave, and learn. Weber produced the classic notion of an ideal bureaucracy as a way to organize human behavior to conduct the business of government. This structure, with its clear lines of hierarchy, authority, and communication, is inhabited by trained professionals who apply their knowledge of the rules and circumstance to the best of their ability.

Simon took the decision as the basic premise for organizational action, building out a structure and concepts for hierarchy, authority, and communication. He proposed designing the flow of information throughout an organization so that the most qualified decision maker makes and implements the call. There are, of course, other pictures of public organizations. Prominent among these is Elinor Ostrom's notion of Institutional Analysis and Development, a broader conception of governance that considers organizational interactions in action spaces. While illuminating, this construct is less helpful for understanding the structure and decision making of large, bureaucratic organizations.

We have argued that as the external environment has continued to create increasingly capable AI and machines, public organizations have begun to make more systematic use of these technologies, leading to an evolution in many organizations. To predict the direction of this evolution, we turned to Weber and Simon for guidance on how the incorporation of AI into organizational decision-making functions may influence its trajectory. For Weber, the threats were (1) a decreased scope of tasks for human bureaucrats, (2) the loss of human managerial control, and 3) an evolving bureaucracy structure. Using Simon's framework, we argued that AI (1) reshapes the contours of the bounded rationality of agents, (2) presents challenges for value alignment of public organizations with their mission, and (3) changes the balance of formal and informal communication. These evolutionary changes point to losses in human control and general dehumanization of the decision-making process and corresponding actions.

Furthermore, again based on the classic organizational pictures of Weber and Simon, we learn the importance of restricting the agency of AI through formal hierarchy and communication. AI should only be applied to decisions that would not benefit from human judgement and do not have the potential further dehumanization and loss of human control. We have also questioned whether these models are as useful as they once were for describing the behavior of modern organizations that conduct the tasks of governments.

With the incorporation of AI and digital networking technologies throughout modern public organizations, it may be better to think of these new entities as hybrids, human and machine entities that are so deeply intertwined as to be inseparable, leading to a decrease in human control and an increase in the dehumanization of decisions and actions. Left unchecked, or

steered poorly, these incredibly powerful and large hybrid public organizations may become, forever, misaligned with the interests of humanity and the freedom of individual humans.

References

- Agarwal, P. K. 2018. "Public Administration Challenges in the World of AI and Bots: Public Administration Challenges in the World of AI and Bots." *Public Administration Review* 78 (6): 917–21. <https://doi.org/10.1111/puar.12979>.
- Arrow, Kenneth J. 1984. "The Economics of Agency. In *Principals and Agents: The Structure of Business*, Ed. John W. Pratt and Richard J. Zeckhauser, 1–35." Boston, MA: Harvard Business School Press.
- Bellotti, Marianne. 2021. "Helping Humans and Computers Fight Together: Military Lessons from Civilian AI." *War on the Rocks*. March 15, 2021. <https://warontherocks.com/2021/03/helping-humans-and-computers-fight-together-military-lessons-from-civilian-ai/>.
- Bertelli, A. M. (2012). *The political economy of public sector governance*. Cambridge University Press.
- Bovens, Mark, and Stavros Zouridis. 2002. "From Street-Level to System-Level Bureaucracies: How Information and Communication Technology Is Transforming Administrative Discretion and Constitutional Control." *Public Administration Review* 62 (2): 174–84.
- Brynjolfsson, Erik, and Andrew McAfee. 2017. "The Business of Artificial Intelligence." *Harvard Business Review*, July 18, 2017. <https://hbr.org/2017/07/the-business-of-artificial-intelligence>.
- Brockman, John, ed. *Possible minds: Twenty-five ways of looking at AI*. Penguin, 2020.
- Bullock, Justin B. 2019. "Artificial Intelligence, Discretion, and Bureaucracy." *The American Review of Public Administration* 49 (7): 751–61. <https://doi.org/10.1177/0275074019856123>.
- Bullock, Justin B., and Kyoung-cheol Kim. 2020. "Creation of Artificial Bureaucrats." *Proceedings of European Conference on the Impact of Artificial Intelligence and Robotics*.
- Bullock, Justin, Matthew M. Young, and Yi-Fan Wang. 2020. "Artificial Intelligence, Bureaucratic Form, and Discretion in Public Service." *Information Polity* 25 (4): 491–506. <https://doi.org/10.3233/IP-200223>.
- Busch, Peter André, and Tom Roar Eikebrokk. 2019. "Digitizing Discretionary Practices in Public Service Provision: An Empirical Study of Public Service Workers' Attitudes."
- Busch, P. A., & Henriksen, H. Z. (2018). Digital discretion: A systematic literature review of ICT and street-level discretion. *Information Polity*, 23(1), 3-28.
- Cho, Wendy K. Tam, and Bruce E. Cain. 2020. "Human-Centered Redistricting Automation in the Age of AI." *Science* 369 (6508): 1179–81. <https://doi.org/10.1126/science.abd1879>.
- Dempsey, Michael. 2021. "The Navy Sub Commanded by Artificial Intelligence." *BBC News*, May 20, 2021, sec. Business. <https://www.bbc.com/news/business-56993035>.
- Dunleavy, P. 2005. "New Public Management Is Dead--Long Live Digital-Era Governance." *Journal of Public Administration Research and Theory* 16 (3): 467–94. <https://doi.org/10.1093/jopart/mui057>.
- Farber, Ken, Kristine Lam, Ellery Taylor, Erica Briscoe, Erica Dretzka, John Scott, John Sprague, et al. 2021. "From Ethics to Operations: Current Federal AI Policy." Advanced Technology Academic Research Center.

- Hambling, David. 2021. "U.S. Navy Destroys Target With Drone Swarm — And Sends A Message To China." *Forbes*. April 30, 2021.
<https://www.forbes.com/sites/davidhambling/2021/04/30/us-navy-destroys-target-with-drone-swarm---and-sends-a-message-to-china/>.
- Huang, Hsini, Kyoung-Cheol (Casey) Kim, Matthew M. Young, and Justin B. Bullock. 2021. "A Matter of Perspective: Differential Evaluations of Artificial Intelligence between Managers and Staff in an Experimental Simulation." *Asia Pacific Journal of Public Administration*, September, 1–19. <https://doi.org/10.1080/23276665.2021.1945468>.
- Huber, J. D., & Shipan, C. R. (2002). *Deliberate discretion?: The institutional foundations of bureaucratic autonomy*. Cambridge University Press
- Janssen, Marijn, and George Kuk. 2016. "The Challenges and Limits of Big Data Algorithms in Technocratic Governance." *Government Information Quarterly* 33: 371–77.
- Kahneman, Daniel, Olivier Sibony, and Cass R. Sunstein. *Noise: A flaw in human judgment*. Little, Brown, 2021.
- Korinek, Anton, and Joseph E. Stiglitz. 2017. "Artificial Intelligence and Its Implications for Income Distribution and Unemployment." In *The Economics of Artificial Intelligence: An Agenda*. Conference Held by National Bureau of Economic Research (2019). University of Chicago Press.
- Lazer, David. 2015. "The Rise of the Social Algorithm." *Science* 348 (6239): 1090–91.
<https://doi.org/10.1126/science.aab1422>.
- Miller, John H., and Scott Page. 2007. *Complex Adaptive Systems: An Introduction to Computational Models of Social Life*. Princeton, N.J: Princeton University Press.
- Noble, Safiya Umoja. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism*. Illustrated edition. New York: NYU Press.
- O'Toole, Laurence J. 1997. "Treating Networks Seriously: Practical and Research-Based Agendas in Public Administration." *Public Administration Review* 57 (1): 45.
<https://doi.org/10.2307/976691>.
- Pala, Ali, and Jun Zhuang. 2019. "Information Sharing in Cybersecurity: A Review." *Decision Analysis*, August. <https://doi.org/10.1287/deca.2018.0387>.
- Putnam, Robert D. 1977. "Elite Transformation in Advanced Industrial Societies: An Empirical Assessment of the Theory of Technocracy." *Comparative Political Studies* 10 (3): 383–412. <https://doi.org/10.1177/001041407701000305>.
- Rourke, Francis E. 1972. *Bureaucratic Power in National Politics*. Second Edition, First Printing. Little, Brown.
- Sætra, Henrik Skaug. 2020. "A Shallow Defence of a Technocracy of Artificial Intelligence: Examining the Political Harms of Algorithmic Governance in the Domain of Government." *Technology in Society* 62 (August): 101283.
<https://doi.org/10.1016/j.techsoc.2020.101283>.
- Schenk, Alicia von. 2021. "Predictably Satisfied: Contributions of Artificial Intelligence to Intra-Organizational Communication." *SSRN Electronic Journal*.
<https://doi.org/10.2139/ssrn.3856479>.
- Simon, Herbert A. 1946. "The Proverbs of Administration." In *Classics of Public Administration*, 7th ed. Boston, MA: Wadsworth/Cengage Learning.
- . 1979. "Rational Decision Making in Business Organizations." *The American Economic Review* 69 (4): 493–513.

- . 1997. *Administrative Behavior*. 4th ed. New York: Simon & Schuster.
- Simon, H. (1996). *The Sciences of the Artificial* 3rd Edition Massachusetts.
- Sun, Wenlong, Olfa Nasraoui, and Patrick Shafto. 2020. “Evolution and Impact of Bias in Human and Machine Learning Algorithm Interaction.” *PLOS ONE* 15 (8): e0235502. <https://doi.org/10.1371/journal.pone.0235502>.
- Waters, Tony, and Dagmar Waters, eds. 2015a. *Weber’s Rationalism and Modern Society*. New York: Palgrave Macmillan US. <https://doi.org/10.1057/9781137365866>.
- , eds. 2015b. *Weber’s Rationalism and Modern Society*. New York: Palgrave Macmillan US. <https://doi.org/10.1057/9781137365866>.
- Weber, Max. 1922. “Bureaucracy.” In *Classics of Public Administration*, 7th ed. Boston, MA: Wadsworth/Cengage Learning.
- Young, Matthew M, Justin B Bullock, and Jesse D Lecy. 2020. “Artificial Discretion as a Tool of Governance: A Framework for Understanding the Impact of Artificial Intelligence on Public Administration.” *Perspectives on Public Management and Governance* 2 (4): 301–13. <https://doi.org/10.1093/ppmgov/gvz014>.
- Young, Matthew M, Johannes Himmelreich, Justin B Bullock, and Kyoung-Cheol Kim. 2021. “Artificial Intelligence and Administrative Evil.” *Perspectives on Public Management and Governance* 4 (3): 244–58. <https://doi.org/10.1093/ppmgov/gvab006>.