Public Law and Legal Theory Research Paper Series
Research Paper No. 22-11

Algorithm vs. Algorithm

Cary Coglianese
UNIVERSITY OF PENNSYLVANIA CAREY LAW SCHOOL

Alicia Lai
UNIVERSITY OF PENNSYLVANIA CAREY LAW SCHOOL

Algorithm vs. Algorithm

Cary Coglianese[†] and Alicia Lai[††]

Abstract

Critics raise alarm bells about governmental use of digital algorithms, charging that they are too complex, inscrutable, and prone to bias. A realistic assessment of digital algorithms, though, must acknowledge that government is already driven by algorithms of arguably greater complexity and potential for abuse: the algorithms implicit in human decision-making. The human brain operates algorithmically through complex neural networks. And when humans make collective decisions, they operate via algorithms too—those reflected in legislative, judicial, and administrative processes. Yet these human algorithms undeniably fail and are far from transparent. On an individual level, human decision-making suffers from memory limitations, fatigue, cognitive biases, and racial prejudices, among other problems. On an organizational level, humans succumb to groupthink and free-riding, along with other collective dysfunctionalities. As a result, human decisions will in some cases prove far more problematic than their digital counterparts. Digital algorithms, such as machine learning, can improve governmental performance by facilitating outcomes that are more accurate, timely, and consistent. Still, when deciding whether to deploy digital algorithms to perform tasks currently completed by humans, public officials should proceed with care on a case-by-case basis. They should consider both whether a particular use would satisfy the basic preconditions for successful machine learning and whether it would in fact lead to demonstrable improvements over the status quo. The question about the future of public administration is not whether digital algorithms are perfect. Rather, it is a question about what will work better: human algorithms or digital ones.

[†] Edward B. Shils Professor of Law and Director, Penn Program on Regulation, University of Pennsylvania Law School.

[††] Judicial Law Clerk, United States Court of Appeals for the Federal Circuit.

The opinions set forth in this article are solely those of the authors and do not represent the views of any other person or institution. The authors gratefully acknowledge the many helpful comments on this project from participants in sessions where draft versions of this manuscript were presented, including at the European Consortium of Political Research, Harvard Kennedy School, Northwestern University School of Law, the University of Pennsylvania Law School, and Vanderbilt University School of Law. We are grateful for helpful comments from participants in these sessions as well as from Richard Berk, Madalina Busuioc, David Lehr, and Aaron Roth, along with input on related projects by Steven Appel and Lavi Ben Dor. The journal's editors—especially Beresford Clarke, Karen Sheng, Emma Ritter, and Drew Langan—offered incisive feedback, while Steven Appel, Emma Ronzetti, Jasmine Wang, and Roshie Xing provided valuable research assistance. This article is based in part on a report initially prepared for the Administrative Conference of the United States ("ACUS"), but the views expressed here are those of the authors and not necessarily ACUS or its members or staff. This article is forthcoming in the *Duke Law Journal*.

Algorithm vs. Algorithm

Cary Coglianese and Alicia Lai

TABLE OF CONTENTS

INTRODUCTION

Computerized algorithms increasingly automate tasks that previously had been performed by humans.[1] They now routinely assist with, or even make, decisions

[1] An algorithm is simply a set of steps designed to solve a problem. A digital algorithm is an algorithm that has its steps executed via computer. Algorithms are not unique to the digital age; they have been part of human societies for millennia. *See* BRIAN CHRISTIAN & TOM GRIFFITHS, ALGORITHMS TO LIVE BY: THE COMPUTER SCIENCE OF HUMAN DECISIONS 2–4 (2016). Today, modern computing power permits humans to take advantage of a distinctive type of digital algorithm known as a machine-learning algorithm—often referred to as artificial intelligence ("AI"). Machine-learning algorithms can learn to identify patterns across the vast quantities of data stored and processed digitally, and these algorithms can detect these patterns autonomously—that is, without human

about business hiring,[2] loan approvals,[3] stock trading,[4] and product marketing.[5] They also drive Internet search and autonomous vehicles, and they provide the backbone for both advanced medical techniques as well as the everyday use of smartphones.[6]

---

specification of the form of a particular model or key variables, and subject mainly to overarching criteria or parameters to be optimized. These algorithms are also often included under the banner of so-called big data. For a discussion of machine learning and how it works, see, e.g., Cary Coglianese & David Lehr, *Regulating by Robot: Administrative Decision Making in the Machine-Learning Era*, 105 GEO. L.J. 1147, 1156–60 (2017) [hereinafter Coglianese & Lehr, *Regulating by Robot*]; David Lehr & Paul Ohm, *Playing with the Data: What Legal Scholars Should Learn About Machine Learning*, 51 U.C. DAVIS L. REV. 653, 669–702 (2017). Machine-learning algorithms come in many forms and are referred to by a variety of terms. *See* Cary Coglianese & David Lehr, *Transparency and Algorithmic Governance*, 71 ADMIN. L. REV. 1, 2 n.2 (2019) [hereinafter Coglianese & Lehr, *Transparency*] ("By 'artificial intelligence' and 'machine learning,' we refer . . . to a broad approach to predictive analytics captured under various umbrella terms, including 'big data analytics,' 'deep learning,' 'reinforcement learning,' 'smart machines,' 'neural networks,' 'natural language processing,' and 'learning algorithms.'"). The particular type of machine-learning algorithm deployed for any specific use will no doubt affect its performance in that setting. For our purposes here we focus generically and broadly on the class of digital algorithms that today can drive automated forecasting and decision-making tools with the potential to substitute for or complement traditional human decision-making within government. For further elaboration of what we mean by machine-learning algorithms and AI, see *infra* Part II.A.

[2] *See* Claire Cain Miller, *Can an Algorithm Hire Better Than a Human?*, N.Y. TIMES (June 25, 2015), https://www.nytimes.com/2015/06/26/upshot/can-an-algorithm-hire-better-than-a-human.html [https://perma.cc/39MF-8QP7].

[3] *See* Scott Zoldi, *How To Build Credit Risk Models Using AI and Machine Learning*, FICO BLOG (Apr. 6, 2017), http://www.fico.com/en/blogs/analytics-optimization/how-to-build-credit-risk-models-using-ai-and-machine-learning [https://perma.cc/N3UW-XKRB].

[4] *See* Jigar Patel, Sahil Shah, Priyank Thakkar & K Kotecha, *Predicting Stock and Stock Price Index Movement Using Trend Deterministic Data Preparation and Machine Learning Techniques*, 42 EXPERT SYS. WITH APPLICATIONS 259, 259 (2015).

[5] *See Using Machine Learning on Computer Engine To Make Product Recommendations*, GOOGLE CLOUD PLATFORM (Feb. 14, 2017), https://cloud.google.com/solutions/recommendations-using-machine-learning-on-compute-engine [https://perma.cc/C4D3-9PCE].

[6] *See, e.g.*, PAUL CERRATO & JOHN HALAMKA, THE DIGITAL RECONSTRUCTION OF HEALTHCARE: TRANSITIONING FROM BRICK AND MORTAR TO VIRTUAL CARE 82–84 (2021); Alexis C. Madrigal, *The Trick That Makes Google's Self-Driving Cars Work*, ATLANTIC (May 15, 2014), http://www.theatlantic.com/technology/archive/2014/05/all-the-world-a-track-the-trick-thatmakes-googles-self-driving-cars-work/370871 [https://perma.cc/9CWC-HTL6]; Nikhil Dandekar, *What Are Some Uses of Machine Learning in Search Engines?*, MEDIUM (Apr. 7, 2016), https://medium.com/@nikhilbd/what-are-some-uses-of-machine-learning-in-search-engines-5770f534d46b [https://perma.cc/7DJD-TDPX]; Steffen Herget, *Machine Learning and AI: How Smartphones Get Even Smarter*, NEXTPIT (Jan. 24, 2018), https://www.androidpit.com/machine-learning-and-ai-on-smartphones [https://perma.cc/XJ3J-CV6L]. For a survey of the state of the art in AI and its varied applications, see generally MICHAEL L. LITTMAN ET AL., GATHERING STRENGTH, GATHERING STORMS: THE ONE HUNDRED YEAR STUDY ON ARTIFICIAL INTELLIGENCE (AI100) 2021 STUDY PANEL REPORT (2021).

The speed and accuracy of these digital algorithms have made them key to developing automated systems that augment or replace humans in a range of tasks.

The use of these digital algorithms extends beyond the private sector. Militaries and intelligence agencies deploy them in conflicts both on the ground and in cyberspace.[7] Law enforcement agencies and criminal courts are turning to various digital algorithms in an effort to investigate or even predict crime, as well as to make decisions about pretrial detention or parole.[8] Other governmental bodies are starting to use digital algorithms to administer social services programs, adjudicate claims for government benefits, and support regulatory functions.[9]

Academics and other commentators have responded by scrutinizing the use of digital algorithms by governmental authorities—particularly the use of the most advanced types of such algorithms, namely those that depend on machine-learning analysis. Much of the scrutiny of advanced digital algorithms has been critical.[10]

---

[7] *See, e.g.*, Ronen Bergman & Farnaz Fassihi, *The Scientist and the A.I.-Assisted, Remote-Control Killing Machine*, N.Y. TIMES (Sept. 18, 2021), https://www.nytimes.com/2021/09/18/world/middleeast/iran-nuclear-fakhrizadeh-assassination-israel.html [https://perma.cc/5SLN-D5XJ]; Patrick Tucker, *Spies Like AI: The Future of Artificial Intelligence for the US Intelligence Community*, DEF. ONE (Jan. 27, 2020), https://www.defenseone.com/technology/2020/01/spies-ai-future-artificial-intelligence-us-intelligence-community/162673 [https://perma.cc/CFQ4-GNEX]; PAUL SCHARRE, ARMY OF NONE: AUTONOMOUS WEAPONS AND THE FUTURE OF WAR 5–6 (2018); Andrew Tarantola, *The Pentagon Is Hunting ISIS Using Big Data and Machine Learning*, ENGADGET (May 15, 2017), https://www.engadget.com/2017/05/15/the-pentagon-is-hunting-isis-using-big-data-and-machine-learning [https://perma.cc/H5UC-VQV9].

[8] For discussion of algorithmic tools in the criminal law context, see generally, e.g., Richard Berk, Lawrence Sherman, Geoffrey Barnes, Ellen Kurtz & Lindsay Ahlman, *Forecasting Murder Within a Population of Probationers and Parolees: A High Stakes Application of Statistical Learning*, 172 J. ROYAL STAT. SOC'Y 191 (2009); Sandra Mayson, *Bias In, Bias Out*, 128 YALE L.J. 2218 (2019); Cary Coglianese & Lavi M. Ben Dor, *AI in Adjudication and Administration*, 86 BROOK. L. REV. 791–838 (2021); RICHARD A. BERK, ARUN KUMAR KUCHIBHOTLA & ERIC TCHETGEN TCHETGEN, IMPROVING FAIRNESS IN CRIMINAL JUSTICE ALGORITHMIC RISK ASSESSMENTS USING OPTIMAL TRANSPORT AND CONFORMAL PREDICTION SETS (2021), https://arxiv.org/pdf/2111.09211.pdf [https://perma.cc/JN8C-WCHS].

[9] *See* Coglianese & Ben Dor, *supra* note 8, at 20–37. *See generally* DAVID FREEMAN ENGSTROM, DANIEL E. HO, CATHERINE M. SHARKEY & MARIANO-FLORENTINO CUELLAR, GOVERNMENT BY ALGORITHM: ARTIFICIAL INTELLIGENCE IN FEDERAL ADMINISTRATIVE AGENCIES 22–29 (2020), https://www-cdn.law.stanford.edu/wp-content/uploads/2020/02/ACUS-AI-Report.pdf [https://perma.cc/TWE9-JLA5] (examining the deployment of AI by federal agencies).

[10] *See, e.g.*, Danielle Keats Citron, *Technological Due Process*, 85 WASH. U. L. REV. 1249, 1313 (2008); danah boyd & Kate Crawford, *Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon*, 15 INFO. COMMC'N & SOC. 662, 673–75 (2012); FRANK PASQUALE, THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION 3 (2015); CATHY O'NEIL, WEAPONS OF MATH DESTRUCTION: HOW BIG DATA INCREASES INEQUALITY AND THREATENS DEMOCRACY 12–13 (2016); Ryan Calo & Danielle K. Citron, *The Automated Administrative State: A Crisis of Legitimacy*, 70 EMORY L.J. 797, 799–804 (2021).

Commentators warn of the ways that machine learning can reproduce human biases baked into existing datasets, thereby generating discriminatory outcomes for members of historically marginalized groups.[11] Critics also worry that machine-learning algorithms—sometimes called "black box" algorithms—can be opaque and inscrutable, failing to provide adequate reasons to individuals about why they are denied government benefits or are predicted to pose a crime risk and thus denied parole.[12]

Despite machine learning's reputation for accuracy, some observers question whether such digital algorithms are accurate and unbiased enough to make decisions with life-altering consequences.[13] Some commentators even treat governmental use of machine learning as an existential threat by warning, for example, that "a wholesale shift toward algorithmic decision-making systems risks eroding the collective moral and cultural fabric upon which democracy and individual freedom rests, thereby undermining the social foundations of liberal democratic political orders."[14]

The resistance to the widespread application of digital algorithms may reflect some degree of "algorithm aversion."[15] Humans may simply trust machines

---

[11] *See, e.g.*, Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CALIF. L. REV. 671, 680–87 (2016); VIRGINIA EUBANKS, AUTOMATING INEQUALITY: HOW HIGH-TECH TOOLS PROFILE, POLICE, AND PUNISH THE POOR 10–13 (2017); Kate Crawford, *Think Again: Big Data*, FOREIGN POL'Y (May 9, 2013), http://www.foreignpolicy.com/articles/2013/05/09/think_again_big_data [https://perma.cc/D9M6-3BBA].

[12] *See, e.g.*, Jenna Burrell, *How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms*, 3 BIG DATA & SOC'Y 1, 1–2 (2016); *2018 Program*, FAT* CONF. (2018), https://fatconference.org/2018/program.html [https://perma.cc/4AG2-UMWX] (discussing work on algorithmic explanation).

[13] *See, e.g.*, Danielle Keats Citron & Frank Pasquale, *The Scored Society: Due Process for Automated Predictions*, 89 WASH. L. REV. 1, 8 (2014); Margaret Hu, *Algorithmic Jim Crow*, 86 FORDHAM L. REV. 633, 643–44 (2017); Dorothy Roberts, *Digitizing the Carceral State*, 132 HARV. L. REV. 1695, 1695, 1697 (2019).

[14] Karen Yeung, *Algorithmic Regulation: A Critical Interrogation*, 12 REGUL. & GOVERNANCE 505, 517 (2018); *see also* Samuel Gibbs, *Elon Musk: Artificial Intelligence Is Our Biggest Existential Threat*, GUARDIAN (Oct. 27, 2014), https://www.theguardian.com/technology/2014/oct/27/elon-musk-artificial-intelligence-ai-biggest-existential-threat [https://perma.cc/8C6P-YZYZ] ("Elon Musk has . . . declar[ed artificial intelligence] the most serious threat to the survival of the human race."); Rory Cellan-Jones, *Stephen Hawking Warns Artificial Intelligence Could End Mankind*, BBC NEWS (Dec. 2, 2014), http://www.bbc.com/news/technology-30290540 [https://perma.cc/S9Y5-ZK7W] ("The development of full artificial intelligence could spell the end of the human race.").

[15] *See, e.g.*, Berkeley J. Dietvorst, Joseph P. Simmons & Cade Massey, *Algorithm Aversion: People Erroneously Avoid Algorithms After Seeing Them Err*, 144 J. EXPERIMENTAL PSYCHOL. 114, 114 (2015); Benjamin Chen, Alexander Stremitzer & Kevin Tobia, *Having Your Day in Robot Court* 4 (UCLA Pub. L., Rsch. Paper 21-20, May 7, 2021),

less than they trust humans, even when the machines are shown to be more accurate and fairer. Humans may be generally less forgiving when machines make mistakes than when humans do.[16] Perhaps unsurprisingly, some commentators now even consider whether governments must honor a "right to a human decision."[17]

Still, critics of digital algorithms do express concerns that merit consideration.[18] It is far from imagined that automated digital systems can suffer from biases, lead to controversies, or precipitate other problems from the way humans design and use them.[19] Yet too often critics dismiss machine learning categorically, as if the mere existence of any imperfections means that artificial intelligence ("AI") should never be used. Such critics can make it seem as if machine-learning algorithms produce problems that are entirely new or distinctively complex, inscrutable, or susceptible to bias. Unfortunately, the complaints leveled against digital algorithms are neither truly distinctive nor entirely new. Human decision-making is prone to many of the same kinds of problems.[20]

Any meaningful assessment of AI in the public sector must therefore start with an acknowledgment that government as it exists today is already grounded in a set of imperfect algorithms. These existing algorithms are inherent in human decision-making. The human brain has its own internal wiring that might be said to operate like a complex algorithmic system in certain respects. Neural networks—one category of machine-learning algorithms—even draw the name from the physical structures underlying human cognition. In addition to the algorithmic nature of individual human judgment, collective human decisions also follow socially and legally established algorithms: namely, legislative, judicial, and administrative procedures.

---

https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3841534 [https://perma.cc/8PBR-N6RF]. A contrary tendency, of course, can be to give too much weight to the outcomes of algorithmic systems. Kate Goddard, Abdul Roudsari & Jeremy Wyatt, *Automation Bias: A Systematic Review of Frequency, Effect Mediators, and Mitigators*, 19 J. AM. MED. INFORMATICS ASSOC. 121, 121 (2012); Jennifer M. Logg, Julia A. Minson & Don A. Moore, *Algorithm Appreciation: People Prefer Algorithmic to Human Judgment*, 151 ORGANIZATIONAL BEH. & HUM. DECISION PROCESSES 90, 90 (2019).

[16] *See generally* CÉSAR A. HIDALGO, DIANA ORGHIAN, JORDI ALBO-CANALS, FILIPA DE ALMEIDA & NATALIA MARTIN, HOW HUMANS JUDGE MACHINES (2021) (examining human biases about machines).

[17] Aziz Z. Huq, *A Right to a Human Decision*, 106 VA. L. REV. 611, 615–20 (2020).

[18] Criticisms also perform a valuable role by drawing attention to pitfalls and encouraging greater care in the deployment of machine-learning algorithms. *See* Cary Coglianese, *Algorithmic Regulation: Machine Learning as a Governance Tool*, *in* THE ALGORITHMIC SOCIETY: TECHNOLOGY, POWER, AND KNOWLEDGE 35, 50 (Marc Schuilenberg & Rik Peeters eds., 2021).

[19] *See infra* Part II.C.

[20] *See infra* Part II.B.

But at both an individual or collective level, human decision-making is prone to a variety of errors and biases that contribute to numerous governing failures both large and small.[21] In fact, in some settings, human decision-making is arguably more prone to inscrutability, bias, and error than are digital algorithms.[22] As a result, when assessing the use of machine learning in governmental settings, any anticipated shortcomings of machine learning must be placed in proper perspective. The choice is not one between digital algorithms and a Platonic ideal. Rather, the choice is one of digital algorithms versus human algorithms, each with their own advantages and disadvantages. And to the extent that automated systems based on digital algorithms would make improvements over human algorithms for specific tasks, they should be adopted.

Part I of this Article begins by offering a counterweight to the criticisms of machine-learning algorithms that tend to dominate legal scholarship. It details the well-documented physical limitations and cognitive biases that afflict individual decision-making by human algorithms, along with additional problems that can arise when humans make collective decisions.

Part II then focuses on machine learning and its promise for improving decision-making. Of course, even though machine learning can mark improvements over human decision-making for some tasks, this does not mean that it always will work better. To the contrary, automated decision systems can fall prey to problems too; they are, after all, designed and operated by humans subject to the limitations described in Part I.

That machine-learning algorithms could fail means that their design and use, especially by governments, should be carried out with due care and attentive oversight. The aim should be to develop and deploy machine-learning algorithms that can improve on the status quo—that is, do a better job than humans of avoiding errors, biases, and other problems. Achieving that aim calls for smart human decision-making about when and how to rely on digital algorithms.

---

[21] As the U.S. Supreme Court has acknowledged, "[i]t is an unalterable fact that our judicial system, like the human beings who administer it, is fallible." Herrera v. Collins, 506 U.S. 390, 415 (1993). A particularly salient example of fallibility in public administration can be found in the human misjudgments in response to the COVID-19 crisis. *See generally* MICHAEL LEWIS, THE PREMONITION: A PANDEMIC STORY 85, 160–85, 295 (2021) (chronicling misperceptions and missteps that impeded successful public health responses). It is possible to identify a vast array of other failures in human-driven government in recent years. *See, e.g.*, PAUL C. LIGHT, A CASCADE OF FAILURES: WHY GOVERNMENT FAILS, AND HOW TO STOP IT 3–7 (Ctr. For Effective Pub. Mgmt., 2014). The law itself—a product of human decision-making—is said to be riddled with incoherencies in its substance and implementation. *See, e.g.*, LEO KATZ, WHY THE LAW IS SO PERVERSE (2011); Cass R. Sunstein, Daniel Kahneman, David Schkade & Ilana Ritov, *Predictably Incoherent Judgments*, 54 STAN. L. REV. 1153, 1154 (2002); MAX H. BAZERMAN & ANN E. TENBRUNSEL, BLIND SPOTS: WHY WE FAIL TO DO WHAT'S RIGHT AND WHAT TO DO ABOUT IT 96–111 (2011). For further discussion of the limitations of human decision-making, see *infra* Part I.

[22] *See infra* Part II.B.

Part III thus presents general considerations to help guide public officials seeking to make sound choices about when and how to use digital algorithms. In addition to focusing officials' attention on the extent to which a shift to digital algorithms will improve upon the status quo, we emphasize in Part III the need to consider whether a new use of digital algorithms would likely satisfy key preconditions for successful deployment of machine learning and whether a system driven by digital algorithms would actually deliver better outcomes. We also emphasize the need to ensure adequate planning, careful procurement of private contractor services, and appropriate opportunities for public participation in the design, development, and ongoing oversight of machine-learning systems.

I. LIMITATIONS OF HUMAN ALGORITHMS

Human judgment exhibits a series of well-documented limitations and biases.[23] Many of these limitations stem from taking shortcuts, relying on heuristics, or leaping to conclusions before gathering information. Others stem from expediency and self-interest. Against these limitations, machine-learning algorithms promise to do better. Chess champion Garry Kasparov once opined that "[a]nything we can do, . . . machines will do it better."[24] To understand whether he may be right, and whether digital algorithms might in fact do better than humans in specific tasks in government, it helps to start with understanding the limitations of the human mind.

In this Part, we review a range of limitations that affect human decision-making, separating physical or biological capacities from cognitive biases.[25] By shedding light on the flaws underlying the status quo processes that rely on human decision-making, our aim is to reveal the key rationale for considering the responsible use of machine-learning algorithms: to improve governmental performance.[26]

---

[23] For recent syntheses of such research, see generally RICHARD H. THALER, MISBEHAVING: THE MAKING OF BEHAVIORAL ECONOMICS (2015) and DANIEL KAHNEMAN, THINKING, FAST AND SLOW (2011). For a synthesis of research on cognitive biases in legal decision-making, see Alicia Lai, *Brain Bait: Effects of Cognitive Biases on Scientific Evidence in Legal Decision-Making* 8–12 (2018) (A.B. thesis, Princeton University) (on file with the Princeton University Library).

[24] DAVID EPSTEIN, RANGE: WHY GENERALISTS TRIUMPH IN A SPECIALIZED WORLD 22 (2019). Kasparov made this declaration after being defeated by the IBM supercomputer Deep Blue. *See id.*

[25] We adopt these categorizations simply for ease of presentation, not because they are airtight or comprehensive. Nothing of consequence hinges on the categories into which we have grouped these human limitations.

[26] We build on others who have recognized that the limitations of human decision-making can impede sound administrative policymaking. *See, e.g.*, Susan E. Dudley & Zhoudan Xie, *Nudging the Nudger: Toward a Choice Architecture for Regulators*, REGUL. & GOVERNANCE 1, 1 (June 14, 2020), https://doi.org/10.1111/rego.12329 [https://perma.cc/Y462-H74D].

A. Physical Limitations

Physical limitations constitute biological ceilings of human performance. As children mature into adults, their brain circuitry is strengthened with use—but they can also be weakened by neglect, injury, illness, or advanced age. Overall, human decision-making is naturally limited by biological constraints. We highlight here physical qualities that can hamper human decision-making.

Memory. Neuroscientists have estimated that human memory has the capacity to store as much as 10^{8432} bits of information—making the human brain a high-capacity storage device.[27] Nevertheless, practical decision-making often depends less on long-term aggregated memory and more on short-term working memory. Typical human working memory is limited to closer to four variables.[28] Decision-makers who attempt to juggle more than about four relevant variables at a time make relatively poor decisions.

Yet for many tasks today, the volume and complexity of modern knowledge exceed humans' working memory.[29] For instance, it is estimated that most of the medical diagnostic errors affecting 12 million adult outpatients per year derive from limits on human memory processing, such as inadequate recollection of patient information and insufficient information recall.[30] One way to overcome the limits on working memory is to rely on an ordinary, nondigital algorithm: a checklist. The World Health Organization, for example, has developed a surgical safety checklist that reduces many key elements of the surgical process into a single page of

---

[27] Yingxu Wang, Dong Liu & Ying Wang, *Discovering the Capacity of Human Memory*, 4 BRAIN & MIND 189, 193–96 (2003). Others have obtained estimates around a billion bits—much lower than 10^{8432} but still substantial. Thomas K. Landauer, *How Much Do People Remember? Some Estimates of the Quantity of Learned Information in Long-Term Memory*, 10 COGNITIVE SCI. 477, 491 (1986).

[28] *See* Nelson Cowan, *The Magical Number 4 in Short-Term Memory: A Reconsideration of Mental Storage Capacity*, 24 BEHAV. BRAIN SCI. 87, 114 (2001). Cowan's work synthesizes a vast literature that usually takes as its starting point George A. Miller, *The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information*, 63 PSYCH. REV. 81 (1956). As extensive commentary published in conjunction with Cowan's article itself indicates, the relevant literature on memory is vast and the issues are complex. We are, by necessity, simplifying issues and distilling relevant research here and throughout our presentation of the various limitations on human decision-making throughout Part I. Although the precise characterization of memory capacity may vary across studies, it is clear that "[t]here are real biological limits to how much information we can process at any given time." LEIDY KLOTZ, SUBTRACT: THE UNTAPPED SCIENCE OF LESS 226 (2021).

[29] *Cf.* ATUL GAWANDE, THE CHECKLIST MANIFESTO 13 (2009) ("[T]he volume and complexity of what we know has exceeded our individual ability to deliver its benefits correctly, safely, or reliably.").

[30] *See* PAUL CERRATO & JOHN HALAMKA, REINVENTING CLINICAL DECISION SUPPORT 1, 6 (2019).

"yes/no" questions.[31] Using checklists has led to significant reductions in morbidity and mortality rates caused by medical error.[32]

Fatigue. Humans perform better when rested. A fatigued individual will be less alert, experience greater difficulties in processing information, have slower reaction times, and suffer from more memory lapses.[33] Fatigue lowers productivity and increases the risk of work-related errors and accidents.[34] Yet workplaces today tend to breed fatigue, often due to shift work.[35] Individuals report being sleepy on an average of three to four days every week.[36] Fatigue and related stresses impair workplace decisions and performance. The National Institute for Occupational Safety and Health notes that "[h]igh levels of fatigue can affect any worker in any occupation or industry with serious consequences for worker safety and health."[37]

Fatigue has been documented to impair human behavior and decision-making in other contexts as well. According to research by the National Transportation Safety Board, 40 percent of highway accidents involve fatigue.[38]

---

[31] *Surgical Safety Checklist*, WORLD HEALTH ORG. (2020), https://apps.who.int/iris/bitstream/handle/10665/44186/9789241598590_eng_Checklist.pdf [https://perma.cc/HZ7C-6TTA].

[32] Eric Nagourney, *Checklist Reduces Deaths in Surgery*, N.Y. TIMES (Jan. 14, 2009), https://www.nytimes.com/2009/01/20/health/20surgery.html?_r=1&ref=health [https://perma.cc/59QW-YQBK] (showing that deaths decline by more than 40 percent and complications by one third).

[33] Paula Alhola & Päivi Polo-Kantola, *Sleep Deprivation: Impact on Cognitive Performance*, 3 NEUROPSYCHIATRIC DISEASE & TREATMENT 553, 553, 556 (2007).

[34] *See, e.g.*, NAT'L SAFETY COUNCIL, CALCULATING THE COST OF POOR SLEEP ~ METHODOLOGY 2 (2017) ("Collectively, costs attributable to sleep deficiency in the U.S. exceeded $410 billion in 2015, equivalent to 2.28% of gross domestic product."); Katrin Uehli, Amar J. Mehta, David Miedinger, Kerstin Hug, Christian Schindler, Edith Holsboer-Trachsler, Jorg D. Leuppi & Nino Kunzli, *Sleep Problems and Work Injuries: A Systematic Review and Meta-Analysis*, 18 SLEEP MED. REV. 61, 61 (2014).

[35] *See, e.g.*, Sarah Kessler & Lauren Hirsch, *Wall Street's Sleepless Nights*, N.Y. TIMES (Mar. 27, 2021), https://www.nytimes.com/2021/03/27/business/dealbook/banker-burnout.html [https://perma.cc/S8AV-LL3A]; Harriet Agerholm, *Amazon Workers Working 55-Hour Weeks and So Exhausted By Targets They 'Fall Asleep Standing Up'*, INDEPENDENT (Nov. 27, 2017), https://www.independent.co.uk/news/uk/home-news/amazon-workers-working-hours-weeks-conditions-targets-online-shopping-delivery-a8079111.html [https://perma.cc/WW9Y-P5RZ].

[36] NAT'L SLEEP FOUND., AMERICANS FEEL SLEEPY 3 DAYS A WEEK, WITH IMPACTS ON ACTIVITIES, MOOD & ACUITY 5 (2020), https://www.sleepfoundation.org/wp-content/uploads/2020/03/SIA-2020-Q1-Report.pdf [https://perma.cc/S4S7-AEYE].

[37] Nat'l Inst. for Occupational Safety & Health, *Work and Fatigue*, CTRS. FOR DISEASE CONTROL & PREVENTION (Jan. 19, 2021), https://www.cdc.gov/niosh/topics/fatigue/default.html [https://perma.cc/PNR2-M3R2].

[38] Jeffrey H. Marcus & Mark R. Rosekind, *Fatigue in Transportation: NTSB Investigations and Safety Recommendations*, 23 INJURY PREVENTION 232, 233 (2017).

Fatigue among orthopedic surgical residents increases risks of medical error by 22 percent.[39]

In the legal system, the treatment that individuals receive also appears to be affected by fatigue-related vagaries of human judgment. One study tracked judicial rulings on parole decisions across three decision sessions, each punctuated by food breaks.[40] At the start of each session, the well-rested judges issued approximately 65 percent favorable decisions on average, which dropped to zero as the judges fatigued.[41] After each food break, the rate reset at 65 percent and the cycle continued.[42]

Aging. As people age, the brain shrinks in volume, and memory and information processing speeds decline.[43] Many older individuals succumb to neurodegenerative disorders, such as Alzheimer's disease or other forms of dementia.[44] Governmental decision-makers are not immune to these effects of aging. Instances have been reported of older judges who could not remember the route to walk out of their own courtrooms, judges who had difficulty reading aloud, judges who seemed to lack memory of previous decisions, and judges who based their decision on nonexistent evidence.[45] Interestingly, although federal judges' health might be scrutinized at the time of their appointments, nothing dictates any routine medical evaluations be conducted for the rest of their careers.[46]

Impulse control. Impulses can have evolutionary advantages in risky situations, but today impulsivity may indicate the symptoms of a range of psychiatric disorders.[47] About 10 percent of the general population is estimated to

---

[39] Frank McCormick, John Kadzielski & Christopher Landrigan, *A Prospective Analysis of the Incidence, Risk, and Intervals of Predicted Fatigue-Related Impairment in Residents*, 147 ARCHIVES SURGERY 430, 433 (2012).

[40] Shai Danziger, Jonathan Levav & Liora Avnaim-Pesso, *Extraneous Factors in Judicial Decisions*, 108 PROC. NAT. ACAD. SCIS. 6889, 6889 (2011).

[41] *Id.* at 6890.

[42] *Id.*

[43] Ruth Peters, *Aging and the Brain*, 82 POSTGRADUATE MED. J. 84, 84 (2006). Of course, age does not affect all individuals the same way. Although information processing speeds tend to decline with age, there exists great variation between individuals in their performance as they age.

[44] *See* Yujun Hou et al., *Ageing as a Risk Factor for Neurodegenerative Disease*, 15 NATURE REV.: NEUROLOGY 565, 565 (2019); ALZHEIMER'S ASS'N, 2021 ALZHEIMER'S DISEASE FACTS AND FIGURES: SPECIAL REPORT: RACE, ETHNICITY AND ALZHEIMER'S IN AMERICA 19 (2021), https://www.alz.org/media/documents/alzheimers-facts-and-figures.pdf [https://perma.cc/V879-HT67].

[45] Joseph Goldstein, *Life Tenure for Federal Judges Raises Issues of Senility, Dementia*, PROPUBLICA (Jan. 18, 2011), https://www.propublica.org/article/life-tenure-for-federal-judges-raises-issues-of-senility-dementia [https://perma.cc/73UW-U7S5].

[46] Francis X. Shen, *Aging Judges*, 81 OHIO ST. L.J. 235, 238–39 (2020).

[47] Such disorders include drug addiction, alcoholism, intermittent explosive disorder, oppositional defiant disorder, and pyromania. T.W. Robbins & J.W. Dalley, *Impulsivity, Risky*

have an impulse control disorder.[48] Attorneys report higher levels of mental health issues such as depression and anxiety, which are not infrequently self-medicated and exacerbated with alcohol or substance abuse, which can then contribute to impulsivity.[49]

Perceptual inaccuracies. Human decisions are affected by mental models of the environment within which individuals act.[50] Given the noisy and chaotic world, if individuals lacked mental models, they would be overwhelmed by the sheer volume of unfiltered information. Humans have thus developed perceptual filters such as selective attention that allows focus on some sensory experiences while tuning out others. But perceptions created from the interaction of different senses can be distorted through the lens of emotions, motivations, desires, and culture. Misperceptions are a common source of mistakes, including those made by governmental actors.[51]

B. Biases

Perhaps in part because of their physical limitations, humans regularly rely on a series of cognitive shortcuts. These shortcuts may reflect traits that have given humans evolutionary advantages. But they can lead to systematic errors in information processing and failures of administrative government.[52] In this Section,

---

Choice, and Impulse Control Disorders: Animal Models, *in* DECISION NEUROSCIENCE: AN INTEGRATIVE APPROACH 81, 81 (Jean-Claude Dreher & Léon Tremblay eds., 2017).

[48] *Table 2. 12-month Prevalence of DSM-IV/WMH-CIDI Disorders by Sex and Cohort*, HARV. MED. SCH., https://www.hcp.med.harvard.edu/ncs/ftpdir/NCS-R_12-month_Prevalence_Estimates.pdf [https://perma.cc/EL3K-GY6T].

[49] A study commissioned by the American Bar Association indicates that more than one third of all attorneys in the United States appear to experience problematic drinking. *Addiction Recovery Poses Special Challenges for Legal Professionals*, BUTLER CTR. FOR RSCH. (Mar. 16, 2017), https://www.hazeldenbettyford.org/education/bcr/addiction-research/substance-abuse-legal-professionals-ru-317 [https://perma.cc/8Y4Q-LZD8].

[50] *See* Daniele Zavagno, Olga Daneyko & Rossana Actis-Grosso, *Mishaps, Errors, and Cognitive Experiences: On the Conceptualization of Perceptual Illusions*, 9 FRONTIERS HUM. NEUROSCIENCE 1, 2 (2015).

[51] For instance, misperceptions can contribute to misidentification of military targets. Eric Schmitt & Anjali Singhvi, *Why American Airstrikes Go Wrong*, N.Y. TIMES (Apr. 14, 2017), https://www.nytimes.com/interactive/2017/04/14/world/middleeast/why-american-airstrikes-go-wrong.html [https://perma.cc/5LZY-BY4H]. They can also undergird conflict and miscommunication in interactions between law enforcement and members of the public. MALCOLM GLADWELL, TALKING TO STRANGERS: WHAT WE SHOULD KNOW ABOUT THE PEOPLE WE DON'T KNOW 342–46 (2019).

[52] *See, e.g.*, Jeffrey J. Rachlinski & Cynthia R. Farina, *Cognitive Psychology and Optimal Government Design*, 87 CORNELL L. REV. 549, 553–54 (2002); Jan Schnellenbach & Christian Schubert, *Behavioral Public Choice: A Survey* 1 (Inst. for Econ. Rsch., Univ. of Freiburg, Working

we detail just a few of the widely documented biases that predictably contribute to errors in human judgment. It may be possible to counteract some of these tendencies through what is known as debiasing—but not always and not necessarily completely.[53]

Availability Heuristic. The availability heuristic or bias refers to the human tendency to treat examples which most easily come to mind as the most important information or the most frequent occurrences.[54] When a hazard is particularly salient or frequently observed, the hazard is more cognitively available and tends to drive decision-making. In the context of legislation and agency decision-making, policy decisions will inevitably become anecdote-driven if preferences are shaped by a set of probability judgments that are themselves affected by the availability bias. For instance, support for government regulation can be driven by recent and memorable instances of harms, such as explosions, fires, or other crises.

Confirmation Bias. Confirmation bias—sometimes referred to as motivated reasoning—is the tendency to search for and favor information that confirms existing beliefs, while simultaneously ignoring or devaluing information that contradicts them.[55] In one study, individuals in two groups—one supportive of capital punishment, the other not—were given purported evidence from the same two fictional studies—one supporting their views, one undermining them.[56] The participants merely ignored the information undermining their preexisting beliefs and focused on what confirmed their initial positions.[57]

Government officials are not immune from motivated reasoning. In a recent Danish study, elected politicians were shown the characteristics of two schools, and

---

Paper No. 14/03, 2014); George Dvorsky, *The 12 Cognitive Biases that Prevent You from Being Rational*, GIZMODO (Jan. 9, 2013), http://io9.com/5974468/the-most-common-cognitive-biases-that-prevent-you-from-being-rational [https://perma.cc/E5YG-75DY]. This is not to say, of course, that these biases always lead to problems. Gerd Gigerenzer, *Heuristics*, *in* HEURISTICS AND THE LAW 17, 40–41 (Gerd Gigerenzer & C. Engel eds., 2006).

[53] Christine Jolls & Cass R. Sunstein, *Debiasing Through Law*, J. LEGAL STUD. 199, 200–02 (2006). A legal requirement that corporate boards include outside members is one example of a debiasing strategy, as it tries to counteract confirmation bias.

[54] Amos Tversky & Daniel Kahneman, *Judgment Under Uncertainty: Heuristics and Biases*, 185 SCIENCE 1124, 1127 (1974) [hereinafter Tversky & Kahneman, *Judgment Under Uncertainty*].

[55] Charles G. Lord, Lee Ross & Mark R. Lepper, *Biased Assimilation and Attitude Polarization: The Effects of Prior Theories on Subsequently Considered Evidence*, 37 J. PERSONALITY & SOC. PSYCHOL. 2098, 2098 (1979).

[56] *Id.*

[57] *Id.* Some research even suggests that as humans acquire domain expertise, they can lose flexibility with regard to problem solving, adaptation, and creative idea generation. Erik Dane, *Reconsidering the Trade-off Between Expertise and Flexibility: A Cognitive Entrenchment Perspective*, 35 ACAD. MGMT. REV. 579, 579 (2010).

asked to choose the better-performing one.[58] When the schools were labeled anonymously, the politicians' answers coalesced around the better-performing school.[59] But when the schools were labeled by their ownership status (i.e., private versus public), the results changed dramatically—because the privatization of schools was a contentious policy issue in Denmark at the time of the study.[60] The additional information led the politicians to cherry-pick evidence that supported their preexisting beliefs and entrenched values. Overall, decision-makers who have established an initial position on an issue—such as by making a speech or other public statement—tend to be less likely to make use of new evidence that might depart from their staked-out views.

Anchoring. Much as with the differences in labeling of options, decisions can be shaped by anchored information.[61] People estimate unknowns by modifying an initial value—whether explicitly given or implicitly in the subconscious—to reach the final answer.[62] Although anchoring typically affects decisions in negotiation, it also can affect how voters evaluate the costs of government programs. One study found that when asked how much they believed a public project would raise their taxes, the majority of participants anchored their estimate according to the number embedded in the question itself (that is, they gave higher estimates in response to mention of a project's need for "$50,000,000" in financing, as opposed to an equivalent "$130 per capita").[63]

System Neglect. Individuals can mistake trees for forests, overweighting individual signals relative to the underlying system which generates these signals.[64] As a result, decisions can be made in isolation, with insufficient regard to the systemic context of the decision. If the environment within which a decision is being made is stable, humans tend to overreact to noisy signals; if the environment is unstable, humans tend to underreact to precise.[65] In general, human decision-makers tend to pay more attention to individual bits of information than to the general system producing

---

[58] Martin Baekgaard, Julian Christensen, Casper Mondrup Dahlmann, Asbjørn Mathiasen & Niels Bjørn Grund Petersen, *The Role of Evidence in Politics: Motivated Reasoning and Persuasion Among Politicians*, 49 BRIT. J. POL. SCI. 1117, 1124 (2019).

[59] *Id.* at 1125.

[60] *Id.* at 1127.

[61] Tversky & Kahneman, *Judgment Under Uncertainty*, *supra* note 54, at 1128.

[62] *See id.*

[63] KENNETH A. KRIZ, ANCHORING AND ADJUSTMENT BIASES AND LOCAL GOVERNMENT REFERENDA LANGUAGE 9, 14 (2014), https://www.ntanet.org/wp-content/uploads/proceedings/2014/078-kriz-anchoring-adjustment-biases-local-government.pdf [https://perma.cc/E6FR-WDCT].

[64] Cade Massey & George Wu, *Detecting Regime Shifts: The Causes of Under- and Overreaction*, 51 MGMT. SCI. 932, 933 (2005).

[65] *Id*. at 945; Mirko Kremer, Brent Moritz & Enno Siemsen, *Demand Forecasting Behavior: System Neglect and Change Detection*, 57 MGMT. SCI. 1827, 1838 (2011).

the information. They can also take into account factors that are outside the affected system and not necessarily relevant to the decision at hand.

Present bias. Individuals tend to resist change. Part of this stems from an endowment effect by which people tend to value retaining their current situation more than gaining an equivalent situation that they do not currently possess. As one well-known study found, participants demanded much more to give up a Cornell University coffee mug than they would be willing to pay to acquire the same mug in the first place.[66] Other research shows that people are more likely to recall the positive attributes of what they possess, focusing on reasons to keep what they already have, while they are more likely to recall the negative attributes of what they do not possess, focusing on the reasons not to buy into change.[67]

A related behavioral tendency, known as loss aversion, also reinforces a present bias. Humans dislike losses more than they like corresponding gains.[68] People also tend to disregard potential gains and focus on the losses associated with an activity. Overall, they face challenges in assessing risks, with difficulties in processing and assigning meaning to probabilities, large numbers, and exponential growth.[69] Subtle changes in the framing of information can affect people's evaluation of risks—even ones that are quantitatively identical. For example, if a health policy is framed in terms of number of lives saved, people are more conservative and risk-averse; if the same policy is framed in terms of number of lives lost, people are much more willing to take risks to try to reduce that number.[70]

Practically speaking, these tendencies help explain why "preventing losses . . . looms larger in government's objective function."[71] Governments are less likely to

---

[66] Daniel Kahneman, Jack L. Knetsch & Richard H. Thaler, *Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias*, 5 J. ECON. PERSPS. 193, 196 (1991).

[67] Michael A. Strahilevitz & George Loewenstein, *The Effect of Ownership History on the Valuation of Objects*, 25 J. CONSUMER RSCH. 276, 285 (1998).

[68] Daniel Kahneman & Amos Tversky, *An Analysis of Decision Under Risk*, 47 ECONOMETRICA 263, 266 (1979).

[69] For a useful collection of essays on this general problem, see generally NUMBERS AND NERVES: INFORMATION, EMOTION, AND MEANING IN A WORLD OF DATA (Scott Slovic & Paul Slovic eds., 2015). People also tend to engage in hyperbolic discounting, preferring immediate rewards to future ones of equal present value. *See, e.g.*, J.D. Trout, *The Psychology of Discounting: A Policy of Balancing Biases*, 21 PUB. AFF. Q. 201, 204 (2007); Jess Benhabib, Alberto Bisin & Andrew Schotter, *Present-Bias, Quasi-Hyperbolic Discounting, and Fixed Costs*, 69 GAMES & ECON. BEHAV. 205, 222 (2010).

[70] Alexander J. Rothman & Peter Salovey, *Shaping Perceptions To Motivate Healthy Behavior: The Role of Message Framing*, 121 PSYCH. BULL. 3, 4–5 (1997).

[71] Caroline Freund & Çağlar Özden, *Trade Policy and Loss Aversion*, 98 AM. ECON. REV. 1675, 1675 (2008); *see also* Robert Jervis, *Political Implications of Loss Aversion*, 13 POL. PSYCH. 187, 187 (1992) [hereinafter Jervis, Political Implications] ("People are loss-averse in the sense that losses loom larger than the corresponding gains."); Jean Galbraith, *Treaty Options: Towards a Behavioral Understanding of Treaty Design*, 53 VA. J. INT'L L. 309, 350, 355 (2013) ("Individuals tend to weigh losses more than gains in decision-making, and so may weigh the risks of switching from a default option more heavily than the possible gains.").

behave aggressively when doing so would produce gains than when the same behavior might prevent losses.[72] Policymakers appear to prefer more cautious measures over more ambitious ones, even if the latter are needed. Such myopia in governmental decision-making creates inertia toward the status quo.[73] It can also lead to the systemic underinvestment in policies and resources needed to prevent future harms.[74]

Susceptibility to Overpersuasion. Although humans may at times be sensibly persuaded to change their opinions when presented with new information, they can also be persuaded by superficial, even irrelevant, changes in environment, context, and framing. Numerous studies reveal biases that can be triggered simply by subtle changes in language or visual imagery.[75] For example, one study found that changes in the gruesomeness of information correlate with conviction rates.[76] With other variables held constant, 34 percent of the subjects who viewed gruesome textual references chose to convict, compared with 14 percent by those who did not.[77] Even delivering unsavory smells to a room can affect decisions having nothing to do with odors.[78]

The visual display of information can also be more influential than expected. Visually gripping demonstratives, such as diagrams, photographs, and animations, can captivate a jury's attention, spark emotions, and prove persuasive.[79] Merely using PowerPoint slides in opening statements in court tends to correspond with achieving more favorable decisions.[80]

---

[72] *See* Jervis, *Political Implications*, *supra* note 71.

[73] Those who stand to lose from new policies often mobilize more strenuously against change than will those who stand to gain. *See* THE POLITICS OF REGULATION 360–61 (James Q. Wilson ed., 1982).

[74] *See* ROBERT MEYER & HOWARD KUNREUTHER, THE OSTRICH PARADOX: WHY WE UNDERPREPARE FOR DISASTERS 2–4 (2017).

[75] *See, e.g.*, Elizabeth F. Loftus & John C. Palmer, *Reconstruction of Automobile Destruction: An Example of the Interaction Between Language and Memory*, 13 J. VERBAL LEARNING & VERBAL BEHAV. 585, 585 (1974); Elizabeth F. Loftus & Guido Zanni, *Eyewitness Testimony: The Influence of the Wording of a Question*, 5 BULL. PSYCHONOMIC SOC'Y 86, 86 (1975). For a review, see Lai, *supra* note 23, at 4.

[76] *See, e.g.*, David A. Bright & Jane Goodman-Delahunty, *Gruesome Evidence and Emotion: Anger, Blame, and Jury Decision-Making*, 30 LAW & HUM. BEHAV. 183, 183 (2006); Beatrice H. Capestany & Lasana T. Harris, *Disgust and Biological Descriptions Bias Logical Reasoning During Legal Decision-Making*, 9 SOC. NEUROSCIENCE 265, 265 (2014).

[77] David A. Bright & Jane Goodman-Delahunty, *The Influence of Gruesome Verbal Evidence on Mock Juror Verdicts*, 11 PSYCHIATRY, PSYCHOLOGY & LAW 154, 154 (2004).

[78] Nicolao Bonini, Constantinos Hadjichristidis, Ketti Mazzocco, Maria Luisa Demattè, Massimiliano Zampini & Andrea Sbarbati, Stefano Magon, Pecunia Olet*: The Role of Incidental Disgust in the Ultimatum Game*, 11 EMOTION 965, 965 (2011).

[79] Lai, *supra* note 23, at 10.

[80] Jaihyun Park & Neal Feigenson, *Effects of a Visual Technology on Mock Juror Decision Making*, 27 APPLIED COGNITIVE PSYCH. 235, 235 (2012). For specific examples, see, e.g., *In re* Pers. Restraint of

Racial and Gender Biases. As with the various cognitive biases noted above, race and gender biases can affect human judgment—even without conscious animus. Implicit biases are a "distorting lens that's a product of both the architecture of our brain and the disparities in our society."[81]

Perceptions about race can be shaped by subtle cues that appear in people's surroundings. One study exposed adult subjects to a series of flashes of light containing letters that were too rapid to be consciously perceived.[82] One group was exposed to flashes with words related to crime, such as "arrest" and "shoot," while the other group was exposed to jumbled letters.[83] But after these flashes, subjects were shown two human faces simultaneously—one Black, one white. The subjects exposed to the crime-related words spent more time staring at the Black face.[84]

In the context of the legal system, studies show evidence of racial bias in the conduct of prosecutors in determining convictions[85] and federal sentences.[86] Racial disparities have been identified as well in the decisions of defense attorneys,[87] police officers,[88] judges,[89] and juries.[90] Similarly, racial disparities can be found in the policies and products of the administrative state. Some of these disparities have resulted from explicit biases reflected in historic race-conscious

---

Glasmann, 286 P.3d 673, 701–03 (Wash. 2012) (en banc) and State v. Robinson, No. 47398-1-I, 2002 WL 258038, at *3 (Wash. Ct. App. Feb. 25, 2002).

[81] JENNIFER L. EBERHARDT, BIASED: UNCOVERING THE HIDDEN PREJUDICE THAT SHAPES WHAT WE SEE, THINK, AND DO 6 (2019); *see also* O. Pascalis, L. S. Scott, D. J. Kelly, R. W. Shannon, E. Nicholson, M. Coleman & C. A. Nelson, *Plasticity of Face Processing in Infancy*, 102 PROC. NAT. ACAD. SCIS. 5297, 5300 (2005) ("[E]xperience with faces early in life may influence and shape the development of a face prototype. The development of this prototype leads to biases in discriminating own-race and own-species faces compared with other-race and other-species faces.").

[82] EBERHARDT, *supra* note 81, at 58–60.

[83] *Id.*

[84] *Id.*

[85] Carly W. Sloan, Racial Bias by Prosecutors: Evidence from Random Assignment 30 (2019) (unpublished manuscript) (on file with author).

[86] M. Marit Rehavi & Sonja B. Starr, *Racial Disparity in Federal Criminal Sentences*, 122 J. POL. ECON. 1320, 1320 (2014).

[87] *See, e.g.*, David S. Abrams & Albert H. Yoon, *The Luck of the Draw: Using Random Case Assignment To Investigate Attorney Ability*, 74 U. CHI. L. REV. 1145, 1145 (2007). *See also* Jeff Adachi, *Public Defenders Can Be Biased, Too, and It Hurts Their Non-White Clients*, WASH. POST (June 7, 2016), https://www.washingtonpost.com/posteverything/wp/2016/06/07/public-defenders-can-be-biased-too-and-it-hurts-their-non-white-clients/.

[88] Kate Antonovics & Brian G. Knight, *A New Look at Racial Profiling: Evidence from the Boston Police Department*, 91 REV. ECON. & STAT. 163, 163 (2009).

[89] *See* Briggs Depew, Ozkan Eren & Naci Mocan, *Judges, Juveniles, and In-Group Bias*, 60 J.L. & ECON. 209, 209 (2017) (finding evidence of negative in-group bias by judges sentencing juvenile offenders).

[90] *See* Shamena Anwar, Patrick Bayer & Randi Hjalmarsson, *The Impact of Jury Race in Criminal Trials*, 127 Q.J. ECON. 1017, 1017 (2012) (finding that all-white juries convict Black defendants 16 percent more frequently than they convict white defendants).

housing policies.[91] Others result from persistent structural racism and implicit biases.[92] Awards of Social Security disability benefits, for example, have been found to exhibit racial disparities, with Black applicants receiving less favorable outcomes compared to white applicants.[93] The Food and Drug Administration's ("FDA") testing protocols result in Black and Latino individuals disproportionately bearing the risks of testing experimental drugs.[94]

C. Group Challenges

To these various problems and limitations of individual decision-making can be added a series of distinctive pathologies associated with group decision-making—the kind of decision-making that prevails throughout much of government.[95] The group setting does not necessarily eliminate the problematic physical and cognitive features that can make individual decision-making go awry. On the contrary, research indicates that it is commonly the case that "groups exaggerate this tendency."[96] Moreover, the group setting adds social dynamics that can create additional problems. It was far from accidental that Otto von Bismarck compared the making of laws to the making of sausages. Some research even

[91] *See, e.g.*, RICHARD ROTHSTEIN, THE COLOR OF LAW: A FORGOTTEN HISTORY OF HOW OUR GOVERNMENT SEGREGATED AMERICA 39 (2017) (explaining how, in the mid-twentieth century especially, "federal, state, and local governments purposely created segregation in every metropolitan area of the nation"); JESSICA TROUNSTINE, SEGREGATION BY DESIGN: LOCAL POLITICS AND INEQUALITY IN AMERICAN CITIES 3 (2018) (noting how segregation emerged from "local governments systematically institutionaliz[ing] discriminatory approaches to the maintenance of housing values and production of public goods").

[92] The racial makeup of the heads of many administrative agencies has also failed to reflect society's racial makeup. Chris Brummer, *What Do the Data Reveal About (the Absence of Black) Financial Regulators?* 8–9 (Brookings Econ. Stud., Working Paper, 2020), https://www.brookings.edu/research/what-do-the-data-reveal-about-the-absence-of-black-financial-regulators [https://perma.cc/LZ4Q-4TUS].

[93] U.S. GEN. ACCT. OFF., GAO/HRD-92-56, SOCIAL SECURITY: RACIAL DIFFERENCE IN DISABILITY DECISIONS WARRANTS FURTHER INVESTIGATION 4 (1992); Erin M. Godtland, Michele Grgich, Carol Dawn Petersen, Douglas M. Sloane & Ann T. Walker, *Racial Disparities in Federal Disability Benefits*, 25 CONTEMP. ECON. POL'Y 27, 27 (2007).

[94] *See* JILL A. FISHER, ADVERSE EVENTS: RACE, INEQUALITY, AND THE TESTING OF NEW PHARMACEUTICALS 4 (2020).

[95] *See, e.g.*, David P. Redlawsk & Richard R. Lau, *Behavioral Decision-Making*, *in* OXFORD HANDBOOK OF POLITICAL PSYCHOLOGY 1, 1–4 (Leonie Huddy, David O. Sears & Jack S. Levy eds., 2d ed. 2013) (discussing the behavioral tendencies of voters' decision-making).

[96] Verlin B. Hinsz, R. Scott Tindale & David A. Vollrath, *The Emerging Conceptualization of Groups as Information Processors*, 121 PSYCH. BULL. 43, 49 (1997).

suggests that as many as half of all decisions made in a group setting result in failure.[97]

Groupthink. One dynamic that arises within groups derives from individuals' psychological drive for consensus that suppresses dissent and the open appraisal of alternatives.[98] When members of a group prize their group membership over the substance of the decision, any individual motivation to appraise alternative courses of action tends to fall to the wayside.[99] Individual doubts and disagreements are effectively censored.[100] Structural faults in the organization—including partial leadership, rigidly established procedures, and homogenous in-groups—tend to lead to harmful symptoms of groupthink, including the illusion of invulnerability, self-censorship, stereotypes of out-groups, and the illusion of unanimity.[101] Groupthink is said to have contributed to a wide range of governmental failures, including the National Aeronautics and Space Administration's fateful decision to launch the Challenger shuttle, President Harry S. Truman's troubled invasion of the Democratic People's Republic of Korea, President John F. Kennedy's failed assault on Bay of Pigs in Cuba, President Lyndon B. Johnson's escalation of U.S. involvement in the Vietnam War, President Richard Nixon's Watergate scandal, and the coverup of the Iran-Contra scandal during the administration of President Ronald Reagan.[102] More recently, groupthink appears to have impeded the

---

[97] *See* Paul C. Nutt, *Surprising but True: Half the Decisions in Organizations Fail*, 13 ACAD. MGMT. EXEC. 75, 75 (1999).

[98] *See* IRVING JANIS, VICTIMS OF GROUPTHINK 3 (1972). The term was coined in George Orwell's *1984* and first used to describe "rationalized conformity" in government organizations. William H. Whyte, Jr., *Groupthink*, FORTUNE (1952), https://fortune.com/2012/07/22/groupthink-fortune-1952 [https://perma.cc/JCN7-Z63E].

[99] Whyte, *supra* note 98.

[100] *Id.*

[101] *Id.*

[102] *See* IRVING L. JANIS, CRUCIAL DECISIONS: LEADERSHIP IN POLICYMAKING AND CRISIS MANAGEMENT 47, 57–58 (1989); EM GRIFFIN, A FIRST LOOK AT COMMUNICATION THEORY 219–28 (1991). For related discussion, see REPORT BY THE PRESIDENTIAL COMMISSION ON THE SPACE SHUTTLE CHALLENGER ACCIDENT 83–119 (1986) (chronicling flawed group decision-making that led to the catastrophic launch of the Challenger space shuttle), https://science.ksc.nasa.gov/shuttle/missions/51-l/docs/rogers-commission/Rogers_Commission_Report_Vol1.pdf [https://perma.cc/P29Y-SQ2D] and RICHARD E. NEUSTADT & ERNEST R. MAY, THINKING IN TIME: THE USES OF HISTORY FOR DECISION-MAKERS 32–33 (1986) (analyzing examples of failed group decision-making across multiple federal administrations). We recognize, of course, that groupthink may not always be the sole driver of organizational failure. *Cf.* DIANE VAUGHAN, THE CHALLENGER LAUNCH DECISION: RISKY TECHNOLOGY, CULTURE, AND DEVIANCE AT NASA 404 (2d ed. 2016) (arguing that "many of the elements of" failure in the Challenger tragedy "have explanations that go beyond the assembled group to cultural and structural sources"). Even Janis recognized that, in situations suffering from groupthink, "other causal factors" may well be at play. JANIS, *supra*, at 275.

government's response to COVID-19[103] and may have contributed to misjudging the circumstances surrounding a U.S. withdrawal from Afghanistan.[104]

Lowest Common Denominator Effect. Another related group decision-making pathology grows out of individual group members' desire to come to agreement: the lowest common denominator effect. When this happens, decisions get made based on what all the group members can agree upon, rather than on what might actually be needed to address the problem confronting the group.[105] Groups that succumb to the lowest common denominator risk setting the bar too low when making decisions and finding solutions.[106]

Garbage Can Decision-making. Groups may also fail to find effective solutions because of "garbage can" decision-making, where group members first identify solutions and search for problems which might justify their preferred solutions—rather than the reverse.[107] Whether a group makes useful choices depends upon a mixture of ideas for solutions, problems to be solved, and decision-makers involved in the group. Often, members of a group will avoid identifying problems in the effort to make decisions. Ultimately, "the nature of the choice, the time [the group] takes, and the problems it solves all depend on a relatively complicated intermeshing of elements" within the organization.[108]

Preference Cycling. Aggregating preferences within groups can also be relatively erratic. According to Arrow's impossibility theorem, when individual

---

[103] Richard Coker, *Coronavirus Can Only Be Beaten If Groups Such as Sage Are Transparent and Accountable*, GUARDIAN (Apr. 27, 2020), https://www.theguardian.com/commentisfree/2020/apr/27/coronavirus-sage-scientific-groupthink [https://perma.cc/NNR4-K3DB]; *see also* Howard Kunreuther & Paul Slovic, *Learning from the COVID-19 Pandemic To Address Climate Change*, MGMT. & BUS. REV. (Winter 2021), https://mbrjournal.com/2021/01/26/learning-from-the-covid-19-pandemic-to-address-climate-change [https://perma.cc/52FV-EG3G] (noting how a "tend[ency] to *follow the herd*, allowing [their] choices to be influenced by other people's behavior, especially when we feel uncertain" influenced key decision-makers' responses to the COVID-19 pandemic).

[104] *See* Tevi Troy, *All the President's Yes-Men*, WALL ST. J. (Aug. 22, 2021), https://www.wsj.com/articles/president-decision-making-biden-kennedy-johnson-taliban-afghanistan-bay-of-pigs-vietnam-saigon-blinkin-sullivan-11629641380 [https://perma.cc/4Y7G-HTEU].

[105] For a discussion of the lowest common denominator effect, see Cary Coglianese, *Is Consensus an Appropriate Basis for Regulatory Policy?*, *in* ENVIRONMENTAL CONTRACTS: COMPARATIVE APPROACHES TO REGULATORY INNOVATION IN THE UNITED STATES AND EUROPE 93, 93–113 (Eric Orts & Kurt Deketelaere eds., 2001).

[106] For ways that groups can fail by trying to make everyone happy, see Cary Coglianese, *Is Satisfaction Success? Evaluating Public Participation in Regulatory Policymaking*, *in* THE PROMISE AND PERFORMANCE OF ENVIRONMENTAL CONFLICT RESOLUTION 69, 69–70 (Rosemary O'Leary & Lisa Bingham eds., 2003).

[107] Michael D. Cohen, James G. March & Johan P. Olsen, *A Garbage Can Model of Organizational Choice*, 17 ADMIN. SCI. Q. 1, 1 (1972).

[108] *Id.* at 16.

preferences are arrayed across more than a single dimension, there may be no clear and stable way to aggregate individual preferences without violating mathematical principles of transitivity.[109] In other words, although a majority of a group may favor option A over option B, and also favor option B over option C, if the group is faced just with a decision that involves a choice just between A and C, it may well rationally choose C. Outcomes "cycle" because the choice that satisfies a majority of group members' preferences can shift depending on the potentially arbitrary way that alternatives are pitted against each other (A versus B, B versus C, or A versus C).[110]

Free Riding or Social Loafing. Members of a group can be less motivated when performing tasks among other group members. Situations when individuals are not individually identifiable lead to lower accountability and responsibility—or social loafing. In one experiment, researchers found that when asked to perform physically exerting tasks of clapping and shouting, participants' efforts sizably decreased when performing in groups as compared to performing alone.[111] This effect also has been documented in industrial production, bystander intervention, and participation in church activities.[112] When individuals need to work cooperatively to achieve collective action, they have the incentive to free ride on the efforts of others—which ultimately undersupplies needed collective goods.[113]

* * *

All of these various characteristics of human decision-making manifest themselves in public policies and outcomes, fueling frequent complaints about government and its performance.[114] When calamities strike, government officials

---

[109] Kenneth J. Arrow, *A Difficulty in the Concept of Social Welfare*, 58 J. POL. ECON. 328, 334–39, 342–43 (1950).

[110] For an accessible introduction to preference cycling, see DANIEL A. FARBER & PHILIP P. FRICKEY, LAW AND PUBLIC CHOICE: A CRITICAL INTRODUCTION 38–39 (1991).

[111] Bibb Latané, Kipling Williams & Stephen Harkins, *Many Hands Make Light the Work: The Causes and Consequences of Social Loafing*, 37 J. PERSONALITY & SOC. PSYCH. 822, 822 (1979).

[112] *Id.* at 831.

[113] MANCUR OLSON, THE LOGIC OF COLLECTIVE ACTION: PUBLIC GOODS AND THE THEORY OF GROUPS 16–22 (1965).

[114] As of 2021, 73 percent of Americans were at least somewhat dissatisfied with government and "how well it works." *Government*, GALLUP, https://news.gallup.com/poll/27286/government.aspx [https://perma.cc/R8U2-7SG7]. Between 2001, when the level was 30 percent, and 2021, dissatisfaction more than doubled. *Id.* For discussions of the effects of human limitations on governmental performance, see, e.g., Eyal Zamir & Raanan Sulitzeanu-Kenan, *Explaining Self-Interested Behavior of Public-Spirited Policy Makers*, 78 PUB. ADMIN. REV. 579, 579 (2017); Michael David Thomas, *Reapplying Behavioral Symmetry: Public Choice and Choice Architecture*, 180 PUB. CHOICE 1, 11 (2019).

receive blame for failing to connect the dots and prevent tragedy from occurring.[115] When problems go unsolved, government again gets blamed, often for being too sclerotic.[116] All along, the persistence of racial, gender, and other biases continue to raise questions about the fairness of government.[117]

Even routine administrative processes driven by humans receive frequent criticisms about delays, inconsistencies, and disparities.[118] Any system that must rely on thousands of humans to make decisions will necessarily be susceptible to such concerns. The Social Security Administration's ("SSA") disability program, for example, depends on about 1500 administrative law judges to process about eight hundred thousand cases each year.[119] Even with this processing throughput, the disability claims system has a backlog of about a million cases.[120] Moreover, inconsistencies across this system's many human decision-makers can be stark.[121] Using just the fifteen most active administrative judges in the Dallas SSA as an example, it has been noted that "the judge grant rates in this single location ranged . . . from less than 10 percent being granted to over 90 percent."[122] Three judges

---

[115] *See, e.g.*, Christopher Carrigan & Cary Coglianese, *Oversight in Hindsight: Assessing the U.S. Regulatory System in the Wake of Calamity*, *in* REGULATORY BREAKDOWN: THE CRISIS OF CONFIDENCE IN U.S. REGULATION 1, 1–6 (Cary Coglianese ed., 2012) (providing examples of calamities that yielded complaints about regulatory decisions).

[116] *See, e.g.*, JONATHAN RAUCH, DEMOSCLEROSIS: THE SILENT KILLER OF AMERICAN GOVERNMENT 17–20 (1994).

[117] *See, e.g.*, Lucie E. White, *Subordination, Rhetorical Survival Skills, and Sunday Shoes: Notes on the Hearing of Mrs. G.*, 38 BUFF. L. REV. 1, 2 (1990); DOROTHY ROBERTS, SHATTERED BONDS: THE COLOR OF CHILD WELFARE 92–99 (2002); HEATHER MCGHEE, THE SUM OF US: WHAT RACISM COSTS EVERYONE AND HOW WE CAN PROSPER TOGETHER 17–40 (2021).

[118] *See, e.g.*, HAROLD J. KRENT & SCOTT MORRIS, ACHIEVING GREATER CONSISTENCY IN SOCIAL SECURITY DISABILITY ADJUDICATION: AN EMPIRICAL STUDY AND SUGGESTED REFORMS 1 (2013); Paul Verkuil, *Meeting the Mashaw Test for Consistency in Administrative Decision-Making*, *in* ADMINISTRATIVE LAW FROM THE INSIDE OUT: ESSAYS ON THEMES IN THE WORK OF JERRY L. MASHAW 239, 239–40 (Nicholas R. Parillo ed., 2017); Aaron Glantz, *For Disabled Veterans Awaiting Benefits Decisions, Location Matters*, PBS NEWSHOUR EXTRA (Mar. 6, 2014), https://www.pbs.org/newshour/extra/app/uploads/2014/03/DisabledVetsWaitingForBenefits.pdf [https://perma.cc/65WV-2PPB]; U.S. GOV'T ACCOUNTABILITY OFF., GAO-16-74, ENERGY EMPLOYEES COMPENSATION: DOL GENERALLY FOLLOWED ITS PROCEDURES TO PROCESS CLAIMS BUT COULD STRENGTHEN SOME INTERNAL CONTROLS 10 (2016); U.S. GOV'T ACCOUNTABILITY OFF., BLACK LUNG BENEFITS PROGRAM: ADMINISTRATIVE AND STRUCTURAL CHANGES COULD IMPROVE MINERS' ABILITY TO PURSUE CLAIMS 10 (2009). For discussion of how delays may sometimes be purposeful, see generally DONALD MOYNIHAN & PAMELA HERD, ADMINISTRATIVE BURDEN: POLICYMAKING BY OTHER MEANS (2018).

[119] Verkuil, *supra* note 118, at 242.

[120] *Id.*

[121] *See, e.g.*, KRENT & MORRIS, *supra* note 118.

[122] TRANSACTIONAL RECS. ACCESS CLEARINGHOUSE, SOCIAL SECURITY AWARDS DEPEND MORE ON JUDGE THAN FACTS (July 4, 2011) [hereinafter SSA REPORT],

awarded benefits to no more than 30 percent of applicants, while three other judges awarded to more than 70 percent.[123]

Physical limitations, cognitive biases, and group pathologies build upon one another to affect human decision-making in ways that can be unpredictable and often undesirable. They contribute—both separately and in combination—to the widely accepted conclusion that government performs poorly.[124]

II. THE PROMISE OF DIGITAL ALGORITHMS

Recognizing the limitations of human decision-making should make both public officials and the public open to the possibility that digital algorithms—whether in the form of simple automation tools or complex machine-learning algorithms—could help improve government's performance.[125] In this Part, we show the promise that digital algorithms hold for making such improvements. We begin by articulating some of digital algorithms' general virtues—especially the virtues of machine-learning algorithms—and then turn to research directly comparing their performance with the status quo.

Although this research confirms that machine-learning algorithms can deliver considerable improvements, we do not claim that digital algorithms will always perform better than current human algorithms. Machine learning cannot eliminate every problem confronting government. The relevant question about digital algorithms is not whether they will be free of all errors or biases. Rather, the question should be whether digital algorithms can perform specific tasks better than humans. Anyone concerned about fairness in government decision-making should

---

https://trac.syr.edu/tracreports/ssa/254 [https://perma.cc/GTR8-6U8N]. The SSA disputed aspects of this study. But others have documented considerable variability in SSA administrative outcomes. *See, e.g.*, Verkuil, *supra* note 118, at 242; KRENT & MORRIS, *supra* note 118.

[123] SSA REPORT, *supra* note 122.

[124] For recent discussions of infirmities in governmental performance, see, e.g., FRANCIS FUKUYAMA, POLITICAL ORDER AND POLITICAL DECAY: FROM THE INDUSTRIAL REVOLUTION TO THE GLOBALIZATION OF DEMOCRACY 484–505 (2014), BO ROTHSTEIN, THE QUALITY OF GOVERNMENT 1–6 (2011), and PETER H. SCHUCK, WHY GOVERNMENT FAILS SO OFTEN: AND HOW IT CAN DO BETTER 30 (2014). Of course, recognizing infirmities is not to deny that government can and does sometimes work well. Scott Douglas et al., *Rising to Ostrom's Challenge: An Invitation To Walk on the Bright Side of Public Governance and Public Service*, 4 POL'Y DESIGN & PRAC. 1, 1 (2021); Cary Coglianese, *Is Government Really Broken?*, 1 U. PA. J.L. & PUB. AFFS. 65, 66–68 (2016).

[125] In fact, federal, state, and local governmental entities have already begun to implement digital algorithms in various ways to support domestic public administration, especially for tedious, voluminous tasks and to parse through data to extract patterns. *E.g.*, Coglianese & Ben Dor, *supra* note 8, at 23–37; ENGSTROM ET AL., *supra* note 9, at 9–11; Kevin C. Desouza, Artificial Intelligence in the Public Sector: A Maturity Model 7–8 (2021), https://www.businessofgovernment.org/sites/default/files/Artificial%20Intelligence%20in%20the%20Public%20Sector_0.pdf.

entertain the possibility that digital algorithms may sometimes prove to be fairer and more consistent than humans. At the very least, it might be easier to remedy biased algorithms than to remove deeply ingrained implicit or cognitive biases from human decision-making.[126]

Nevertheless, because the design and operation of digital algorithms depend on humans, public officials should approach their use with due care. We therefore conclude this Part by highlighting a few problems and controversies that have arisen when governments have shifted to a reliance on digital algorithms. By appreciating that risks remain with the use of digital algorithms, it becomes evident that government officials need to be suitably cautious and make smart decisions about when and how to choose digital versus human algorithms, the issue Part III takes up.

A. Digital Algorithms and Their Virtues

Statistical and other mathematical algorithms have been pivotal to nearly every major advance in science and technology. In recent decades, major developments in computing power now allow business leaders, medical and other professionals, and government officials to improve what they do by taking advantage of a distinctive type of digital algorithm known as a machine-learning algorithm.

Machine-learning algorithms learn autonomously by deciphering patterns and generating inferences in large datasets that contain images, numbers, dense text, and natural languages.[127] These algorithms can assume many different forms, but are often grouped into two main categories.[128] In the "supervised learning" category, algorithms are provided with numerous labeled examples—for example, images categorized as "dog" or "cat"—and then generate a model to identify unlabeled images of dogs and cats. By contrast, in "unsupervised learning," algorithms can learn without the benefit of labeled data. When an unsupervised learning algorithm is fed an increasing number of images of dogs and cats, it builds predictive models for how to distinguish the two.[129]

---

[126] *See generally* MICHAEL KEARNS & AARON ROTH, THE ETHICAL ALGORITHM: THE SCIENCE OF SOCIALLY AWARE ALGORITHM DESIGN (2019) (discussing ways that digital science can incorporate adherence to ethical principles into machine-learning technologies).

[127] Coglianese & Lehr, *Regulating by Robot*, *supra* note 1, at 1156–57; Lehr & Ohm, *supra* note 1, at 655.

[128] Our discussion of machine learning here is, by necessity, both brief and basic, and machine-learning algorithms can fall into additional categories, such as semi-supervised and reinforcement learning algorithms.

[129] Coglianese & Lehr, *Regulating by Robot*, *supra* note 1, at 1158 n.37.

Unlike traditional statistical analysis techniques, machine learning does not require humans to specify at the outset which variables to use.[130] Of course, humans are never truly and completely out of the picture, as they must still select the learning algorithm's objective and meta-design, feed it its training data, and tweak the algorithm's optimization process for analyzing test data. Nevertheless, machine-learning algorithms largely design their own predictive models based on existing data, finding patterns in the data that can be used to generate predictions that can be quite accurate.[131]

As the amount of data generated on a daily basis has increased dramatically in recent years, and the cost of computing power has decreased, machine-learning algorithms have grown increasingly feasible to use. Their use in performing a wide variety of tasks in the private sector, health professions, and, increasingly, government stems from a desire to reap several key benefits that they offer, including increased accuracy, more consistent outcomes, faster computational speeds, and greater productivity. These benefits might even be characterized as inherent to digital algorithms.

Accuracy. By definition, algorithms consist of logical steps and equations; mathematical equations dutifully carry out rules created for them and produce outputs that are within those bounds. As a result, their accuracy can be assessed via metrics that are expressed clearly and numerically.[132] Data analysts can compare multiple different types of machine-learning algorithms to see which ones yield the most accurate results when performing similar tasks. In this way, different algorithms compete with one another to establish which has the lowest error rates. A survey of nearly two thousand machine-learning algorithms used for breast cancer risk prediction revealed one algorithm as the most accurate of the group.[133]

Consistency. Consistency underlies the conception of any fair system of government,[134] and deploying a single algorithm can help achieve consistent results. As digital algorithms comprise a set of established steps to approach their

---

[130] Lehr & Ohm, *supra* note 1, at 676.

[131] Typically, machine-learning analysis does not support causal claims. But sometimes it can be incorporated into, and assist with, broader analysis of causal connections. For related discussion, see Sendhil Mullainathan & Jann Spiess, *Machine Learning: An Applied Econometric Approach*, 31 J. ECON. PERSPS. 87, 96 (2017).

[132] Aditya Mishra, *Metrics To Evaluate Your Machine Learning Algorithm*, TOWARDS DATA SCIENCE (Feb. 24, 2018), https://towardsdatascience.com/metrics-to-evaluate-your-machine-learning-algorithm-f10ba6e38234 [https://perma.cc/LB64-8J4L].

[133] Ricvan Dana Nindrea, Teguh Aryandono, Lutfan Lazuardi & Iwan Dwiprahasto, *Diagnostic Accuracy of Different Machine Learning Algorithms for Breast Cancer Risk Calculation: A Meta-Analysis*, 19 ASIAN PAC. J. CANCER PREVENTION 1747, 1747 (2018), https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6165638/pdf/APJCP-19-1747.pdf [https://perma.cc/Q3KU-Z7JZ] (finding that an algorithm known as Super Vector Machine was superior in its forecasting ability).

[134] Amanda Frost, *Overvaluing Uniformity*, 94 VA. L. REV. 1567, 1568–69 (2008).

objective in a systematic manner, they are almost by definition approaching the same task in the same way each time.[135] They also lend themselves to high replicability of outcomes when applied to the same data and following the same computational procedures.[136]

Speed. Computers can return speedy results, which are especially valuable when time is of the essence. They can be useful with real-time tracking and reporting, such as in the FDA's use of microbial sources tracking to assess foodborne outbreaks in real time.[137] Their speed has made them valuable to private investors making high-frequency trades in securities markets[138]—and it is that same speed that can make them valuable to regulators overseeing these markets.[139]

Productivity. Computers are not only fast, but they can also handle a large volume of tasks at once, helping to expand any organization's productivity. Algorithms have the capacity to handle as many variables as their processing power allows. A modern computer typically has sixteen gigabytes of RAM— allowing for datasets of millions, possibly billions, of data points—more than enough for many algorithmic tasks.[140] Given the daunting tasks that government agencies must complete with limited budgets and time, digital systems' productivity improvements make them greatly appealing. This is undoubtedly part of the reason why the Internal Revenue Service uses data mining algorithms to predict fraud and abuse,[141] the General Service Administration has automated "administrative 'cutting and pasting' tasks,"[142] and the U.S. Patent and Trademark Office is

---

[135] We are assuming here digital algorithms that do not have stochasticity—or randomness—deliberately programmed into them. *See generally* James C. Spall, *Stochastic Optimization*, in HANDBOOK OF COMPUTATIONAL STATISTICS: CONCEPTS AND METHODS at 173 (2nd ed., J. Gentle, W. Härdle, and Y. Mori, eds.) (2012).

[136] Yash Raj Shrestha, Shiko M. Ben-Menahem & Georg von Krogh, *Organizational Decision-Making Structures in the Age of Artificial Intelligence*, 6 CAL. MGMT. REV. 66, 68, 70 (2019).

[137] FOOD & DRUG ADMIN., FDA COMMISSIONER'S FELLOWSHIP PROGRAM (2011), https://www.fda.gov/media/83569/download [https://perma.cc/8K9A-K337].

[138] STAFFS OF THE CFTC & SEC, FINDINGS REGARDING THE MARKET EVENTS OF MAY 6, 2010 2–3 (2010), https://www.sec.gov/news/studies/2010/marketevents-report.pdf [https://perma.cc/MB6Z-R3CD].

[139] *See* Cary Coglianese, *Optimizing Regulation for an Optimizing Economy*, 4 J.L. & PUB. AFFS. 1, 1–2 (2018) [hereinafter Coglianese, *Optimizing*].

[140] Håkon Hapnes Strand, *How Do Machine Learning Algorithms Handle Such Large Amounts Of Data?*, FORBES (Apr. 10, 2018), https://www.forbes.com/sites/quora/2018/04/10/how-do-machine-learning-algorithms-handle-such-large-amounts-of-data/#1442c680730d [https://perma.cc/FVU7-VMDA].

[141] DAVID DEBARR & MAURY HARWOOD, IRS, RELATIONAL MINING FOR COMPLIANCE RISK 177–78 (2004), https://www.irs.gov/pub/irs-soi/04debarr.pdf [https://perma.cc/UA6X-QJRZ].

[142] Jory Heckman, *How GSA Turned an Automation Project into an Acquisition Time-Saver*, FED. NEWS NETWORK (Mar. 29, 2018), https://federalnewsnetwork.com/technology-main/2018/03/how-gsa-turned-an-automation-project-into-a-acquisition-time-saver [https://perma.cc/H9LW-L5N2].

developing electronic examination tools to substitute for time-consuming manual document reviews.[143] The Federal Communications Commission and other agencies are using algorithmic natural language processing tools to review rulemaking dockets filled with hundreds of thousands, even millions, of public comments.[144]

B. Digital Algorithms Versus Human Algorithms

Digital algorithms are able to perform a variety of tasks better than humans can.[145] Digital algorithms, for example, can recall stored content faster and more accurately than humans.[146] Unlike humans, who are vulnerable to memory limitations when faced with more than four variables, algorithms have practically unlimited capacity for data storage and the handling of heavy information-processing workloads.[147]

Moreover, a single digital system can replace many different human decision-makers, allowing for greater consistency over a series of repeated decisions. When different humans must make governmental decisions, discrepancies and inconsistencies can arise between their judgments. By contrast, algorithms that are fixed—ones that accept the same inputs and training data—will be much more likely to produce consistent outputs.[148]

---

[143] U.S. PAT. & TRADEMARK OFF., FY 2019 UNITED STATES PATENT AND TRADEMARK OFFICE PERFORMANCE AND ACCOUNTABILITY REPORT 20 (2020), https://www.uspto.gov/sites/default/files/documents/USPTOFY19PAR.pdf [https://perma.cc/2UMX-86ZG]; Lea Helmers, Franziska Horn, Franziska Biegler, Tim Oppermann & Klaus-Robert Müller, *Automating the Search for a Patent's Prior Art with a Full Text Similarity Search*, PLOS ONE 1, 1 (Mar. 4, 2019).

[144] ENGSTROM ET AL., *supra* note 9, at 59–60; David A. Bray, *An Update on the Volume of Open Internet Comments Submitted to the FCC*, FED. COMMC'NS COMM'N (Sep. 17, 2014), https://www.fcc.gov/news-events/blog/2014/09/17/update-volume-open-internet-comments-submitted-fcc [https://perma.cc/ZH58-UEQC].

[145] For an overview of the relative advantages of digital algorithms, see generally AJAY AGRAWAL, JOSHUA GANS & AVI GOLDFARB, PREDICTION MACHINES: THE SIMPLE ECONOMICS OF ARTIFICIAL INTELLIGENCE (2018).

[146] *E.g.*, Soham Banerjee, Pradeep Kumar Singh & Jaya Bajpai, *A Comparative Study on Decision-Making Capability Between Human and Artificial Intelligence*, *in* 652 NATURE INSPIRED COMPUTING 203, 209 (Bijaya Ketan Panigrahi, M.N. Hoda, Vinod Sharma & Shivendra Goel eds., 2018).

[147] *See* MAX TEGMARK, LIFE 3.0: BEING HUMAN IN THE AGE OF ARTIFICIAL INTELLIGENCE 105–06 (2017) (discussing the information processing advantages that digital algorithms hold over human judges).

[148] Admittedly, this consistency also leads to a concern about digital algorithms: if they are wrongly designed, they can put in place flaws or biases that will then apply across all cases, as opposed to just some, as with an inconsistently distributed system dependent on human decision-

This is not to deny that humans will remain better at some tasks than will digital algorithms. The human mind, for example, is well-suited to making reflexive, reactionary decisions in response to sensory inputs.[149] Thus, a human automobile driver may be able to respond reflexively in less time than an algorithm when swerving to avoid an accident. But a human analyst would not be able to thoroughly comb through thousands of pages of documents as quickly as an algorithm. Many governmental tasks are more similar to the latter example. For instance, the U.S. Bureau of Labor Statistics ("BLS") collects data on workplace injuries from hundreds of thousands of businesses to enable the U.S. Department of Labor to identify methods for preventing workplace injuries. In the past, human analysts have needed to read and assign to each incident report a series of codes for occupation, event, injury, injury location, and injury source.[150] But by relying on a machine-learning system, the BLS can now have at least 80 percent of these codes assigned digitally in a manner quicker and more accurate on average than a trained human coder.[151] Digital algorithms' comparative speed and efficiency in tasks like these give them the potential to eliminate many backlogs and unfair delays in governmental processes.[152]

To understand machine-learning algorithms' comparative advantages and disadvantages, various efforts have been made to compare these digital algorithms'

---

makers. Consistency, in other words, is of little virtue if it only leads to ineffectual or problematic results delivered consistently. Yet if there exist some humans who can make accurate and unbiased decisions in a given context, that itself provides reason to think that humans can design digital systems to yield results that are both high quality and consistent. The key is ensuring that the human decision-makers who design digital algorithmic systems are smart and make high quality decisions about the design and operation of digital algorithms. In much the same way, a system that uses a consistent approach may also be easier to modify and fix when errors or biases arise.

[149] But new research seems continually to draw into question such claims about the inherent superiority of humans at given tasks. Development of "neuromorphic" hardware that mimics the human brain is starting to run brain-like software. Sara Reardon, *Artificial Neurons Compute Faster Than the Human Brain*, NATURE (Jan. 26, 2018), https://www.nature.com/articles/d41586-018-01290-0 [https://perma.cc/5H9N-56FF].

[150] P'SHIP FOR PUB. SERV. & IBM CTR. FOR BUS. GOV'T, THE FUTURE HAS BEGUN: USING ARTIFICIAL INTELLIGENCE TO TRANSFORM GOVERNMENT 8 (2018), https://ourpublicservice.org/wp-content/uploads/2018/01/0c1b8914d59b94dc0a5115b739376c90-1515436519.pdf [https://perma.cc/6M24-EJHR] [hereinafter THE FUTURE HAS BEGUN].

[151] *Automated Coding of Injury and Illness Data*, U.S. BUREAU OF LAB. STAT. (Sept. 21, 2020), https://www.bls.gov/iif/autocoding.htm [https://perma.cc/CJX6-ZRRN]; *see also* P'SHIP FOR PUB. SERV. & IBM CTR. FOR BUS. GOV'T, *supra* note 150, at 8 (discussing BLS reliance on AI to assist with coding data).

[152] *See* ENGSTROM ET AL., *supra* note 9, at 854 ("Managed well, algorithmic governance tools can modernize public administration, promoting more efficient, accurate, and equitable forms of state action.").

performance directly to that of humans.[153] The most famous of these efforts have pitted digital algorithms against humans in games such as chess and Go.[154] Others focused on medical and business decisions. For example, with respect to clinical diagnoses of certain skin lesions, state-of-the-art machine-learning classifiers have been shown to be more accurate than board-certified dermatologists and other physicians.[155] In mortgage lending, automated underwriting algorithms apparently "more accurately predict[] default" than human underwriters do, resulting in "higher borrower approval rates, especially for underserved applicants."[156]

Other studies have compared machine-learning algorithms' performance with status quo results and found improved performance in a variety of distinctively public sector tasks:

- Greek border officials deployed a machine-learning system to screen travelers for COVID-19.[157] Researchers found that the digital algorithm identified about two- to four-times as many asymptomatic travelers during peak travel than traditional screening protocols.[158]

- Only about 10 percent of the more than three hundred thousand facilities subject to U.S. Environmental Protection Agency water pollution regulations receive government inspections in any given year, and normally only about 7 percent of inspected facilities are found noncompliant.[159] But when using a machine-learning algorithm, inspectors could undertake the same number of inspections but find more than six times the number of

---

[153] DANIEL KAHNEMAN, OLIVIER SIBONY & CASS R. SUNSTEIN, NOISE: A FLAW IN HUMAN JUDGMENT 336 (2021) ("A great deal of evidence suggests that algorithms can outperform human beings on whatever combination of criteria we select.").

[154] *See, e.g.*, David Silver et al., *Mastering the Game of Go with Deep Neural Networks and Tree Search*, 529 NATURE 484, 488 (2016) (reporting that the AlphaGo computer program beat a human champion in five straight games).

[155] Philipp Tschandl et al., *Comparison of the Accuracy of Human Readers Versus Machine-Learning Algorithms for Pigmented Skin Lesion Classification: An Open, Web-Based, International, Diagnostic Study*, 20 LANCET ONCOLOGY 938, 943 (2019). *But see* Taku Harada et al., *A Perspective from a Case Conference on Comparing the Diagnostic Process: Human Diagnostic Thinking vs. Artificial Intelligence (AI) Decision Support Tools*, INT'L J. ENV'T RSCH. & PUB. HEALTH (2020), https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7504543 [https://perma.cc/2ZDD-RJX4].

[156] Susan Wharton Gates, Vanessa Gail Perry & Peter M. Zorn, *Automated Underwriting in Mortgage Lending: Good News for the Underserved?*, 13 HOUS. POL'Y DEBATE 369, 370 (2002).

[157] Hamsa Bastani, Kimon Drakopolous, Vishl Gupta, Jon Vlachogiannis, Christos Hadjicristodoulou, Pagona Lagiou, Gkikas Magiorkinis, Dimitrios Paraskevis & Sotrios Tsiodras, *Efficient and Targeted COVID-19 Border Testing Via Reinforcement Learning*, 599 NATURE 108, 108 (2021).

[158] *Id.*

[159] Miyuki Hino, Elinor Benami & Nina Brooks, *Machine Learning for Environmental Monitoring*, 1 NATURE SUSTAINABILITY 583, 583–84 (2018).

regulatory violators—increasing the rate of violation detection to about 50 percent of all inspections.[160]

- Human judges worry that defendants who are released from jail will commit crimes while out on bail. A machine-learning algorithm could grant or deny bail at the same rate as judges but reduce crime by 25 percent—or they could keep crime rates the same and reduce jailing by 42 percent.[161] These improvements can be obtained even while reducing racial disparities in jailing rates.[162]

- Replacing humans with machine learning for arraignment decisions in domestic violence cases could cut in half the number of rearrests for domestic violence within two years of release.[163]

These examples indicate the considerable potential machine learning holds for improving governmental performance.

Demonstrating that machine learning can outperform humans in the completion of some tasks does not mean that they will outperform humans in every task. Machine learning tends to perform best with tasks involving pattern recognition and high levels of repetition. This means that even if humans remain distinctively advantaged for tasks requiring creativity and solving unique problems, digital algorithms still hold great promise for reducing much of the drudgery work in government.[164]

Nevertheless, many commentators still oppose the use of machine-learning algorithms. These critics charge that machine-learning algorithms are too opaque and prone to bias.[165] Yet even with respect to the qualities of transparency and lack of bias, humans do not necessarily compare favorably to machine learning.

---

[160] *Id.*

[161] Jon Kleinberg, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig & Sendhil Mullainathan, *Human Decisions and Machine Predictions*, 133 Q.J. ECON. 237, 241 (2017).

[162] *Id.*

[163] Richard A. Berk, Susan B. Sorenson & Geoffrey Barnes, *Forecasting Domestic Violence: A Machine Learning Approach To Help Inform Arraignment Decisions*, 13 J. EMPIRICAL L. STUD. 94, 105 (2016).

[164] *E.g.*, P'SHIP FOR PUB. SERV. & IBM CTR. FOR BUS. GOV'T, MORE THAN MEETS AI: ASSESSING THE IMPACT OF ARTIFICIAL INTELLIGENCE ON THE WORK OF GOVERNMENT 3 (Feb. 27, 2019), https://ourpublicservice.org/wp-content/uploads/2019/02/More-Than-Meets-AI.pdf [https://perma.cc/3PW3-8EVZ]; Emma Martinho-Truswell, *How AI Could Help the Public Sector*, HARV. BUS. REV. (Jan. 26, 2018), https://hbr.org/2018/01/how-ai-could-help-the-public-sector [https://perma.cc/XU7N-PJS7].

[165] *See supra* notes 10–12 and accompanying text.

It is true that so-called black-box machine-learning algorithms do not offer an intuitive basis for understanding why they reach their outcomes. But data scientists are extensively researching algorithmic explainability and finding techniques to understand and explain the results of machine-learning algorithms.[166] Moreover, humans are themselves far from transparent.[167] Expert judgments are often "cryptic and mysterious" to those affected by their judgments.[168] Even when humans explain their decisions, these accounts can be as much rationalizations as true reasons—a point legal realists made nearly a century ago with respect to judicial decision-making.[169] People themselves often do not really know the reasons why they decided as they did. In many contexts, the resulting decisions can come about from "implicit biases about which we are often unaware ourselves."[170] Indeed, for this reason, "[i]n many ways, human cognition forms the ultimate black box, even to the person engag[ed] in the cognitive activity."[171]

When it comes to bias, the issue again is not whether machine-learning algorithms can escape bias altogether, but rather whether they can perform better than humans. Again, well-designed and responsibly administered digital algorithms can sometimes do better than humans—even when trained on datasets with baked-in human biases.

---

[166] For a discussion of some of these developments, see Coglianese & Lehr, *Transparency*, *supra* note 1, at 50–55.

[167] *See* Sendhil Mullainathan, *Biased Algorithms Are Easier To Fix than Biased People*, N.Y. TIMES (Dec. 6, 2019), https://www.nytimes.com/2019/12/06/business/algorithm-bias-fix.html [https://perma.cc/F2L8-Z69D] ("Humans are inscrutable in a way that algorithms are not. Our explanations for our behavior are shifting and constructed after the fact."); John Zerilli, Alistair Knott, James Maclaurin & Colin Gavaghan, *Transparency in Algorithmic and Human Decision-Making: Is There a Double Standard?*, 32 PHIL. & TECH. 661, 663 (2019) ("[M]uch human decision-making is fraught with transparency problems . . . ."). Michael Lewis has tellingly compared the use, in response to pandemics, of computer-based disease models to human judgment by experts, observing that the latter have implicitly "used models" too. LEWIS, *supra* note 21, at 85. He has aptly noted that the experts relied on models or

> abstractions to inform their judgments. Those abstractions just happened to be inside their heads. Experts took the models in their minds as the essence of reality, but the biggest difference between their models and the ones inside the computer was that their models were less explicit and harder to check. Experts made all sorts of assumptions about the world, just as computer models did, but those assumptions were invisible.

Id.

[168] Jay Hegdé & Evgeniy Bart, *Making Expert Decisions Easier To Fathom: On the Explainability of Visual Object Recognition Expertise*, 12 FRONTIERS NEUROSCIENCE 1, 1 (Oct. 12, 2018).

[169] *See* WILLIAM TWINING, KARL LLEWELLYN AND THE REALIST MOVEMENT 229–31 (1973).

[170] Jon Kleinberg, Jens Ludwig, Sendhil Mullainathan & Cass R. Sunstein, *Algorithms as Discrimination Detectors*, 117 PROC. NAT'L. ACAD. SCIS. U.S.A. 30096, 30097 (2020).

[171] *Id.*; *see also id.* at 30100 ("It is tempting to think that human decision making is transparent and that algorithms are opaque . . . [, but] the opposite is true—or could be true.").

One reason why digital algorithms can fare better in avoiding bias is that algorithms necessarily demand the centralized compilation of large volumes of data. As a result, the use of digital algorithms necessarily brings with it the information needed to detect unwanted biases.[172] By comparison, governmental processes that depend on a distributed series of one-off decisions by different humans may never even produce the kind of aggregate data that would make unwanted disparate treatment visible. It is typically only with big data of the kind that fuels machine-learning algorithms that researchers can even ferret out the discrimination that humans perpetrate.

Another reason digital algorithms fare better than human algorithms when it comes to bias is that, once bias is detected (whether in humans or machines), the digital algorithms can be easier to debias. Debiasing humans, after all, can be quite difficult.[173] By contrast, with digital algorithms it will always be possible in principle to make mathematical adjustments that reduce unwanted biases. These adjustments can even be made while avoiding unlawful forms of "reverse discrimination."

Overall, digital algorithms can outperform human algorithms, exhibiting positive qualities such as accuracy, consistency, speed, and productivity. And even with respect to negative concerns, such as opacity and bias, digital algorithms may again fare much better than humans, even if they are not altogether perfect or error-free.

C. Human Errors with Digital Algorithms

Digital algorithms' biggest weakness may well stem from the fact that they need to be designed and operated by humans. All of the human foibles discussed in Part I can come into play with the development and deployment of digital algorithms. Humans may rush to put in place digital systems that are insufficiently thought-through and vetted. Humans may also be inattentive to the full range of values affected and consequences created by digital systems. If humans remain inattentive to or unconcerned about the possibility of bias, then digital algorithms' advantages with respect to de-biasing will never materialize.

---

[172] *Id.* at 30098 ("[Digital] algorithms . . . have the potential to become a force for social justice by serving as powerful detectors of human discrimination.").

[173] *See, e.g.*, Mullainathan, *supra* note 167 ("Changing people's hearts and minds is no simple matter."); Edward H. Chang, Katherine L. Milkman, Dena M. Gromet, Robert W. Rebele, Cade Massey, Angela L. Duckworth & Adam M. Grant, *The Mixed Effects of Online Diversity Training*, 116 PROC. NAT'L. ACAD. SCIS. U.S.A. 7778, 7781 (2019) (finding modest effects at best from diversity training, but with no effects on the individuals that "policymakers typically hope to influence most with such interventions"). The difficulty in eliminating bias from human should be evident from, if nothing else, the persistence of racist and misogynistic beliefs and outcomes in society.

As Part I suggests, human decision-making has been responsible for an untold number of mistakes, injustices, and calamities. This unfortunately includes, at times, failures in deploying digital systems, such as:

- Stanford University's initial digital algorithm for allocating COVID-19 vaccines excluded nearly all of its medical residents from the initial priority group, even though many of them regularly treat COVID-19 patients.[174] Although it was a digital algorithm that established the preliminary vaccine allocation decisions, the human administrators who reviewed and approved the ultimate plan were untested in novel situations and showed an "utter disconnect [from] . . . front line workers."[175]

- Many states are using data mining algorithms to identify fraud in food stamp benefits, unemployment insurance, and Medicaid.[176] In Michigan, a digital fraud detection system adopted in 2013 made roughly 48,000 fraud accusations against unemployment insurance recipients and forced repayment and high penalties through garnished wages, levied bank accounts, and seized tax refunds.[177] Later, a state review determined that 93 percent of these fraud determinations were incorrect.[178]

---

[174] Laurel Wamsley, *Stanford Apologizes After Vaccine Allocation Leaves Out Nearly All Medical Residents*, NPR (Dec. 18, 2020, 8:04 PM), https://www.npr.org/sections/coronavirus-live-updates/2020/12/18/948176807/stanford-apologizes-after-vaccine-allocation-leaves-out-nearly-all-medical-resid [https://perma.cc/7UV4-2AWP].

[175] *Id.*

[176] Michele Gilman, *AI Algorithms Intended To Root Out Welfare Fraud Often End Up Punishing the Poor Instead*, CONVERSATION (Feb. 14, 2020, 8:45 AM), https://theconversation.com/ai-algorithms-intended-to-root-out-welfare-fraud-often-end-up-punishing-the-poor-instead-131625 [https://perma.cc/LRS5-KCY4].

[177] Allie Gross, *Update: UIA Lawsuit Shows How the State Criminalizes the Unemployed*, DET. METRO TIMES, https://www.metrotimes.com/news-hits/archives/2015/10/05/uia-lawsuit-shows-how-the-state-criminalizes-the-unemployed [https://perma.cc/T77R-77TR] (last updated Oct. 5, 2015, 12:06 PM); Jonathan Oosting, *Michigan Refunds $21M in False Jobless Fraud Claims*, DET. NEWS (Aug. 11, 2017, 2:00 PM), https://www.detroitnews.com/story/news/politics/2017/08/11/michigan-unemployment-fraud/104501978/ [https://perma.cc/T3B5-7BZK].

[178] Sarah Cwiek, *State Review: 93% of State Unemployment Fraud Findings Were Wrong*, MICH. RADIO (Dec. 16, 2016, 6:03 PM), https://www.michiganradio.org/politics-government/2016-12-16/state-review-93-of-state-unemployment-fraud-findings-were-wrong [https://perma.cc/VT38-9U6Z]. Controversy also emerged in recent years over an automated fraud detection system in Australia. Luke Henriques-Gomes, *Robodebt Class Action: Coalition Agrees To Pay $1.2bn To Settle Lawsuit*, GUARDIAN (Nov. 16, 2020, 4:42 AM), https://www.theguardian.com/australia-news/2020/nov/16/robodebt-class-action-coalition-agrees-to-pay-12bn-to-settle-lawsuit [https://perma.cc/33V5-JJZK]. And in the Netherlands in 2020, a court ruled that a digital system

- Between 2015 and 2020, at least twenty federal agencies as varied as the U.S. Immigration and Customs Enforcement and the U.S. Postal Inspection Service, used or owned facial recognition software.[179] According to the Department of Commerce's National Institute of Standards and Technology, facial recognition software shows widespread evidence of racial bias, with some algorithms generating results that are up to one hundred times more likely to confuse two different individuals of color than two different white individuals.[180]

- When COVID-19 kept students in England from sitting for their university admissions exams, the government's Office of Qualifications and Examinations Regulation ("Ofqual") opted to create an algorithm to impute scores to students "based on evidence of their likely performance in the exams had they gone ahead."[181] The algorithm was intended to adjust for grade inflation, but it actually lowered the scores for 40 percent of students compared with their teacher-awarded grades.[182] Following heated public

---

used to detect fraud in social benefits claims violated the European Convention on Human Rights. Rb. Den Haag 2 mei 2020, ECLI:NL:RBDHA:2020:865 (NJCM/Netherlands) (Neth.), ¶ 6.7.

[179] Rachel Metz, *Facial Recognition Tech Has Been Widely Used Across the US Government for Years, a New Report Shows*, CNN BUS., https://www.cnn.com/2021/06/30/tech/government-facial-recognition-use-gao-report/index.html [https://perma.cc/DFQ8-SQ5M] (last updated June 30, 2021, 1:15 PM).

[180] *NIST Study Evaluates Effects of Race, Age, Sex on Face Recognition Software*, NAT'L INST. FOR STANDARDS & TECH. (Dec. 19, 2019), https://www.nist.gov/news-events/news/2019/12/nist-study-evaluates-effects-race-age-sex-face-recognition-software [https://perma.cc/LR2U-WBL2]; *see also* Brian Fung, *Facial Recognition Systems Show Rampant Racial Bias, Government Study Finds*, CNN BUS., https://www.cnn.com/2019/12/19/tech/facial-recognition-study-racial-bias/index.html [https://perma.cc/6L8R-9Z9C] (last updated Dec. 19, 2019, 6:37 PM). In noting these important concerns about bias with facial recognition algorithms, we do not overlook the limitations and biases involved in relying on human recognition and recall. *See generally* SEAN M. LANE & KATE A. HOUSTON, UNDERSTANDING EYEWITNESS MEMORY: THEORY AND APPLICATIONS (2021).

[181] OFQUAL, AWARDING GCSE, AS, A LEVEL, ADVANCED EXTENSION AWARDS AND EXTENDED PROJECT QUALIFICATIONS IN SUMMER 2020: INTERIM REPORT 11–12 (2020) (U.K.), https://www.gov.uk/government/publications/awarding-gcse-as-a-levels-in-summer-2020-interim-report [https://perma.cc/T9LN-VM5M].

[182] *See* Richard Adams, Sally Weale & Caelainn Barr, *A-Level Results: Almost 40% of Teacher Assessments in England Downgraded*, GUARDIAN (Aug. 13, 2020, 6:39 AM), https://www.theguardian.com/education/2020/aug/13/almost-40-of-english-students-have-a-level-results-downgraded [https://perma.cc/D6Q6-5FBZ].

> uproar, Ofqual withdrew the algorithm-determined scores and let teachers' grade estimates prevail.[183]

Not all of these examples involved what might be termed true machine-learning algorithms, but they nevertheless serve as a reminder that failures can arise from digital algorithms—and as a reminder of the need for humans to learn from these failures. In some of these failed cases, government officials have neglected to engage in sufficient public vetting of their algorithmic tools.[184] Ofqual's efforts, for example, have been described as "proprietary, secretive and opaque," with overlooked "[o]pportunities for meaningful public accountability."[185]

In some instances of digital failure, it is possible that alternative systems based entirely on humans would have failed too. Still, it remains the case that no digital algorithm will itself be infallible. These algorithms will make their own mistakes—perhaps even ones that humans would not make. What they do promise, though, is to make fewer mistakes overall. That said, they can only achieve this promise if they are used with care. Just as humans can fail when making a decision that calls for purely human judgment, due to the limitations noted in Part I, they can also fail when making human judgments about the design and use of digital algorithms. The key is for humans to engage in smart decision-making about when and how to deploy digital algorithms.

---

[183] *See* Adam Satariano, *British Grading Debacle Shows Pitfalls of Automating Government*, N.Y. TIMES (Aug. 20, 2020), https://www.nytimes.com/2020/08/20/world/europe/uk-england-grading-algorithm.html [https://perma.cc/K2X8-XLUV].

[184] In 2017, the city of Boston sought to reconfigure its school bus schedules using a digital algorithm aimed at improving the "sleep health of high school kids, getting elementary school kids home before dark, supporting kids with special needs, lowering costs, and increasing equity overall." Joi Ito, *What the Boston School Bus Schedule Can Teach Us About AI*, WIRED (Nov. 5, 2018, 8:00 AM), https://www.wired.com/story/joi-ito-ai-and-bus-routes [https://perma.cc/H83T-FYDH]. But its initial plan was met with resistance by many angry parents who preferred the status quo—suggesting that better communication and engagement may have helped. *E.g.*, *id.*; Ellen P. Goodman, *Smart Algorithmic Change Requires a Collaborative Political Process*, REG. REV. (Feb. 12, 2019), https://www.theregreview.org/2019/02/12/goodman-smart-algorithmic-change-requires-collaborative-political-process [https://perma.cc/V36K-QY8M]. Although the city dropped its most ambitious plan to change bus schedules, it nevertheless used digital algorithms to optimize school bus routes, which reduced vehicle emissions and fuel costs considerably. Sean Fleming, *This US City Put an Algorithm in Charge of Its School Bus Routes and Saved $5 Million*, WORLD ECON. F. (Aug. 22, 2019), https://www.weforum.org/agenda/2019/08/this-us-city-put-an-algorithm-in-charge-of-its-school-bus-routes-and-saved-5-million [https://perma.cc/PL6W-L98E].

[185] Louise Amoore, *Why 'Ditch the Algorithm' Is the Future of Political Protest*, GUARDIAN (Aug. 19, 2020, 6:47 AM), https://www.theguardian.com/commentisfree/2020/aug/19/ditch-the-algorithm-generation-students-a-levels-politics [https://perma.cc/LNX5-2GZR].

III. DECIDING TO DEPLOY DIGITAL ALGORITHMS

When contemplating a shift from human to digital decision-making, the choice will be between one type of algorithm (human) and another type (digital). Choosing between a human or a digital algorithm always will itself require a process of some kind—or what we might call a meta-process, to distinguish it from the processes under consideration to perform a specific governmental task. That meta-process will unavoidably be one that humans must undertake.

This final Part thus focuses on how humans—namely, government officials—should approach choosing between a human and a digital algorithm. Careful decision-making will be needed to avoid humans making mistakes about the design and deployment of digital algorithms. By no means should government decision-makers rush unthinkingly into adopting and relying on machine-learning algorithms—no more than they should unthinkingly rush to shift from one type of human-driven process to another human-driven process.[186] The core question will always be whether a shift to using a digital algorithm would be better than the status quo that relies on human algorithms.

A. Selecting a Multicriteria Decision Framework

What constitutes "better" will not always be easy, straightforward, or uncontroversial. Moreover, a judgment that machine learning will (or will not be) better than human decision-making can never be meaningfully made in the abstract or across-the-board. The advisability of using machine learning will vary across different contexts and different tasks and problems. In some cases, machine learning will prove better than human decision-making; in other cases, it will not.[187]

---

[186] The overall need for care in choosing to digitize a governmental process is basically the same as is needed when making any decision to redesign a process. *See* Cary Coglianese, *Process Choice*, 5 REGUL. & GOVERNANCE 250, 255–57 (2011) (noting that, just as substantive choices about regulations need analysis, so too do choices about process). *See generally* CARY COGLIANESE, ORG. FOR ECON. COOP. & DEV., MEASURING REGULATORY PERFORMANCE: EVALUATING THE IMPACT OF REGULATION AND REGULATORY POLICY (2012), https://www.oecd.org/gov/regulatory-policy/1_coglianese%20web.pdf [https://perma.cc/7VC7-4B9E] [hereinafter COGLIANESE, MEASURING REGULATORY PERFORMANCE] (showing how regulatory procedures and processes can be evaluated empirically).

[187] In still other cases, systems which involve humans working in collaboration with digital systems may well prove the most optimal. For presentation purposes, this article has been framed around a binary choice between human algorithms and digital algorithms; however, the best option in some cases might involve a combination of the two. *Cf.* Tim Wu, *Will Artificial Intelligence Eat the Law? The Rise of Hybrid Social-Ordering Systems*, 119 COLUM. L. REV. 2001, 2026–28 (2019). The decision framework and factors presented throughout Part III could in principle be applied just as well to any option involving a hybrid system of human–machine collaboration.

Even when machine learning is better, this will not necessarily mean it will be better in every relevant respect. Machine learning is not perfect. These algorithms still make mistakes and present downsides. They can make demonstrable improvements in speed and accuracy, but perhaps at some loss in the intuitive explainability of decisions. Nevertheless, a full consideration of machine-learning algorithms' relative strengths and weaknesses may still lead to the judgment that, all things considered, machine learning is overall better than human decision-making for a given task and in a given context.[188]

Deciding whether to rely on machine learning will necessitate balancing different, and perhaps often competing, values. This kind of balancing could take one of at least three forms: due process balancing, benefit-cost analysis, or multicriteria policy analysis. The last of these is likely to be the best approach for administrators to use when facing the meta-question of whether and when to use machine-learning tools to automate tasks previously handled by humans.

Due Process Balancing. The first kind of balancing approach is reflected in the prevailing law of procedural due process, as articulated by the Supreme Court in its decision in *Mathews v. Eldridge*.[189] The *Mathews* test seeks to balance the government's interests affected by a particular procedure (such as the costs of administering the procedure) with the degree of improved accuracy the procedure would deliver and the private interests at stake.[190] Although the *Mathews* formula is often used by courts to assess a single process under challenge, it could be adapted by administrators as a framework for choosing between a status quo human-based process and a proposed shift to a digitally algorithmic process. The question would be which system delivers the most value on net, taking into account decisional accuracy and private stakes and then deducting the government's costs.

Well-designed machine-learning systems would seem almost inherently superior to human systems under a *Mathews* calculus: they are likely to be less costly than systems that must rely on hundreds, if not thousands, of human decision-makers, and their main appeal is that they can be more accurate than humans. The private interests at stake are essentially exogenous and will be presumably unaffected by the choice of whether to use a human or digital algorithm. As a result, reliance on the Mathews calculus would often collapse the choice between human systems and digital ones into a single question: Which will produce more accurate decisions? The *Mathews* calculus thus almost seems hardwired to support the digital algorithm, provided that the specific machine-learning application in question can be shown to produce more accurate decisions than human decision-

[188] That is, digital algorithms "can be far less imperfect than noisy and often-biased human judgment." KAHNEMAN, SIBONY & SUNSTEIN, *supra* note 153, at 337.

[189] Mathews v. Eldridge, 424 U.S. 319 (1976).

[190] *See id.* at 333–35.

makers.[191] Yet even though improvements in accuracy can be vital, the decision to shift to a machine-learning algorithm will surely entail other considerations beyond accuracy.

Benefit-Cost Analysis. A second balancing approach would sweep more broadly and account for both accuracy and all other consequences that a shift to machine learning might entail. It would call for administrators to make an all-things-considered judgment about the use of machine learning: essentially, to conduct a benefit-cost analysis. Machine learning would be justified under this approach when it can deliver net benefits (i.e., benefits minus costs) that are greater than those under the status quo. One advantage of this approach is that it accounts for more factors than the *Mathews* calculus. The *Mathews* factors are clearly important, but sometimes they will be incomplete. By contrast, benefit-cost analysis is, in principle, always complete, because it calls for a quantification and monetization of all consequences.[192] But benefit-cost analysis will also have its practical limits in this setting—at least if it is to be approached in a hard fashion that seeks to place every consequence into a common monetary equivalent that yields an estimate of net benefits.[193] It will likely be infeasible in most cases for administrators to conduct a hard benefit-cost analysis because some of the consequences of adopting machine learning will not be capable of being placed in a common unit. For example, if a particular machine learning application would be more accurate and efficient but would result in a greater and more disproportionate number of adverse errors for individuals in historically marginalized groups, it may be neither meaningful nor justifiable to put the efficiency gains and the equity losses in the same units.[194]

Multicriteria Decision Analysis. A third balancing approach—a variation on the first two—will more feasibly accommodate a range of values and consequences: multicriteria decision analysis.[195] This approach is also sometimes

---

[191] Coglianese & Lehr, *Regulating by Robot*, *supra* note 1, at 1185–89.

[192] For comprehensive treatments of benefit-cost analysis methods, see generally EDWARD M. GRAMLICH, A GUIDE TO BENEFIT-COST ANALYSIS (2d ed. 1997) and ANTHONY E. BOARDMAN, DAVID H. GREENBERG, AIDAN R. VINING & DAVID L. WEIMER, COST-BENEFIT ANALYSIS: CONCEPTS AND PRACTICE (5th ed. 2018).

[193] Even with respect to other issues, agencies do not always have enough information to monetize all benefits and costs. *See, e.g.*, Michigan v. EPA, 576 U.S. 743, 759 (2015) (stating that an agency is not required to "conduct a formal cost-benefit analysis in which each advantage and disadvantage is assigned a monetary value"); Amy Sinden, *Formality and Informality in Cost-Benefit Analysis*, 2015 UTAH L. REV. 93, 101.

[194] *See generally* ARTHUR M. OKUN, EQUALITY AND EFFICIENCY: THE BIG TRADEOFF (1975) (addressing the tension between equality and efficiency).

[195] Sometimes this is referred to as multigoal analysis. DAVID L. WEIMER & AIDAN R. VINING, POLICY ANALYSIS: CONCEPTS AND PRACTICE 355 (6th ed. 2017). For a brief introduction to methods of analyzing outcomes using criteria that cannot be converted into a common metric, see

called a qualitative or soft benefit-cost analysis.[196] Essentially, it calls for the decision-maker to run through a checklist of criteria against which both the human-based status quo and the digital alternative should be judged. These criteria will be more extensive than the three *Mathews* factors, but they need not be placed in the same precise common units as in a hard benefit-cost analysis. The decision-maker then compares how well each alternative will fare against each criterion, without necessarily converting any estimates into a common unit.

When choosing between digital and human-based options, it is important to gather and present as much information as possible about each alternative. Each can then be quantitatively (even if not monetarily) rated on each criterion (for example, number of errors). Where quantification is not possible, alternatives can at least be qualitatively rated with respect to each criterion. Even a rough qualitative metric, such as a three-point scale ("+" for positive, "+/-" for neutral, and "-" for negative), might be used to illustrate the strengths and weaknesses of each alternative when assessed against each criterion, with the ratings then placed in a summary table. A decision-maker can then better visualize the relative advantages and disadvantages of each alternative and proceed to make a reasoned judgment.[197]

This multicriteria analytic approach is likely to be the most practical and best approach for administrators to follow in deciding when to proceed with making a shift from a human-based status quo to a digital-based alternative.[198] The main question will be what criteria such an approach should include.

---

id. at 352–58. A branch within the field of operations research provides a suite of sophisticated mathematical tools that can be used in conducting multicriteria decision analysis. For perspectives on this analytic approach, see generally RALPH L. KEENEY & HOWARD RAIFFA, DECISIONS WITH MULTIPLE OBJECTIVES: PREFERENCES AND VALUE TRADEOFFS (1993) and MURAT KOKSALAN, JYRKI WALLENIUS & STANLEY ZIONTS, MULTIPLE CRITERIA DECISION MAKING: FROM EARLY HISTORY TO THE 21ST CENTURY (2011).

[196] *See, e.g.*, *id.* at 352–53 (discussing qualitative benefit-cost analysis); Sinden, *supra* note 193, at 107–29 (discussing differences between hard and soft, or formal and informal, benefit-cost analysis).

[197] In drawing upon such a qualitative scalar rating, it is important for decision-makers to use caution. Rather than relying simply on a summing up of the ratings, a decision-maker needs to consider the evidence fully and engage in sustained reasoning about each option. Not every criterion will deserve to be treated equally, as would occur with a summation of ratings. Furthermore, the uniform distance between different points on a scale likely will not reflect fully the true relevant differences between the strengths and weaknesses of different options.

[198] With respect to choosing whether to use machine learning, a multicriteria framework can be used at different stages of the development process when different information is available. That is, it can be used at the outset in deciding whether an agency should even invest in the development of a machine-learning based system, as well as later, whenever such system has been developed, in deciding whether to deploy the system. It can provide a basis for subsequent evaluation of the system in operation and making decisions about future modification of the system.

B. Key Criteria in Choosing Digital Algorithms

The actual criteria will vary to some degree from use to use, depending on the tasks that a machine-learning system would take over from humans. The precise criteria for a system used to read the handwriting on U.S. postal mail, for example, will differ from those that might be appropriate for deciding whether to use a machine-learning system to automate decisions about whether to grant license applications for commercial airline pilots.[199] Nevertheless, in general, two key categories of criteria should affect agencies' choices about whether to shift to a process based on machine learning: (1) preconditions for successful use and (2) improved outcomes.[200]

Preconditions for Use. Agencies will first need access to adequate human expertise as well as data storage and processing technologies. Analysts' and data scientists' expertise and time are needed to tailor and train algorithms to each specific task. This process of customizing each algorithm to each task can be labor-intensive. It also is technologically sophisticated. Unfortunately, government agencies must compete with the private sector to attract the necessary talent.[201] Without sufficient technical skills, agencies will be limited in their ability to realize the full potential of machine-learning algorithms.[202]

Digital algorithms are also dependent upon an analytic infrastructure—the hardware, software, and network resources needed to support the analysis of large

---

[199] The latter use is a hypothetical discussed at length in Coglianese & Lehr, *Transparency*, *supra* note 1, at 10, 17, 52–53.

[200] For discussion on which this section draws, see generally CARY COGLIANESE, A FRAMEWORK FOR GOVERNMENTAL USE OF MACHINE LEARNING 66–72 (2020), https://www.acus.gov/sites/default/files/documents/Coglianese%20ACUS%20Final%20Report.pdf [https://perma.cc/CW3H-WUFP] and Cary Coglianese & Alicia Lai, *Assessing Automated Administration*, *in* OXFORD HANDBOOK FOR AI GOVERNANCE (Justin Bullock et al. eds., forthcoming 2022). For a related discussion of issues for government agencies to consider when seeking to use AI tools successfully, see Desouza, *supra* note 125, at 11–18.

[201] Coglianese, *Optimizing*, *supra* note 139, at 10; *see also* Shelly Hagan, *More Robots Mean 120 Million Workers Need To Be Retrained*, BLOOMBERG (Sept. 6, 2019, 12:00 AM), https://www.bloomberg.com/news/articles/2019-09-06/robots-displacing-jobs-means-120-million-workers-need-retraining [https://perma.cc/ALN6-7XMC] (noting that AI advancements will require upskilling workers amid an existing talent shortage). Furthermore, the process of public sector hiring can be slow. Eric Katz, *The Federal Government Has Gotten Slower at Hiring New Employees for Five Consecutive Years*, GOV'T EXEC. (Mar. 1, 2018), https://www.govexec.com/management/2018/03/federal-government-has-gotten-slower-hiring-new-employees-five-consecutive-years/146348 [https://perma.cc/AAD6-RQ54].

[202] There are some positive indications. Under the Foundations for Evidence-Based Policymaking Act, signed into law in 2019, agencies must appoint "Chief Data Officers" and "Evaluation Officers" to understand and promote data, laying the stage for AI. Foundations for Evidence-Based Policymaking Act of 2018, Pub. L. No. 115-435, §§ 313, 3520(c), 132 Stat. 5529, 5531, 5541–42 (2019).

volumes of data. Agencies need storage systems that can house datasets and protect them from physical deterioration.[203] These storage systems and the networks used to analyze agency data must also be protected from hackers.[204] Some agencies have begun to realize the need to build this infrastructure.[205] However, many other agencies are still funneling resources into maintaining legacy systems that are largely becoming obsolete and remain too susceptible to cybersecurity risks.[206]

In addition to these tangible human and technology resources, which agencies will either need to have in place or secure through government contracts, there are more fundamental preconditions for government to rely on machine-learning tools. Currently these tools produce "narrow" AI, given their focus on specific, human-specified goals for well-defined problems. This is contrasted with "general" AI which, like humans, would exhibit creativity, flexibility, and learning beyond the confines of a well-defined task.[207] Where the preconditions for narrow AI are very poorly met, machine learning is unlikely even to be feasible for an agency to consider. The following three preconditions can be thought of as a necessary, even if not sufficient, condition for a potential shift from a human- to machine-based process:

---

[203] *Cf.* Ian Sample, *Google Boss Warns of 'Forgotten Century' with Email and Photos at Risk*, GUARDIAN (Feb. 13, 2015, 4:16 AM), https://www.theguardian.com/technology/2015/feb/13/google-boss-warns-forgotten-century-email-photos-vint-cerf [https://perma.cc/6GZN-YK45] (describing the risks posed by obsolescence of digital storage technologies).

[204] *See, e.g.*, OFF. OF THE INSPECTOR GEN., U.S. OFF. OF PERS. MGMT., SEMIANNUAL REPORT TO CONGRESS 8 (2019), https://www.opm.gov/news/reports-publications/semi-annual-reports/sar61.pdf [https://perma.cc/F5X2-CYWT] (describing "the implementation and maintenance of mature cybersecurity programs [as] a critical need for OPM and its contractors").

[205] The Federal Aviation Administration, Federal Deposit Insurance Corporation, and Federal Communications Commission have released statements of their efforts to create large data sets to support agency function. The Office of Financial Research within the U.S. Department of Treasury created the global Legal Entity Identifier program in an effort to make big data more readily analyzable for financial market regulators. The FDA, Environmental Protection Agency, and Securities Exchange Commission have begun to leverage cloud storage systems to store, consolidate, and analyze enormous data sets. For discussion of these agencies' efforts, see Coglianese & Lehr, *Regulating by Robot*, *supra* note 1, at 1162–66.

[206] Coglianese, *Optimizing*, *supra* note 139, at 11. *See also* Kevin C. Desouza, Delivering Artificial Intelligence in Government: Challenges and Opportunities 21–22 (2018), https://www.businessofgovernment.org/sites/default/files/Delivering%20Artificial%20Intelligence%20in%20Government.pdf (discussing the need for agencies "to replace, modify, and retire systems to accommodate modern systems that provide a platform to develop and deploy AI").

[207] For a helpful discussion of the distinction between narrow and general AI, see STUART RUSSELL, HUMAN COMPATIBLE: ARTIFICIAL INTELLIGENCE AND THE PROBLEM OF CONTROL 42–48 (2019).

- *Goal clarity and precision*. Machine-learning algorithms operate by optimizing with respect to a specified objective. Furthermore, an algorithm's objective function must, by definition, be mathematically defined. What this means is that machine-learning tools will only be appropriate for an operating task where the objective can be clearly defined.[208] For example, if the goal is simply to make the most accurate decisions about claimants' eligibility for benefits, the algorithm's goal can be specified in terms of reducing forecasting error.

  But if the goal is understood both to make accurate forecasts about who will be eligible while also minimizing unfairness to applicants from a racial minority group, then the degree of clarity may be insufficient for two reasons. First, it may be unclear what fairness exactly entails.[209] Must the benefits awarded be proportionate to the distribution of each racial group in society overall or in the applicant pool? Or, perhaps what must be proportionate is the degree of false negative errors? Second, even if fairness is defined with sufficient clarity, given how machine learning works, there will almost surely be a tradeoff between maximizing accuracy (the minimization of forecasting error) and addressing fairness. But in making such tradeoffs, agencies may have insufficient statutory direction or social consensus around how to define such a tradeoff in precise mathematical terms.[210] Exactly how much unfairness should be tolerated to avoid how much diminution in accuracy?

  In their need for goal clarity, machine-learning algorithms share many affinities with performance-based regulation—sometimes called regulation by objectives.[211] But, as has been noted elsewhere, it may not

[208] *See, e.g.,* Coglianese, *supra* note 18, at 47–49 (discussing the importance of "value completeness" and "value precision" in defining the objectives of an algorithmic tool).

[209] For helpful discussion of various options, see Mayson, *supra* note 8, at 2233–35.

[210] In human decision-making systems, the existence of such tradeoffs may be obscured and their resolution effectuated through what Cass Sunstein has called "incompletely theorized agreements." Cass R. Sunstein, *Incompletely Theorized Agreements*, 108 HARV. L. REV. 1733, 1735 (1995). But machine-learning algorithms demand more than such incomplete agreements, such as about what may be "reasonable." They need the value choices reflected in the algorithm's objective to be stated with mathematical precision.

[211] By presidential order, executive agencies are instructed that, when issuing regulations, they "shall, to the extent feasible, specify performance objectives, rather than specifying the behavior or manner of compliance that regulated entities must adopt." Exec. Order No. 12,866, § l(b)(8), 58 Fed. Reg. 51,735, 51,736 (Oct. 4, 1993).

always be clear what the full social objective is.[212] For example, for years federal regulators seeking to reduce accidental poisonings relied on a performance-based approach to standards for child-resistant packages containing drugs and household chemicals.[213] But these standards that optimized for child resistance also prevented adults from opening such containers easily—and thus induced many adults, once they managed to open these containers, to leave them open and thus left their contents easily accessible to children.[214] Only after seeing poisonings increase did regulators redefine their objectives and revise the standards to ensure that packaging would be resistant to opening by children but still easy for adults to use.[215] This example suggests that, at least in some cases, one of the most vexing preconditions for the use of machine learning will be to define a goal that is both acceptable on policy grounds and can be defined mathematically.

- *Data availability*. Machine learning achieves accurate forecasts by discerning patterns in large amounts of relevant data. If large amounts of data are unavailable, then a necessary ingredient will be missing and the use of machine learning to automate a task will simply not be viable. The necessary data may be unavailable for various administrative or technical reasons. For example, even though the data exist, they may only have been recorded and stored by an agency in paper, rather than digital, form.[216] Or, disparate digitally stored datasets may lack sufficient means to allow data for each business or individual in the different datasets to be linked to each other, such as through a common entity identifier.

  More fundamentally, sufficient data may be lacking because there simply is an insufficient number of narrow, repeated events around

[212] Cary Coglianese, *The Limits of Performance-Based Regulation*, 50 U. MICH. J.L. REFORM 525, 562 (2017).

[213] *Id.* at 532, 555.

[214] *See, e.g.,* W. Kip Viscusi, *The Lulling Effect: The Impact of Child-Resistant Packaging on Aspirin and Analgesic Ingestions*, 74 AEA PAPERS & PROCEEDINGS 324, 326 (1984) (describing how the standards ultimately resulted in "a sharp increase in the proportion of aspirin-related poisonings associated with protective packaging").

[215] Coglianese, *supra* note 212*,* at 555–56.

[216] Cary Coglianese, *Deploying Machine Learning for a Sustainable Future*, *in* A BETTER PLANET: 40 BIG IDEAS FOR A SUSTAINABLE FUTURE 200, 204 (Daniel C. Esty ed. 2019) (discussing the need for converting paper records to electronic format to provide data for machine-learning analysis); *cf.* Coglianese, *Optimizing*, *supra* note 139, at 11 (describing the prevalence of legacy IT systems in the federal government).

which data exist. It may be easier to find data to support machine-learning analysis of x-rays to determine if a coal miner qualifies for black lung benefits, but harder to find common data that could be used to determine whether asylum applicants satisfy the test of having a "well-founded fear of future persecution."[217] The latter requires both a "subjectively genuine and an objectively reasonable fear,"[218] which can encompass many unique circumstances.[219]

Similarly, data may be available to show the probability that a particular defendant's DNA could be contained within a mixed DNA sample from a crime scene.[220] But in the absence of any DNA samples, it may be impossible to have a large data set that can help determine a key fact in a criminal case, such as whether the defendant was driving a yellow convertible that passed through the intersection of Fourth and Chestnut Streets at 12:35 a.m. on November 17. In short, for questions that are truly one-of-a-kind, it will be inherently difficult to find a sufficiently large data set of the type needed to make machine learning a viable task.[221]

- *External validity*. Related to data availability is a question of the available data's representativeness of the population to which the algorithm will be applied. The world is ever-changing, so at a minimum, to make machine-learning systems viable, a government agency will need to have access to a steady stream of new data to keep updating an algorithm and retraining it as conditions in the world—and the data about those conditions—keeps changing. If the relevant parts of the world change more quickly than an algorithm's underlying datasets can be replenished with current data, then the algorithm will be

[217] 8 C.F.R. § 1208.13(b) (2021);); *see also* 8 U.S.C. 1101(a)(42) (specifying asylum qualification based on "a well-founded fear of persecution on account of race, religion, nationality, membership in a particular social group, or political opinion").

[218] De Belbruno v. Ashcroft, 362 F.3d 272, 284 (4th Cir. 2004).

[219] INS v. Cardoza-Fonseca, 480 U.S. 421, 448 (1987) ("[A] term like 'well founded fear'. . . can only be given concrete meaning through a process of case-by-case adjudication.").

[220] *See* Christopher Rigano, *Using Artificial Intelligence To Address Criminal Justice Needs*, NAT'L INST. OF JUST. (Oct. 8, 2018), https://nij.ojp.gov/topics/articles/using-artificial-intelligence-address-criminal-justice-needs [https://perma.cc/PD2L-WTHD].

[221] *Cf.* Gary Marcus & Ernest Davis, *A.I. Is Harder Than You Think*, N.Y. TIMES (May 18, 2018), https://www.nytimes.com/2018/05/18/opinion/artificial-intelligence-challenges.html [https://perma.cc/9AGR-UHVV] ("No matter how much data you have and how many patterns you discern, your data will never match the creativity of human beings or the fluidity of the real world."). For an earlier philosophical discussion, see HUBERT L. DREYFUS, WHAT COMPUTERS STILL CAN'T DO: A CRITIQUE OF ARTIFICIAL REASON (MIT Press rev. ed. 1992) (1972).

> "brittle"[222]—that is, it will suffer from what statisticians call an external validity problem. A machine-learning algorithm used to forecast employment levels in the economy, for example, might not be capable of producing an accurate forecast during an unprecedented, pandemic-induced recession.
>
> Of course, any kind of forecasting and decision-making tool—even human judgment—will be limited in unprecedented times or periods of rapid dynamism. Circumstances of true unknown unknowns—or what Professor Robin Hogarth calls "coconut uncertainty" [223] —present inherent levels of uncertainty. The key question is whether, under such circumstances, machine-learning algorithms will prove more or less brittle than other types of analysis, including human judgment. It is certainly conceivable that with the right kind of data acquisition and feedback process, an algorithmic system could be designed so that it fares better than human alternatives in periods of disruption. The high level of uncertainty endemic to such periods, though, will make it hard to be confident that machine learning—or anything else, for that matter—fares better than alternatives.

Taking these three preconditional factors together, machine-learning systems will realistically only amount to a plausible substitute for human judgment for tasks where the objective can be defined with precision, tasks that are repeated over a large number of instances (such that large quantities of data can be compiled), and tasks where data collection and algorithm training and retraining can keep pace with relevant changing patterns in the world. This is not to say that these preconditions must be perfectly satisfied nor that they are the only considerations to take into account. But if they are not even minimally satisfied for a given use case, it will make little sense to contemplate deploying digital algorithms. On the other hand, where these preconditions are sufficiently satisfied, there can be some reason for an administrator to think that machine learning could improve on the status quo and that it will be worth taking further steps to assess the possibility of deploying an algorithmic system.

---

[222] M.L. CUMMINGS, WOMEN CORP. DIRS., THE SURPRISING BRITTLENESS OF AI 2 (2020), https://www.womencorporatedirectors.org/WCD/News/JAN-Feb2020/Reality%20Light.pdf [https://perma.cc/LU5C-R5TH].

[223] ROBIN HOGARTH, ON COCONUTS IN FOGGY MINE-FIELDS: AN APPROACH TO STUDYING FUTURE-CHOICE DECISIONS 6 (2008), https://www.researchgate.net/publication/228499901_On_Coconuts_in_Foggy_Mine-Fields_An_approach_to_studying_future-choice_decisions [https://perma.cc/FDS7-WLC5].

Performance in Improving Outcomes. The next step, after determining if the necessary preconditions for a machine-learning option can be satisfied, is to assess a digital system's likely performance in improving outcomes. This is the ultimate test for machine learning: how it performs compared to the status quo.

As Part I makes clear, the human status quo leaves plenty of room for improvement. Whether a machine-learning system is realistically expected to fare better will constitute a centerpiece of any multicriteria analysis aimed at deciding whether to adopt machine learning. The precise definition of "better" will need to be informed by each specific task, whether that task involves forecasting the weather, identifying tax fraud, or determining eligibility for licenses or benefits. Although the specific relevant criteria will vary across different uses, it is possible to identify three general types of impacts that should be considered in determining whether machine learning improves outcomes:

- *Goal performance*. Current systems operated by humans have goals that they are meant to achieve. The first set of outcome-oriented criteria for deciding whether to use machine learning should be guided by those prevailing goals. The relevant factors can be captured by a series of straightforward questions: Would machine learning prove more accurate in achieving an administrative agency's goals? Would it operate more quickly? Would it cost less? Would it yield a greater degree of consistency? These questions can be asked from the standpoint of the current statutory purpose or operational goal of a human-driven system. Decision-makers can also step back and use the possibility of automation to consider current goals afresh. They will do well to consider more precisely the underlying problem that the system is supposed to solve and seek to measure the degree to which the digital algorithm helps solve it. The key will be to determine whether—and by how much—machine learning will help an administrative agency do its job better.[224] As indicated in Part II.B, in important instances digital algorithms can indeed achieve improvements in the attainment of basic administrative and policy goals. This does not mean, of course, that they will always result in improvements.

- *Impacts on those directly affected*. The ways that machine learning might help an agency do its job better are only one way to consider machine learning's impacts. Unless already fully captured in the agency's own performance goals, it is also important to assess the

[224] For a discussion of regulatory outcomes and their evaluation, see COGLIANESE, MEASURING REGULATORY PERFORMANCE, *supra* note 186, at 9–13.

effects of machine learning on those individuals or businesses who would be directly affected by a specific machine-learning system, such as the applicants for government benefits or licenses. How would a machine-learning system treat them? Would their data be kept private? Would some directly affected parties gain or suffer disproportionately to others? Would those directly affected by a machine-learning system feel like that system has served them fairly? Recall that algorithmic systems do not need to be perfect or completely problem free—just better than the status quo. If the status quo for some tasks is dependent on human personnel to answer telephones and thus keeps members of the public waiting on hold for hours before they speak to a person who can assist them, a machine-learning chatbot could be much better, relatively speaking. Indeed, the private firm eBay uses a fully automated customer dispute resolution system that works so well that customers who experience disputes are reportedly more inclined to do business with eBay again than are those who never experience a dispute inthe first place.[225]

- *Impacts on broader public*. Unless already factored into the agency's own performance goals, administrators contemplating the introduction of a digital algorithmic system should include broader societal effects in any multicriteria analysis. How would machine learning affect those who might not be directly interacting with or be affected by the system? Will the errors that remain with machine learning prove to have broader societal consequences? Few such spillover effects might exist, for example, with an automated mailing sorting system. But they would certainly be present with a digital system that determines who can receive a commercial pilot's license. There, the impact on air travelers surely would need to be considered. Ultimately, the most crucial question will again be a comparative one: Will the broader consequences of the machine-learning system prove to be more or less positive than the consequences prevailing under the status quo?

  It is conceivable that a machine-learning system could deliver improved outcomes across all types of outcomes. Yet probably few processes—digital or otherwise—will perform better than the status quo on each and every possible type of outcome. As a result, efforts must be made to characterize the degree of improvements and performance losses

[225] *See* BENJAMIN H. BARTON & STEPHANOS BIBAS, REBOOTING JUSTICE: MORE TECHNOLOGY, FEWER LAWYERS, AND THE FUTURE OF LAW 113 (2017); ETHAN KATSH & ORNA RABINOVICH-EINY, DIGITAL JUSTICE: TECHNOLOGY AND THE INTERNET OF DISPUTES 34–35 (2017).

> resulting from a shift to machine learning. Administrators, in other words, should ask not only *whether* machine learning improves accuracy, but by *how much* and *at what cost*.

Decision-makers will need to establish priorities among these different types of outcomes—goal performance as well as impacts on those directly affected and on the broader public. If using machine learning for a particular task turns out to lower the administrative costs of a performing that task but will also result in a slight loss of accuracy compared with the status quo, it will be necessary to ask how important accuracy is for the given task. Are any errors that occur with machine learning all that consequential? It may be fine, for example, for the U.S. Postal Service ("USPS") to accept some degree of loss in the accuracy of letter-sorting if doing so could dramatically lower the costs of handling the mail. But it will be much less acceptable to tolerate a similar tradeoff between administrative cost savings and predictive accuracy with a system designed to identify catastrophic safety risks in oil and gas pipelines.

Before choosing to rely on a digital system, decision-makers should ensure that they have carefully validated its performance—assessing statistically whether machine learning can be expected to lead to improved outcomes.[226] Such validation efforts should be undertaken when training and testing an algorithm on historic data, conducted before adopting any digital system wholesale. Agencies may also consider setting up pilot programs to run a digital system in parallel with the current human-driven process for a length of time to study how it will operate in practice.[227] Even though validation efforts are needed before deciding to deploy a digital system, these efforts should continue even after it replaces a human-driven system. Indeed, it would be prudent to evaluate the system relatively early in its use before any loss of human skill becomes entrenched. It will also be appropriate to audit performance on a regular basis at specified intervals. Future upgrades to any digital system would benefit from further auditing efforts to ensure that each new version improves on the one that preceded it—or at least does not create any new unacceptable side effects or other problems.

Assessing how well a new digital system will meet the preconditions for success and determining whether it will improve outcomes is simply being smart and responsible. Failing to think through decisions to digitize can have real and even tragic consequences for the public. Public officials must be aware of and

---

[226] *Cf.* Adoption of Recommendations, 82 Fed. Reg. 61,728, 61,738 (Dec. 29, 2017) (explaining the importance of agencies trying to "learn whether outcomes are improved in those time periods or jurisdictions with the regulatory obligation").

[227] Professors David Engstrom and Daniel Ho call this approach "prospective benchmarking." David Freeman Engstrom & Daniel E. Ho, *Algorithmic Accountability in the Administrative State*, 37 YALE J. ON REG. 800, 849–53 (2020).

deliberate about combatting their own physical and cognitive limitations and avoiding any potential pitfalls from collective decision-making over the use of artificial intelligence.[228]

Failure to take due care can also leave an agency susceptible to public controversy and litigation. While real, these risks of conflict and litigation are not truly distinctive.[229] The various objections to governmental use of machine learning—opacity, bias, and such—have their analogues in legal principles that agencies have had to comply with for decades.[230] As a result, nothing intrinsic about machine learning should lead government agencies to eschew consideration of digital algorithms due to legal risks.[231] Standard principles of administrative law can readily accommodate use of machine-learning tools as long as agencies pursue their use responsibly.[232]

In fact, agencies could even find that sometimes their legal positions and relationships with the public improve when they implement well-designed digital tools.[233] After all, if these tools can perform better than humans in delivering accurate, prompt, and fair outcomes, agencies may have a legal obligation to deploy them to enhance administrative justice.[234] The upshot is that agency officials who act responsibly in deciding to rely on machine learning should be able to manage litigation risks and avoid needless controversy—all while delivering real public value.

---

[228] Decision-makers would do well in this regard to consider the guidance offered by public administration scholars about the need for ensuring legitimacy and accountability in governmental uses of AI. *See generally* Madalina Busuioc, *Accountable Artificial Intelligence: Holding Algorithms to Account*, 81 PUB. ADMIN. REV. 825 (2020) (providing recommendations on how to address AI's accountability issues); Matthew M. Young, Justin B. Bullock, & J. Lecy, *Artificial Discretion as a Tool of Governance: A Framework for Understanding the Impact of Artificial Intelligence on Public Administration*, 2 PERSPS. ON PUB. MGMT. & GOVERNANCE 301 (2019) ("provid[ing] a framework for defining, characterizing, and evaluating artificial discretion as a technology that both augments and competes with traditional bureaucratic discretion").

[229] For a review of the litigation to date over governmental authorities' use of mathematical algorithms, see Coglianese & Ben Dor, *supra* note 8, at 35–37.

[230] *See* Coglianese & Lehr, *Transparency*, *supra* note 1, at 30 ("[N]efarious governmental action can take place entirely independently of any application of machine learning.").

[231] *Id.*; Coglianese & Lehr, *Regulating by Robot*, *supra* note 1, at 1202; Steven M. Appel & Cary Coglianese, *Algorithmic Administrative Justice*, *in* THE OXFORD HANDBOOK OF ADMINISTRATIVE JUSTICE (Marc Hertogh et al. eds., 2021). Some of this work forms a basis for the discussion contained in this Part.

[232] Coglianese & Lehr, *Regulating by Robot*, *supra* note 1, at 1215; Coglianese & Lehr, *Transparency*, *supra* note 1, at 42, 55.

[233] Cary Coglianese & Kathryn Hefter, *From Negative to Positive Algorithm Rights*, WM. & MARY BILL OF RTS. J. (forthcoming 2022).

[234] *See id.*; Appel & Coglianese, *supra* note 231, at 15.

C. Putting Digital Algorithms in Place

The key ultimately is for government officials to make sound decisions about putting digital algorithms in place. Three principal strategies are available to help agencies achieve this objective: planning, public participation, and procurement provisions.

First, planning entails going through the types of assessments outlined in Parts III.A and III.B. By conducting algorithmic audits and validation studies, and by completing a multicriteria analysis, agency officials can assure that they will be making better informed decisions about their agencies' use of digital systems.[235] In engaging in this planning, agencies can rely on an extensive array of guidelines.[236] This includes the Organization for Economic Cooperation and Development's "principles for AI,"[237] the Administrative Conference of the United States's statement on "issues agencies should consider when adopting or modifying AI systems,"[238] an executive order promoting governmental use of AI that "fosters

[235] Private sector firms increasingly recognize the importance of full, robust vetting of new forms of AI. Los Alamos National Laboratory, *How Artificial Intelligence and Machine Learning Transform the Human Condition*, YOUTUBE, at 31:26 (Aug. 2, 2021), https://www.youtube.com/watch?v=HyuqxdfC4oE [https://perma.cc/K53U-5Q8R] (address by Andrew Moore, Director of Google Cloud AI). For guidance on auditing digital algorithms, see Joshua A. Kroll, Joanna Huey, Solon Barocas, Edward W. Felten, Joel. R. Reidenberg, David G. Robinson & Harlan Yu, *Accountable Algorithms*, 165 U. PA. L. REV. 633, 660–61 (2017); MILES BRUNDAGE ET AL., TOWARD TRUSTWORTHY AI DEVELOPMENT: MECHANISMS FOR SUPPORTING VERIFIABLE CLAIMS 24–25 (2020), https://arxiv.org/pdf/2004.07213.pdf [https://perma.cc/T86W-LEGX]; SUPREME AUDIT INSTITUTIONS OF FINLAND, GERMANY, THE NETHERLANDS, NORWAY, AND THE UK, AUDITING MACHINE LEARNING ALGORITHMS: A WHITE PAPER FOR PUBLIC AUDITORS 15–17 (2020), https://www.auditingalgorithms.net/auditing-ml.pdf [https://perma.cc/8WHB-VWZ2].

[236] For a general overview of regulatory principles, proposals, and other initiatives related to AI in the United States, see Christopher S. Yoo & Alicia Lai, *Regulation of Algorithmic Tools in the United States*, 13 J.L. & ECON. REG. 7, 7–9 (2020). In addition to the guidelines noted in the paragraph, the National Institute of Standards and Technology within the U.S. Department of text has been charged with developing a voluntary Artificial Intelligence Risk Management Framework, which it embarked on developing in 2021. Artificial Intelligence Risk Management Framework, 86 Fed. Reg. 40,810, 40,810 (July 29, 2021). The head of the White House Office of Science and Technology Policy has indicated a further desire to develop its own set of principles for governmental use of AI. Eric Lander & Alondra Nelson, *Americans Need a Bill of Rights for an AI-Powered World*, WIRED (Oct. 8, 2021, 8:00 AM), https://www.wired.com/story/opinion-bill-of-rights-artificial-intelligence [https://perma.cc/4FRF-S2GY].

[237] ORG. FOR ECON. COOP. & DEV., RECOMMENDATION OF THE COUNCIL ON ARTIFICIAL INTELLIGENCE (2019), https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449 [https://perma.cc/5PH8-C8YA].

[238] Agency Use of Artificial Intelligence, 86 Fed. Reg. 6,616, 6,616 (Jan. 22, 2021).

public trust and confidence,"[239] and an "accountability framework" offered by the Government Accountability Office for agency use of AI tools.[240]

Second, agencies should seek public input on their digitization decisions. This could take the form of convening public hearings or workshops, soliciting public comments on draft proposals, or consulting with outside experts, third-party auditors, or advisory committees.[241] By encouraging public participation, agency officials can help counteract any tendencies toward the groupthink that are more likely to afflict more closed decision-making processes.[242] They may also learn about a fuller range of values and interests that could be affected by any digital algorithms they design and implement. Government officials can benefit overall from tapping into the distributed knowledge held by experts, activists, and others in the broader public at various stages of project management, from planning to ongoing use and continued improvement.[243]

Finally, when agencies contract out with third-party vendors for the development and operation of algorithmic decision-making systems, they should consider the need to access and disclose sufficient information about the algorithm, the underlying data, and the validation results to satisfy subsequent expectations for transparency. [244] In establishing contract terms and conditions with external contractors, administrators can insert provisions to ensure that contractors will provide sufficient information to the agency and the public and will adhere to basic

---

[239] Exec. Order No. 13,960, 85 Fed. Reg. 78,939, 78,939 (Dec. 3, 2020).

[240] U.S. GOV'T ACCOUNTABILITY OFF., ARTIFICIAL INTELLIGENCE: AN ACCOUNTABILITY FRAMEWORK FOR FEDERAL AGENCIES AND OTHER ENTITIES (2021).

[241] *See* Michael Sant'Ambrogio & Glen Staszewski, *Democratizing Rule Development*, 98 WASH. U. L. REV. 793, 832–33 (2021).

[242] Public participation can offer agencies a chief advantage that economist Roger Porter has attributed to a "multiple advocacy" model of presidential decision-making: namely, the full presentation of competing viewpoints. ROGER B. PORTER, PRESIDENTIAL DECISION MAKING: THE ECONOMIC POLICY BOARD 241–47 (1982). Participation can also reinforce the "active open-mindedness" that is important for successful decision-making in any organizational setting. PHILIP E. TETLOCK & DAN GARDNER, SUPERFORECASTING: THE ART AND SCIENCE OF PREDICTION 126–27, 207–08 (2015).

[243] *See* Cary Coglianese, Heather Kilmartin & Evan Mendelson, *Transparency and Public Participation in the Federal Rulemaking Process*, 77 GEO. WASH. L. REV. 924, 932 (2009).

[244] *See, e.g.*, Coglianese & Lehr, *Transparency*, *supra* note 1, at 21; Cary Coglianese & Erik Lampmann, *Contracting for Algorithmic Accountability*, 6 ADMIN. L. REV. ACCORD 175, 186 (2021); David S. Rubenstein, *Acquiring Ethical AI*, 73 FLA. L. REV. 747, 799–803 (2021). Consideration should also be paid to privacy protections for any data shared between contractors and to the use of any privacy-enhancing technology. KAITLIN ASROW & SPIRO SAMONAS, FED. RSRV. BANK OF S.F., PRIVACY ENHANCING TECHNOLOGIES: CATEGORIES, USE CASES, AND CONSIDERATIONS (2021), https://www.frbsf.org/economic-research/events/2021/august/bard-harstad-climate-economics-seminar/files/Privacy-Enhancing-Technologies-Categories-Use-Cases-and-Considerations.pdf [https://perma.cc/JJ7B-8Q5L] (discussing various forms of privacy-enhancing technologies).

principles of responsible action in the development of algorithmic tools.[245] Furthermore, given that human frailties can affect all human decisions—including the decision of how and whether to procure digital services—administrators should remain vigilant and avoid being unduly persuaded by contractors' sale pitches.[246]

In recommending careful and robust planning, public participation, and procurement practices, we do not mean to suggest that agency officials must give equal rigor to these implementation strategies in every case.[247] To the contrary, just as agencies are expected to conduct more extensive regulatory impact analyses for more significant rulemakings, the amount of time and effort devoted to planning for digitization can and should vary as well. The nature and extent of public participation can also vary depending on the use case for a digital system. If a digital system is intended to guide enforcement targeting, agencies may be fully justified in not openly inviting comment from the regulated industry or even from the general public. But this would not preclude the agency from seeking to retain a third-party auditor or setting up a peer review process involving outside experts who have entered into confidentiality agreements.

In general, the degree of time and rigor that agencies devote to planning, public participation, and procurement provisions can vary depending on the overall level of risk a government agency would likely face with a particular use case for a digital algorithm.[248] That level of organizational risk will be affected by two major

---

[245] Lavi M. Ben Dor & Cary Coglianese, *Procurement as AI Governance*, 2 IEEE TRANSACTIONS ON TECH. & SOC'Y 192, 194 (2021).

[246] *See* Omer Dekel & Amos Schurr, *Cognitive Biases in Government Procurement – An Experimental Study*, 10 REV. L. & ECON. 169, 170–71 (2014) (describing systemic biases influencing competitive bidding in governmental contracts).

[247] What Porter has to say about structuring White House decision-making applies in any governmental context, including agency decision-making about the use of digital tools: "Different circumstances require different organizational responses. An executive should weigh carefully the strengths and limitations of alternative decision-making processes in fitting them to particular circumstances and available resources." PORTER, *supra* note 242, at 252.

[248] As our aim in this section is to offer guidance to decision-makers within administrative agencies, the overarching risk considered here is that presented to the governmental entity contemplating a shift to the use of a digital algorithm. This is not to suggest that an agency's decision-making should be devoid of consideration of the risks posed by a contemplated use to affected individuals or to society overall. On the contrary, the consideration of these risks should be paramount. But, consistent with our analysis, a shift to digital algorithms might actually lower the risks to affected individuals or society when compared with a status quo based on human algorithms—and yet, even so, a government agency could still face risks of controversy and legal contestation associated with making such a shift. Those organizational risks will necessitate greater attention to the issues of planning, participation, and procurement highlighted in this subsection. For a helpful discussion of the differences between organizational risk, such as to governmental entities, and the risks to society, see generally GREG PAOLI & ANNE WILES, PENN PROGRAM ON REGUL., KEY ANALYTIC CAPABILITIES OF A BEST-IN-CLASS REGULATOR (2015),

factors: the degree to which machine learning determines agency action, and the stakes, financial and otherwise, associated with the use case in question.

When it comes to determining the degree to which a digital algorithm determines an agency's action, we distinguish different ways that the results of a machine-learning algorithm could play a role:

- *Input*: The result produced by a digital algorithm could provide information to the human agency decision-maker, making the algorithm but one factor in the agency's decision.
- *Default*: A digital algorithm could be part of an automated system that generates a default decision that can be overridden by a human—a human-in-the-loop system.
- *Decision*: A digital algorithm could make a final decision subject only to judicial review—a human-out-of-the-loop system.

All things being equal, agencies can expect that uses of machine learning that only provide inputs into agency decisions will pose fewer organizational risks compared with uses that generate defaults or make decisions.[249]

Second, the higher the stakes of the action to which machine learning is directly connected, the higher the risk to the agency.[250] Among the uses with the least significant stakes will likely be those that assist with or perform only internal staff functions at an agency. For example, consider an IT department within a government agency that chooses to deploy a machine-learning algorithm as part of a chatbot that answers calls from staff for technology assistance. That chatbot could work autonomously to process password reset requests on its own, without any

---

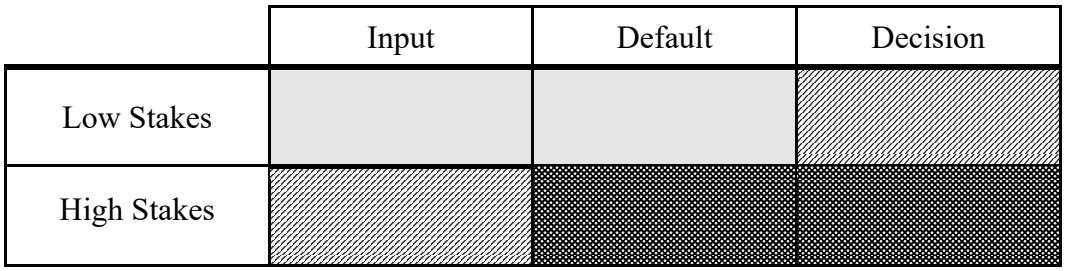
https://www.law.upenn.edu/live/files/4710-paoliwiles-ppr-researchpaper062015pdf [https://perma.cc/ZM3Z-6KPN].

[249] Again, the notion of risk here is that to the governmental entity rather than to society or to affected individuals. For all the reasons articulated in Parts I and II, the risks of error and of adverse consequences to affected individuals or society may well be markedly greater when machine learning only provides an input into otherwise flawed human judgment.

[250] The European Union has proposed making similar distinctions between high-risk and low-risk uses of AI and then imposing greater regulatory obligations on those organizations that develop high-risk forms of AI. *See generally Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts*, COM (2021) 206 final (Apr. 21, 2021) (outlining the proposal). What the European Union proposal contemplates by "risk" approximates what we discuss here as the "stakes" associated with any particular use case. But by using the term "stakes," we self-consciously contemplate the possibility that a shift to an AI-based system in high-stakes circumstances might lower the probability of error and thus *reduce* the level of risk (understood as probability multiplied by the consequences) to those individuals or entities affected by the AI system. The use of AI could perhaps even convert otherwise high-risk circumstances to ones of low-risk for affected individuals or entities. Nevertheless, for the government agency, the existence of high stakes in the form of substantial potential consequences to the affected individuals could still present the agency with greater organizational risk of conflict and controversy as it contemplates a shift even to such an efficacious AI system.

human intervention; however, notwithstanding the system's full level of determination, the stakes to members of the public could hardly be lower.[251] On the other hand, digital systems that help process applications for licenses or permits of high economic value to private businesses will necessarily involve high stakes—and thus may pose some considerable risk of conflict, even if only used as an input or default decision.

Figure 1: Agency Risks from Use of Machine Learning



Combining the two factors (level of the stakes and the degree of decisional determination), Figure 1 visualizes the risks arising from different uses of machine learning.[252] The degree of shading indicates the degree of caution and care that agencies should use when designing and deploying digital algorithms: the darkest shaded cells pose the greatest risks and imply the need for the greatest rigor and care; the lightest shaded cells pose the least risk and do not demand as extensive planning, public participation, or procurement protocols.

For instance, USPS's use of machine learning to help read handwriting when sorting letters and packages would fall within the *low-stakes* row and the *default* column because a postal worker can always intervene to redirect a mistakenly sorted piece of mail. On the other hand, the use of machine learning as part of a digital system to make criminal sentencing recommendations would clearly fall into the *high-stakes* row. But the risk of such a system would be reduced if the results of a digital algorithm only provide judges with one of many factors in a sentencing decision. In *State v. Loomis*,[253] the Wisconsin Supreme Court upheld the state's use of a risk assessment algorithm in the sentencing process in large part because it was merely one

[251] *See* Jessica Mulholland, *Chatbots Debut in North Carolina, Allow IT Personnel To Focus on Strategic Tasks*, GOV'T TECH. (Oct. 11, 2016), https://www.govtech.com/Chatbots-Debut-in-North-Carolina-Allow-IT-Personnel-to-Focus-on-Strategic-Tasks.html [https://perma.cc/8FJA-CXB6].

[252] Although this figure uses discrete cells for ease of illustration, both axes should be conceived as continua: from low stakes to high stakes, and from low levels of determination to high levels.

[253] State v. Loomis, 881 N.W.2d 749 (Wis. 2016).

input into the sentencing decision.[254] The court specifically emphasized that the sentencing decision in Loomis's case "was supported by other independent factors" and that the algorithm's "use was not determinative."[255]

Figure 1 is a heuristic that is intended to guide agency officials in thinking about their risk management of digital algorithms. The Figure is not itself determinative of when to use digital algorithms. Even when a digital algorithm would be decisive in high-stakes matters, this would not mean that the algorithm should be avoided. To the contrary, the heightened stakes may well make it more imperative for an agency to determine if a digital algorithm would make a significant improvement in accuracy, consistency, speed, or other performance goals. After all, when the stakes are high, the government should do all it can to maximize its decision-making performance—and sometimes the need for high performance will weigh in favor of machine learning if a digital algorithm will yield better outcomes than the human one.[256] Even in those contexts, it will be important for agencies to manage the potential risks of digital deployment by engaging in careful planning and validation efforts, close review of procurement provisions, and appropriate forms of public engagement.

CONCLUSION

Administrative agencies face choices about whether and when to rely on automated decision-making systems. The increasing use of machine-learning algorithms to drive automation in business, medicine, transportation, and other facets of society portends a future of increased use of machine-learning tools by government. Indeed, already government agencies have been developing and relying upon digital algorithms to assist with enforcement, benefits administration, and other important government tasks.

Moving toward governance aided by digital algorithms naturally gives rise to concerns about how these new digital tools will affect the effectiveness, fairness, and openness of governmental decision-making. This Article shows that concerns about machine-learning systems should be kept in perspective. The status quo that relies on human algorithms is itself far from perfect. If the responsible use of machine learning can usher in a government that—at least for certain uses—achieves better results than the status quo at constant or even fewer costs, then both governmental officials and the public would do well to support such use.

---

[254] *See id.* at 753.

[255] *Id.*

[256] *See generally* Coglianese & Hefter, *supra* note 233 (discussing both positive and negative consequences of AI decision-making and contemplating a shift in social acceptance of algorithmic tools by governmental entities).

The challenge for agencies will be to decide when and how to use digital algorithms to reap their advantages. Agency officials should take appropriate caution when making decisions about digital algorithms—especially because these decisions can be affected by the same foibles and limitations that can affect any human decision. Officials should consider whether a potential use of a digital algorithm will satisfy the general preconditions for the success of such algorithms, and then they should seek to test whether such algorithms will indeed deliver improved outcomes. With sound planning and risk management, government agencies can make the most of what digital algorithms can deliver by way of improvements over existing human algorithms.