2019

A Right to a Human Decision

Aziz Z. Huq

A Right to a Human Decision

Aziz Z. Huq*

[105 Virginia Law Review – (forthcoming May 2020)]

Abstract

Recent advances in computational technologies have spurred anxiety about a shift of power from human to machine decision-makers. From prison sentences to loan approvals to college applications, corporate and state actors increasingly lean on machine learning tools (a subset of artificial intelligence) to allocate goods and to assign coercion. Machine-learning tools are perceived to be eclipsing, even extinguishing, human agency in ways that sacrifice important individual interests. An emerging legal response to such worries is a right to a human decision. European law has already embraced the idea in the General Data Protection Regulation. American law, especially in the criminal justice domain, is already moving in the same direction. But no jurisdiction has defined with precision what that right entails, or furnished a clear justification for its creation.

This Article investigates the legal possibilities of a right to a human decision. I first define the conditions of technological plausibility for that right as applied against state action. To understand its technological predicates, I specify the margins along which machine decisions are distinct from human ones. Such technological contextualization enables a nuanced exploration of why, or indeed whether, the gaps that do separate human and machine decisions might have normative import. Based on this technological accounting, I then analyze the normative stakes of a right to a human decision. I consider three potential normative justifications: (a) an appeal to individual interests to participation and reason-giving; (b) worries about the insufficiently reasoned or individuated quality of state action; and (c) arguments based on negative externalities. A careful analysis of these three grounds suggests that there is no general justification for adopting a right to a human decision by the state. Normative concerns about insufficiently reasoned or accurate decisions, which have a particularly powerful hold on the legal imagination, are best addressed in other ways. Similarly, concerns about the ways that algorithmic tools create asymmetries of social power are not parried by a right to a human decision. Indeed, rather than firmly supporting a right to a human decision, available evidence tentatively points toward a countervailing 'right to a well-calibrated machine decision' as ultimately more normatively well- grounded.

* Frank and Bernice J. Greenberg Professor of Law, University of Chicago Law School. Thanks to Faith Laken for terrific research aid. Thanks to Tony Casey, David Driesen, Lauryn Gouldin, Daniel Hemel, Darryl Li, Anup Malani, Richard McAdams, Eric Posner, Julie Roin, and Lior Strahilevitz for thoughtful conversation. Workshop participants at the University of Chicago, the University of Houston, and Syracuse University School of Law also provided thoughtful feedback. All errors are mine.

Table of Contents

A Right to a Human Decision

Introduction

Every tectonic technological change—from the first grain domesticated to the first smartphone set abuzz[1]—begets a new society. Among the birth pangs of a new society are novel anxieties about how power is distributed—how it is to be gained, and how it will be lost. A spate of sudden advances in the computational technology known as machine learning has stimulated the most recent rush of inky public anxiety. These new technologies apply complex algorithms,[2] called machine-learning instruments, to vast pools of public and government data so as to execute tasks previously beyond mere humans' ability.[3] Corporate and state actors increasingly lean on these tools to make "decisions that affect people's lives and livelihoods – from loan approvals, to recruiting, legal sentencing, and college admissions."[4]

As a result, many people feel a loss of control over key life decisions.[5] Machines, they fear, resolve questions of critical importance on grounds that are beyond individuals' ken or control.[6] As a result, individuals experience a loss of elementary human agency, and a corresponding vulnerability to an inhuman and inhumane machine logic. For some, "the very idea" of an algorithmic system making

---

[1] For recent treatments of these technological causes of social transformations, see JAMES C. SCOTT. AGAINST THE GRAIN: A DEEP HISTORY OF THE EARLIEST STATES (2018), and RAVI AGRAWAL, INDIA CONNECTED: HOW THE SMARTPHONE IS TRANSFORMING THE WORLD'S LARGEST DEMOCRACY (2018).

[2] An algorithm is simply a "well-defined set[s] of steps for accomplishing a certain goal." Joshua A. Kroll et al., *Accountable Algorithms*, 165 U. PA. L. REV. 633, 640 n.14 (2017); *see also* THOMAS H. CORMEN ET AL., INTRODUCTION TO ALGORITHMS 10 (2d ed. 2001) (defining an algorithm as "any well-defined computation process that takes some value, or set of values, as input and produces some value, or set of values, as output" (emphases omitted)). The task of computing, at its atomic level, comprises the execution of serial algorithms. MARTIN ERWIG, ONCE UPON AN ALGORITHM: HOW STORIES EXPLAIN COMPUTING 2-4 (2017).

[3] Machine learning is a general purpose technology that, in broad terms, encompasses "algorithms and systems that improve their knowledge or performance with experience." PETER FLACH, MACHINE LEARNING: THE ART AND SCIENCE OF ALGORITHMS THAT MAKE SENSE OF DATA 3 (2012); *see also* ETHEM ALPAYDIN, INTRODUCTION TO MACHINE LEARNING 2 (3d ed. 2014) (defining machine learning in similar terms). For the uses of machine learning, see Susan Athey, *Beyond Prediction: Using Big Data for Policy Problems*, SCIENCE, Feb 3, 2017, at 483, http://science.sciencemag.org/content/355/6324/483.full (noting the use of structured machine learning to solve prediction problems). I discuss the technological scope of the project, and define relevant terms, *infra* at text accompanying note 81. I will use the terms "algorithmic tools" and "machine learning" interchangeably, even though the class of algorithms is technically much larger.

[4] Kartik Hosanagar and Vivian Jair, *We Need Transparency in Algorithms, But Too Much Can Backfire*, HARV. BUS. REV., July 23, 2018; Cary Coglianese & David Lehr, *Regulating by Robot: Administrative Decision Making in the Machine-Learning Era*, 105 GEO. L.J. 1147, 1149 (2017) ("[T]oday, advanced machine-learning algorithms are creating a vastly automated society, transforming many facets of life. Many products and services, including email spam filters, medical diagnoses, product marketing, and self-driving cars, now depend on machine-learning algorithms and their ability to deliver astonishing forecasting power and speed.").

[5] Shoshana Zuboff, *Big Other: Surveillance Capitalism and the Prospects of an Information Civilization*, 20 J. INFO. TECH. 75, 75 (2015) (describing a "new form of information capitalism [that] aims to predict and modify human behavior as a means to produce revenue and market control").

[6] *See, e.g.,* Rachel Courtland, *The bias detectives*, SCIENCE, June 2018, 357, 357 (documenting concerns among the public about algorithmic risk scores for detecting child abuse as failing to account for an "effort to turn [a] life around").

an important decision on the basis of historical data seems intrinsically "unfair."[7] Machines, it is said, want fatally for "empathy."[8] For others, machine decisions seem dangerously inscrutable, nontransparent, and hence hazardously unpredictable.[9] Worse, governments and companies wield these tools freely to taxonomize their populations, predict individual behavior, and even manipulate behavior and preferences in ways that give them a new advantage over the human subjects of algorithmic classification. Even the basic terms of political choice seem compromised.[10] At the same time that machine learning is poised to change the basic terms of interaction between citizen and government (or big tech), advances in robotics as well as machine learning appear to be about to displace huge tranches of both blue-collar and white-collar labor markets.[11] A feared future looms characterized by massive economic dislocation, wherein people have lost control many central life choices, and basic consumer and political preferences are no longer really one's own.

This article is about one nascent and still inchoate legal response to these fears: the possibility that, under certain conditions, one has a right to a human decision, rather than a decision reached by a purely automated process (which I call a 'machine decision'). European law has embraced the idea. American law, especially in the criminal justice domain, is flirting with it.[12] My aim in this article is to test this burgeoning proposal, to investigate its relationship with technological possibilities, and to ascertain whether it is a cogent response to growing distributional, political, and epistemic anxieties. Taking seriously concerns about equity, privacy, and power raised by an algorithmic future, I aim here to analyze whether a right to a human decision by state actors is a cogent response.

To motivate the inquiry, consider some of the anxieties unfurling already in public debate: A nursing union, for instance, launches a campaign urging patients to demand human medical judgments rather than technological assessment.[13] And a majority of patients surveyed in a 2018 Accenture survey flatly refused to be treated by a machine, rather than a human.[14] When California replaced money bail

---

[7] Reuben Binns et al., *'It's Reducing a Human Being to a Percentage': Perceptions of Justice in Algorithmic Decisions*, PROC. OF THE 2018 CHI CONFERENCE ON HUMAN FACTORS IN COMPUTING SYSTEMS 377, 385 (2018).

[8] VIRGINIA EUBANKS, AUTOMATING INEQUALITY: HOW HIGH-TECH TOOLS PROFILE, POLICE, AND PUNISH THE POOR 30 (2018).

[9] Will Knight, *The Dark Secret at the Heart of AI*, MIT TECH. REV., 2017, April 11, 2017, https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/ ("The computers that run those services have programmed themselves, and they have done it in ways we cannot understand. Even the engineers who build these apps cannot fully explain their behavior.").

[10] *See, e.g.,* Daniel Kreiss and Shannon C. McGregor, *Technology firms shape political communication: The work of Microsoft, Facebook, Twitter, and Google with campaigns during the 2016 US presidential cycle*, 35 POL. COMM. 155, 156-57 (2018) (describing the role of social media employees in shaping campaigns).

[11] For what has become the standard view, *see* Larry Elliott, *Robots will take our jobs. We'd better plan now, before it's too late*, GUARDIAN, Feb. 1, 2018, https://www.theguardian.com/commentisfree/2018/feb/01/robots-take-our-jobs-amazon-go-seattle. For a more nuanced account, see MARTIN FORD, RISE OF THE ROBOTS: TECHNOLOGY AND THE THREAT OF A JOBLESS FUTURE 282-83 (2017) (offering a more nuanced and conditional analysis).

[12] *See infra* text accompanying note 81.

[13] National Nurses United, *When it matters most, insist on a registered nurse*, May 14, 2013, https://www.nationalnursesunited.org/insist-registered-nurse.

[14] Accenture, *Accenture 2018 Consumer Survey on Digital Health*, https://www.accenture.com/us-en/insight-new-2018-consumer-survey-digital-health?utm_source=newsletter&utm_medium=email&utm_campaign=newsletter_axiosvitals&stream=top-stories

with a "risk-based pretrial assessment" tool, state court judges warned that "[t]echnology cannot replace the depth of judicial knowledge, experience, and expertise in law enforcement that prosecutors and defendants' attorneys possess."[15] In 2018, the city of Flint, MI, discontinued the use of a highly effective machine learning tool to identify defective water pipes, reverting under community pressure to human decision-making with a far lower hit rate for detecting defective pipes.[16] Or consider the worry congealed in an anecdote told by data scientist Cathy O'Neill: An Arkansas women Catherine Taylor is denied federal housing assistance because she fails an automated, "web-crawling, data-gathering" background check.[17] It is only when "one conscientious human being" takes the trouble to look into the quality of this machine result that it is discovered that Taylor has been red-flagged in error.[18] O'Neill's anecdote nicely captures the fear that machines will be unfair, incomprehensive, or incompatible with the flexing of elementary human agency.

Varying judgments about the value of human as opposed to automated decision-making are also apparent in the divergent approaches major aircraft manufacturers—and their regulators—take to "artificial intelligence" autopilots for commercial airlines.[19] Whereas Airbus favors heavy automation, Boeing vests more control in human pilots[20] (although perhaps not always enough?[21]). American law favors the human over the machine. Autopilot functionalities that cannot be "quickly and positively disengaged by the pilots to prevent it from interfering with their control" are prohibited in federally regulated planes.[22] In effect, the regulation fashions a narrowly gauged right to a human pilot, at least in the last instance.

The most important formulation of a right to a human decision to date is found in European law. In April 2016, the European Parliament enacted a new regime of data protection in the form of

---

[15] Quentin L. Kopp, *Replacing Judges with Computers is Risky*, HARV. L. REV. BLOG, Feb 18, 2018, https://blog.harvardlawreview.org/replacing-judges-with-computers-is-risky/.

[16] Alexis C. Madrigal, *How a Feel-Good AI Story Went Wrong in Flint*, THE ATLANTIC, Jan. 3, 2019, https://www.theatlantic.com/technology/archive/2019/01/how-machine-learning-found-flints-lead-pipes/578692/

[17] CATHY O'NEILL, WEAPONS OF MATH DESTRUCTION: HOW BIG DATA INCREASES INEQUALITY AND THREATENS DEMOCRACY 152-53 (2016).

[18] *Id.* at 153.

[19] Jack Stewart, *Don't Freak Out Over Boeing's Self-Flying Plane—Robots Already Run the Skies*, WIRED, June 9, 2017, https://www.wired.com/story/boeing-autonomous-plane-autopilot/.

[20] *Id.*

[21] James Glanz et al., *Why Regulators Fear the Two Boeing 737 Crashes Happened for the Same Reason*, N.Y. TIMES, Mar. 13, 2019, https://www.nytimes.com/interactive/2019/03/13/world/boeing-737-crash-investigation.html.

[22] 14 C.F.R. § 21.1329(a)(1); *see also* 14 C.F.R. § 91.3 ("The pilot in command of an aircraft is directly responsible for, and is the final authority as to, the operation of the aircraft."). Fatal crashes do occur as a result of automated systems' failure. In June 2008, Air France Flight 447 crashed into the Atlantic Ocean, killing all 228 people aboard. The investigation subsequently carried out by French aviation authorities found that the sequence of events leading to the accident included unanticipated disconnection by the autopilot. Bureau d'Enquêtes et d'Analyses, *Final Report on the accident on 1st June 2009 to the Airbus A330-203 registered F-GZCP operated by Air France flight AF 447 Rio de Janeiro*, 17 (July 2012), https://www.bea.aero/docspa/2009/f-cp090601.en/pdf/f-cp090601.en.pdf.

a General Data Protection Regulation ("GDPR").[23] Unlike the legal regime it superseded,[24] the GDPR as implemented in May 2018 is legally mandatory even in the absence of implementing legislation by member states of the European Union ("EU"). Hence, it can be directly enforced in court through hefty financial penalties.[25] Article 22 of the GDPR endows natural individuals with "the right not to be subject to a decision based solely on automated processing including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her."[26] That right covers private and some (but not all) state entities.[27] On its face, it fashions an opt-out of quite general scope from automated to human decision-making.[28]

The GDPR also has extraterritorial effect. It reaches platforms such as Google and Facebook that offer services within the EU.[29] And American law is also making tentative moves toward a similar right to a human decision. In 2016, for example, the Wisconsin Supreme Court held that an algorithmically generated risk score "may not be considered as the determinative factor in deciding whether the offender can be supervised safely and effectively in the community" as a matter of due process.[30] That decision precludes full automation of bail determinations. There must be a human judge in the loop. The Wisconsin Court's holding is unlikely to prove unique. State deployment of machine-learning has, more generally, elicited sharp complaint sounding in procedural justice and fairness terms.[31] Further, the Sixth Amendment's right to a jury trial has to date principally been

---

23 Commission Regulation 2016/679, 2016 O.J. (L 119), https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX% 3A32016R0679 [hereinafter "General Data Protection Regulation" or "GDPR"]; *see also* Christina Tikkinen-Piri, Anna Rohunen, and Jouni Markkula, *EU General Data Protection Regulation: Changes and implications for personal data collecting companies*, 34 COMP. L. & SEC. REV. 134, 134-35 (2018) (documenting enactment process of GDPR).

24 *See* Directive 95/46, of the European Parliament and of the Council of 24 October 1995 on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data, art. 1, 1995 O.J. (L 281) (EC) [hereinafter "Directive 95/46"].

25 Bryce Goodman & Seth Flaxman, *European Union regulations on algorithmic decision-making and a "right to explanation,"* arXiv preprint 1606.08813, at 2 (2016) (explaining the difference between a nonbinding Directive and a legally binding Regulation under European law).

26 GDPR §22(1).

27 GDPR §4(7) & (8) (defining "controller and "processor" as key scope terms). The Regulation, however, does not apply to criminal and security investigations. *See* GDPR §2(d).

28 As I explain below, this is not the only provision of the GDPR that can be interpreted to create a right to a human decision. *See infra* text accompanying notes 63 to 68.

29 There is sharp divergence in the scholarship over the GDPR's extraterritorial scope, which ranges from the measured, see Griffin Drake, *Navigating the Atlantic: Understanding EU Data Privacy Compliance Amidst a Sea of Uncertainty*, 91 S. CAL. L. REV. 163, 166 (2017) (documenting new legal risks to American companies pursuant to the GDPR), to the alarmist, see Mira Burri, *The Governance of Data and Data Flows in Trade Agreements: The Pitfalls of Legal Adaptation*, 51 U.C. Davis L. Rev. 65, 92 (2017) ("The GDPR is, in many senses, excessively burdensome and with sizeable extraterritorial effects.").

30 State v. Loomis, 881 N.W.2d 749, 760 (2016).

31 *See, e.g.,* Julia Angwin, Jeff Larson, Surya Mattu & Lauren Kirchner, *Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And It's Biased Against Blacks*, PROPUBLICA (May 23, 2016), https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing [https://perma.cc/6L7T-ELPG] (criticizing machine-learning instruments in the criminal justice context).

deployed to resist *judicial* fact-finding.[32] But there is no conceptual reason why the Sixth Amendment could not be invoked to preclude at least some forms of algorithmically generated outcomes in the criminal sentencing context. Indeed, it would seem to follow a fortiori that a right that precludes a jury's substitution with a judge would also block its displacement by a mere machine.

These are just tips of the iceberg. We are in the early days yet so far as machine learning's spread goes. The first truly functional "deep learning" instruments, capable of complex visual and textual pattern recognition, were tested successfully only a decade ago.[33] Many more human capabilities will prove redundant over time.[34] The impact of potential displacements is likely to be "profound."[35] In 2016, for instance, a team of researchers at University College, London, developed a machine-learning tool to predict decisions of the European Court of Justice with a success rate of 79 percent.[36] The time of machine judges is not yet upon us. But it awaits at the horizon.[37] So even if the right to a human decision seems fanciful today, the day will soon come when it seems the sole riposte to an irremediable technological sea change. So the time is ripe to ask what, exactly, it means.

In this Article, I first situate a right to a human decision first in a contemporary technological context. I can thereby specify with precision the margins along which machine decisions are distinct from human ones. My focus is on a group of related technologies known as machine learning. This is

---

[32] *See, e.g.*, Apprendi v. New Jersey, 530 U.S. 466, 477 (2000) (explaining that the Fifth and Sixth Amendments "indisputably entitle a criminal defendant to a jury determination that [he] is guilty of every element of the crime with which he is charged, beyond a reasonable doubt" (alteration in original) (internal quotation marks omitted) (quoting United States v. Gaudin, 515 U.S. 506, 510 (1995))).

[33] Jürgen Schmidhuber, *Deep learning in neural networks: An overview*, 61 NEURAL NETWORKS 85, 96-97 (2015); Yoshua Bengio, *Machines Who Learn*, SCIENTIFIC AMERICAN, June 2016, at 46, 48 (noting that deep learning "came into its own" in 2005). For an account of deep learning, see Yann LeCun et al., *Deep Learning*, 521 SCIENCE 436, 437 (2015) (describing deep learning in terms of its use of stochastic gradient ascent to improve predictive quality continuously); *see also infra* text accompanying note 116.

[34] Yet it is easy to overstate the magnitude of likely short-term changes, and miss more consequential long-term dynamics, see Rodney Brooks, *The Seven Deadly Sins of AI Prediction*, MIT TECH. REV., Oct 6, 2017, https://www.technologyreview.com/s/609048/the-seven-deadly-sins-of-ai-predictions/.

[35] *See* Erik Brynjolfsson and Tom Mitchell, *What can machine learning do? Workforce implications*, 358 SCIENCE 1530, 1530-31 (2017) (predicting "an even larger and more rapid transformation [in labor markets] due to recent advances in machine learning"). The number of companies using machine learning of various sorts has grown extremely rapidly in the past three to four years. Andrew Nusca, *The Current State of Artificial Intelligence, According to Nvidia's CEO*, FORTUNE, Mar. 22, 2016, 9:00 AM), http://fortune.com/2016/03/22/artificial-intelligence-nvidia/.

[36] Nikolaos Aletras, Dimitrios Tsarapatsanis, Daniel Preoţiuc-Pietro & Vasileios Lampos, *Predicting judicial decisions of the European Court of Human Rights: A natural language processing perspective*, 2 PEER J. COMP. SCI. e93 (2016); *see also* Chris Johnston, *Artificial intelligence judge developed by UCL scientists*, THE GUARDIAN, Oct. 24, 2016, https://www.theguardian.com/technology/2016/oct/24/artificial-intelligence-judge-university-college-london-computer-scientists#https://www.theguardian.com/technology/2016/oct/24/artificial-intelligence-judge-university-college-london-computer-sci ("We don't see AI replacing judges or lawyers, but we think they'd find it useful for rapidly identifying patterns in cases that lead to certain outcomes." (quoting creator of algorithm)).

[37] Sean Braswell, *All Rise for Chief Justice Robot!*, OZY (June 7, 2015), http://www.ozy.com/immodest-proposal/all-rise-for-chief-justice-robot/41131; Eugene Volokh, *Chief Justice Robots*, 68 DUKE L.J. 1135, 1142 (2019) (developing the idea of "creating an AI judge that we can use for legal decisions"). For a nuanced discussion of the possibility that "machine-learning-based decisions" as administrative adjudicators, see Coglianese & Lehr, *supra* note 4, at 1186-91. For a cogent argument against these ideas, see Emily Berman, *A Government of Laws and Not of Machines*, 98 B.U. L. Rev. 1277 (2018).

the form of artificial intelligence diffusing most rapidly today.[38] A right to a human decision cannot be defined or evaluated without understanding the technical differences between human decision-making and decisions reached by these machine learning technologies. Indeed, careful analysis of how machine learning is designed and implemented reveals that the idea of a right to a human decision is not straightforward in practice. Nor can it be justified on certain grounds given present technological constraints. Some common complaints about machine decisions' opacity and autonomy, I show, are compromised or thwarted by technical facts. Claims about a right to human input, I suggest, are better understood to turn on the timing, and not the sheer fact, of such involvement.

With this technical foundation in hand, I analyze the right to a human decision in relation to the normative ends it might plausibly be understood to further. I analyze three slightly different justifications. My first line of analysis takes up the interests of an individual exposed to a machine decision. The most pertinent of these hinge upon an individual's participation in decision-making and her opportunity to offer reasons. A second analytic salient tracks ways that a machine instrument might be intrinsically objectionable (without accounting for an individual legal right). I focus here on worries about the absence of individualized consideration and a machine's failure to offer reasoned judgments. Finally, I consider dynamic, system-level effects in the form of negative spillover, in particular in relation to social power. None of these kinds of lines of argument in the end provides sure ground for a legal right to a human decision. Rather, I offer the (so far seemingly counter-intuitive) possibility of *a right to a well-calibrated machine decision*. That is, machine decisions are often capable of classification with a smaller number of false positives and false negatives than humans, and have the potential to act with fewer distorting biases. But, as critics fairly observe, theory is not practice. Many of the instruments implemented by government tend to be highly flawed.[39] A right to a well-calibrated machine decision, and not a reversion to equally flawed human decision-making, as plausible and as powerful in normative terms as its more familiar antonym. This conclusion, to be clear, is emphatically not a repudiation of the distributional, political, and normative objections raised against machine learning's spread. I rather hope to show why a right to a human decision is not a fitting response. By clarifying its tight normative bounds, I hope to head-off legal investments in an inapt idea before it becomes irrevocably embedded in the law.

My analysis here focuses on how such a right might apply to state action, and not private action for three reasons. First, salient U.S. legal frameworks, unlike the GDPR's coverage, are largely (although not exclusively) trained on state action. Accordingly, a focus on state action makes sense in terms of explaining and evaluating the current U.S. regulatory landscape. Second, whereas the range of private uses of algorithmic tools is vast and heterogenous, the public uses of algorithmic tools are to date more homogenous. This makes them a more straightforward object of analysis. Algorithms are now deployed in a far broader array of private activities. These range from Google's PageRank

---

[38] *See infra* text accompanying note 95 (defining machine learning). I am not alone in this focus. Legal scholars are paying increasing attention to new algorithmic technologies. For leading examples, see Kroll et al., *supra* note 2; Michael L. Rich, *Machine Learning, Automated Suspicion Algorithms, and the Fourth Amendment*, 164 U. PA. L. REV. 871, 929 (2016) (developing a "framework" for integrating machine-learning technologies into Fourth Amendment analysis); Andrew Guthrie Ferguson, *Big Data and Predictive Reasonable Suspicion*, 163 U. PA. L. REV. 327, 383-84 (2015) (similar); Kate Crawford & Jason Schultz, *Big Data and Due Process: Toward A Framework to Redress Predictive Privacy Harms*, 55 B.C. L. REV. 93, 109 (2014) (arguing for "procedural data due process [to] regulate the fairness of Big Data's analytical processes with regard to how they use personal data (or metadata …)").

[39] AI Now, *AI Now Report 2018*, at 18-22 (2018), https://ainowinstitute.org/reports.html.

algorithm,[40] to "fintech" used to generate new revenue streams,[41] to medical instruments used to calculate stroke risk,[42] to engineers' identification of new stable inorganic compounds.[43] Algorithmic tools are also embedded within new applications such as voice recognition software, translation software, and visual recognition systems.[44] In contrast, the state is to date an unimaginative user of machine learning. Criminal justice appears to be a leading adaptor.[45] Other regulatory domains so far lag behind. In a handful of states, welfare bureaucracies have started to employ algorithmic tools to sort between recipients.[46] The same is true at the federal level. The Environmental Protection Agency, for instance, uses machine learning to evaluate the effects of some toxins.[47] The Internal Revenue Service uses a machine-learning instrument to predict fraud and abuse.[48] Because these adoptions seem to be the exception rather than the rule, state uses of machine learning provide a sharper target of inquiry. Third, where the state does use algorithmic tools, it often results in deprivations of liberty, the freedom of movement, bodily integrity, or basic income. Accordingly, these machine decision present arguably the most compelling circumstances for adopting a right to a human decision. By focusing the inquiry where a right to a human decision is prima facie most compelling, I can most effectively stress test arguments for and against that right.

The Article proceeds in three steps. Part I catalogs ways in which law has crafted, or could craft, a right to a human decision. This taxonomical enterprise demonstrates that such a right is far from hypothetical or fanciful. Part II defines the class of computational tools to be considered, and explores the manner in which such instruments are distinct from human decisions. While this Part is

---

40 *See, e.g.,* David Segal, *The Dirty Secrets of Search*, N.Y. TIMES, Feb. 12, 2011 at BU1.

41 *See* Falguni Desai, *The Age of Artificial Intelligence in Fintech*, FORBES, June 30, 2016, 10:42 PM, http://www.forbes.com/sites/falgunidesai/2016/06/30/the-age-of-artificial-intelligence-in-fintech [https://perma.cc/EK89-DD3Y] (describing how fintech firms use artificial intelligence to improve investment strategies and analyze consumer financial activity).

42 *See, e.g.*, Benjamin Letham, Cynthia Rudin, Tyler H. McCormick, & David Madigan, *Interpretable classifiers using rules and Bayesian analysis: Building a better stroke prediction model*, 9 ANNALS APP. STAT. 1350 (2015).

43 *See, e.g.,* Paul Raccuglia et al., *Machine-learning-assisted materials discovery using failed experiments*, 533 NATURE 73 (2016) (identifying new vanadium compounds).

44 LeCun et al., *supra* note 33, at 438.

45 *See* Aziz Z. Huq, *Racial Equity in Algorithmic Criminal Justice*, 68 DUKE L. J. 1043 (2019) (collecting examples); *see also* Andrew Guthrie Ferguson, *Policing Predictive Policing*, 94 WASH. U.L. REV. 1109, 1122-44 (2017) (providing a careful catalogue of predictive policing tools). Such usages are now extending into the immigration domain. Spencer Woodman, *Palantir Provides the Engine for Donald Trump's Deportation Machine*, INTERCEPT (Mar. 2, 2017, 1:18 PM), https://theintercept.com/2017/03/02/palantir-provides-the-engine-for-donald-trumps-deportation-machine/ (reporting that the Department of Homeland Security ("DHS") awarded a private contractor a $41 million contract to build an "Investigative Case Management" system to allow DHS to "access a vast 'ecosystem' of data to facilitate immigration officials in both discovering targets and then creating and administering cases against them").

46 *See* EUBANKS, *supra* note 8, at 14-38.

47 *See* U.S. Envtl. Prot. Agency, *ToxCast Fact Sheet* (2013), http://www.epa.gov/sites/production/files/2013-12/documents/toxcast-fact-sheet.pdf.

48 David DeBarr & Maury Harwood, *Relational Mining for Compliance Risk*, Presented at the Internal Revenue Service Research Conference (2004), http://www.irs.gov/pub/irs-soi/04debarr.pdf. For a similar effort in the food safety domain, see *Commissioner's Fellowship Program: Final Report Abstracts*, U.S. Food & Drug Admin., https://www.fda.gov/AboutFDA/WorkingatFDA/FellowshipInternshipGraduateFacultyProgr/CommissionersFellowshipProgram/ucm413253.htm.

grounded in description, it also analyzes ways in which technological context either shapes or forecloses certain formulations of the right. Part III then turns to the potential normative foundations of such a right. It provides a careful taxonomy of those grounds. It then shows why they all fall short. Finally, a brief conclusion inverts the analytic lens of the paper to consider whether a *right to a well-calibrated machine decision* can be imagined, and even defended on more persuasive terms than a right to a human decision.

I. Legal Articulations of a Right to a Human Decision

This Part documents ways in which law creates something like a right to a human decision. I use the term "law" here capaciously to extend beyond U.S. law to European law, and to range across both private and public law domains, capturing the regulation of state and nonstate action. I take this wide-angle view so as to develop an understanding of several aspects of this putative right: the reasons for which it is articulated; the contexts in which it is applied; and the limits with which it is hedged. That inquiry is largely descriptive. By surveying the current legal landscape, I offer a 'proof of concept' to the effect that a right to a human decision is not so outlandish a notion as to be dismissed out of hand. At the same time, the opacities and limits of current law provide further evidence of the difficulties packed into any effort to vest such a right in individuals.

A. The European Right to a Human Decision

European law has, in some form, recognized something akin to a right to a human decision since 1978. Although that right has to date not had much practical legal impact, the GDPR's enactment may generate a concrete effect. Historical antecedents to Art. 22 of the GDPR also cast some light on the difficulties of fashioning such a right, and of discerning its justifications.

1. Antecedents

An early antecedent of a right to a human decision is France's 1978 law on "computing, files, and liberties."[49] Article 2 of this law, as originally enacted, prohibited official use of automated profiling or personality screening.[50] So far as I can tell, this measure was never applied to machine learning, which, in any case, were not in general usage at the time. Seventeen years later, the European Parliament and Council promulgated the Data Protection Directive.[51] The latter lacked independent legal effect on individuals, but obliged European Union member states to enact conforming laws. Article 15 of the Directive obligated member states to create a right "not to be subject to a decision which produces legal effects … or significantly affects him and which is based solely on automated processing of data intended to evaluate certain personal aspects."[52] Countries implemented this

---

[49] Loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés, https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=LEGITEXT000006068624&dateTexte=19781031.

[50] "Aucune décision de justice impliquant une appréciation sur un comportement humain ne peut avoir pour fondement un traitement automatisé d'informations donnant une définition du profil ou de la personnalité de l'intéressé." *Id.* at Art. 2. A second clause extended the same rule to administrative decisions. *Id.*

[51] Directive 95/16.

[52] *Id.* Art. 15(1).

provision in various ways, at times differentiating between private and public actors.[53] As with the 1978 French law, however, these measures appear to have had little effect on the ground.[54] There is also little evidence of the concerns that motivated Article 15's inclusion in the Data Protection Directive.[55] Accordingly, its history yields scant guidance or illumination as to how to conceptualize or justify a right to a human decision.

2. *Article 22 of the GDPR*

In 2017, the European Commission declared that it was time to take "an essential step to strengthening citizens' fundamental rights in the digital age."[56] The result was the General Data Protection Regulation ("GDPR"). In effect since May 2018, the GDPR is a comprehensive reworking of European data privacy and protection rules. Unlike the earlier Directive, it acts directly on companies and state institutions that handle covered forms of data. It also contains penalty provisions envisaging fines running into the millions of euros.[57] Its ninety-nine articles also cover a broad array of other topics, and my discussion homes on the relevant language rather than providing a synoptic overview.

Article 22(1) of the GDPR vests natural persons with a "right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her."[58] According to the European Commission Data Protection Working Party created by the European Union, Article 22(1) applies only if "there is no human involvement in the decision process."[59] In some tension with that formulation, the Working Party guidance document also suggests that only "meaningful" and "ex post" review of an automated

---

[53] *See, e.g.,* Decreto legislative 30 giugno 2003, no, 196: Codce in Materia du Protezine dei Dati Personali, http://www.camera.it/parlam/leggi/deleghe/03196dl.htm (distinguishing public actors, and imposing an absolute prohibition on them, but creating a right to object to automated decisions by private actors).

[54] *See* Lee Bygrave, *Minding the Machine: Article 15 of the EC data protection directive and automated profiling*, 17 COMP. L. & SEC. REV. 17, 21 (2001) (describing the right as a "house of cards"). Only one German decision seems to have touched on this right. See Isak Mendoz & Lee A. Bygrave, *The Right not to be Subject to Automated Decisions based on Profiling*, *in* EU INTERNET LAW 11 n.36 (Tatiana-Eleni Synodinou et al. eds. 2017) (describing a German decision that found that credit-scoring systems fall outside the scope of Article 15 and implementing domestic legislation).

[55] Bygrave asserts that the Article was motivated by "the potential for … automisation to diminish the role played by persons in shaping important decision-making processes." Bygrave, *supra* note 54, at 18. But this is a tautology, not an explanation of why machine decisions are to be disfavored in relation to human ones.

[56] European Commission, *Questions and Answers—Data Protection Reform Package* (2017),

europa.eu/rapid/press-release_MEMO-17-1441_en.pdf

[57] Jan Philipp Albrecht, *How the GDPR will change the world*, 2 EUR. DATA PROT. L. REV. 287, 287 (2016) (describing the sanctions regime).

[58] GDPR Art. 22(1).

[59] Article 29 Data Protection Working Party, *Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679*, at 20 (Feb. 6, 2018), https://iapp.org/resources/article/wp29-draft-guidelines-on-automated-individual-decision-making-and-profiling/ [hereinafter "Working Party Guidelines"].

decision would remove it from the scope of Article 22(1).[60] The GDPR, though, does not explain how the quality of such review is to be assessed.[61] The precise range of automated machine-learning tools captured by the prohibition remains up for grabs. Article 22(2) goes on from this to exclude processing "necessary for entering into, or performance of, a contract"; otherwise authorized by "Union or Member State law"; or "based on the data subject's explicit consent."[62] To date, there does not appear to be any European state that has enacted an exception under Article 22(2). The scope of that exception, therefore, appears limited,[63] whereas the extent of the other exceptions remains open to interpretation. Depending on how the term "explicit consent" is defined, moreover, the provision might take on either a broader or a narrower scope. It is quite possible to imagine private entities acquiring such consent as a matter of course; whether state employers or welfare agencies could do the same, however, is a different matter. Accordingly, it is difficult to predict the precise scope or practical effect of Article 22(1)'s right.[64]

Article 22 is not the only element of the GDPR that might be glossed as a right against processing. Article 18 allows natural persons to "obtain restrictions" on unlawful or inaccurate data processing,[65] while Article 21's "right to object" mandates that an entity "no longer process" a person's data once "compelling legitimate grounds" have been invoked.[66] Given the broad definitions of "processing"[67] and "profiling,"[68] these other provisions might sweep widely indeed. Again, the absence of implementation by national governments or enforcement actions lodged either by private, national,

---

[60] *Id.* (excluding from Article 22(1) instances in which a human "reviews and takes account of other factors in making a final decision").

[61] It is not clear what "meaningful" supervision entails. See Michael Veale & Lilian Edwards, *Clarity, surprises, and further questions in the Article 29 Working Party draft guidance on automated decision-making and profiling*, 34 COMP. L. & SEC. REV. 398, 401(2018) ("How this expanded notion of 'solely' could practically be assessed from the point of view of the data controller or the data subject is one of the significant grey areas th[e] guidance leaves in its wake."). In addition, if an automated process does not change "legal rights" or have an "equivalent or similarly significant" effect, the Working Party suggests that it is not covered by Article 22(1). Working Party Guidelines, supra note 59, at 21-22 (noting that while targeted advertising is not typically covered, the "intrusiveness" of the targeting, an individual's expectations, and the operator's knowledge of the "vulnerabilities" of the person might render it covered by Article 22(1)).

[62] GDPR Art 22(2).

[63] The GDPR also defines consent in very narrow and demanding terms.

[64] The antecedent to the GDPR, the first Data Protection Directive, was interpreted in light of the proportionality principle employed across European public law. *See* Charlotte Bagger Tranberg, *Proportionality and data protection in the case law of the European Court of Justice* 1 INT'L. DATA PRIVACY L. 239, 239-40 (2011) (summarizing case law). The uncertainty over how proportionality review would be applied to the GDPR adds yet more difficulty to predicting the law's path.

[65] GDPR Art. 18.

[66] *Id.* Art. 21(1).

[67] This is defined in relevant part to include "any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means." *Id.* Art. 4(2).

[68] This is defined to include "any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements." *Id.* Art. 4(4).

or supranational authorities means that there remains considerable uncertainty about what any of these provisions mean, let alone whether they are properly glossed to include a right to a human decision.

The principal element of the GDPR to attract attention so far is the potential right to an explanation of algorithmic decisions, which is located elsewhere in the document.[69] The right to a human decision, whether anchored in Article 22(1) or elsewhere, has evinced contrastingly little scholarly attention to date.[70]

B. American Law and the Right to a Human Decision

There is no precise analog in U.S. law to the GDPR. This section adumbrates three legal domains in which hints can be discerned. Knitting these together reveals the inchoate shadow of a right to a human decision lurking in the interstices of federal and state law.

First, constitutional law at both the state and the federal level already creates individual rights to forms of decision-making that are inconsistent with at least certain forms of machine-learning. Perhaps the most prominent of these is the Constitution's creation of various entitlements to a jury's decision, most prominently the Sixth Amendment right to a jury trial in criminal cases.[71] Ignoring for present purposes the ubiquity of plea-bargaining, the Constitution's entitlement to a jury determination is one kind of a right to a human decision. In the original Constitution's scheme, the embedding of juries at critical instances in which state power was being exercised manifested a commitment to democracy.[72] The obligation of the citizen as a juror was necessarily "an active one" that presupposed the autonomous exercise of human judgment in the face of potentially tyrannical state power.[73] If the Sixth Amendment is violated by the substitution of a judge for a body of human citizens, it is hard to see how the pivotal decisional role in a criminal trial could be played a machine. So the right to a jury trial under the Sixth Amendment, and parallel constitutional provisions that purportedly ensure access to a jury, likely encompasses a right to a human decision. And the use of a machine-learning tool's output to increase a defendant's sentence beyond a statutory maximum would violate a defendant's Sixth Amendment right to have a jury decide those facts if the facts upon which

---

[69] There is some debate on whether the GDPR should be interpreted to create such a right. *Compare* Sandra Wachter, Brent Mittelstadt & Luciano Floridi, *Why a right to explanation of automated decision-making does not exist in the general data protection regulation*, 2 INT'L DATA PRIVACY L. 76 (2017) (arguing against the inference of a right to an explanation) *with* Andrew Selbst & Julia Powles, *Meaningful information and the right to explanation*, 4 INT'L DATA PRIVACY L. 233, *23*4 (2017) (offering a "positive conception of that right").

[70] *See* Veale & Edwards, *supra* note 61, at 400-01 (noting ambiguities in current formulation); *see also* Meg Leta Jones, *The right to a human in the loop: Political constructions of computer automation and personhood*, 47 SOC. STUD. SCIENCE 216, 224 (2017) (flagging the existence of Article 22).

[71] U.S. Const. art. III § 2, cl. 3 ("The Trial of all Crimes … shall be by Jury ….."); amd VI (right to "an impartial Jury"); amd VII (right to "trial by jury" in certain civil cases); *see also* Apprendi v. New Jersey, 530 U.S. 466, 477 (2000) (affirming the Sixth Amendment jury trial right in contradistinction to judicial fact-finding).

[72] *See* Akhil Reed Amar, *The Bill of Rights as a Constitution*, 100 YALE L.J. 1131, 1183 (1991); *see also* Mark DeWolfe Howe, *Juries as Judges of Criminal Law*, 52 HARV. L. REV. 582, 584-85 (1939) (emphasizing the broad scope of juries' decisional power on questions of both fact and law in the Founding Era).

[73] Jenny E. Carroll, *Nullification As Law*, 102 GEO. L.J. 579, 589 (2014); *see also* JEFFREY ABRAMSON, WE, THE JURY: THE JURY SYSTEM AND THE IDEAL DEMOCRACY 22-24, 68-75 (1994) (discussing the history of the development of the jury in colonial America as a political institution).

that output had been made were not themselves found by a jury.[74] In addition, the Sixth Amendment may impose a further friction (if not prohibition) on machine decisions in the form of the Confrontation Clause.[75]

Second, the idea of due process might also be basis for a human rather than a machine judgment. At its core, the idea of procedural due process is thought to entail "notice and some kind of hearing."[76] There is some debate about the timing and the content of a hearing, at least so far as the Constitution's Due Process guarantee is concerned.[77] But it is not hard to see how a question could arise whether due process is supplied by a machine decision. Indeed, it is arguably difficult to make sense of the idea of a "hearing" in the absence of a natural person who is either physically present for verbal arguments, or who reads and evaluates written submissions.[78]

To date, there has not been a frontal challenges to algorithmic tools on Sixth Amendment or due process grounds. This is perhaps because such instruments have to date been used to support human decision-making rather than formally ousting it. But the Wisconsin Supreme Court, in a 2016 decision of *State v. Loomis*, resolved a Due Process challenge hinging on a criminal defendant's challenge to a sentencing algorithm's criteria (as distinct from its very use).[79] That challenge in part rested on the defendant's limited ability to challenge the algorithm's terms and in part upon the kind of data (general rather than individualized) upon which the algorithm relied to reach its recommendation.[80] The Wisconsin Court rejected the defendant's constitutional arguments. It reasoned that the algorithm employed only publicly available data, as well as data that a defendant has supplied, and that the defendant could have denied or explained any information that was employed to develop his prediction.[81] The Court flagged the group-based, rather than individualized nature of the data used to train the sentencing algorithm in that case, but refused to view it as unconstitutional

---

[74] Booker v. United States, 543 U.S. 220, 245-46 (2005). So-called evidence-based tools are widely used in sentencing contexts already, albeit without rigorous direction and safeguards. Erin Collins, *Punishing Risk*, 107 GEO. L.J. 57, 66 (2018) (describing the deployment of actuarial sentencing tools, often "without guidance as to the purposes for which [they] may be used"). As a general matter, they appear to be employed to guide judges' exercise of discretion over sentence length within the bounds set by the Sixth Amendment. John Monahan & Jennifer Skeem, *Risk Assessment in Criminal Sentencing*, 12 ANNU. REV. CLIN. PSYCH. 489, 500-01 (2016).

[75] Andrea Roth, *Machine Testimony*, 126 YALE L. J. 1972, 2039-51 (2017).

[76] Richard H. Fallon, Jr., *Some Confusions About Due Process, Judicial Review, and Constitutional Remedies*, 93 COLUM. L. REV. 309, 330 (1993). For a class exposition of this idea, Henry J. Friendly, "*Some Kind of Hearing*," 123 U. PA. L. REV. 1267 (1975) (discussing the characteristic elements of a fair hearing and assessing their relative importance).

[77] *Cf.* United States v. Florida East Coast Ry., 410 U.S. 224, 239 (1973) ("The term 'hearing' in its legal context undoubtedly has a host of meanings.").

[78] Friendly, *supra* note 76, at 1270 ("Although the term "hearing" has an oral connotation, I see no reason why in some circumstances a "hearing" may not be had on written materials only.").

[79] State v. Loomis, 881 N.W.2d 749 (2016); *see also* Huq, *supra* note 45, at -- (discussing the sentencing instrument used in Wisconsin).

[80] As the defendant's expert witness Dr. David Thompson explained: "The Court does not know how the COMPAS compares that individual's history with the population that it's comparing them with. The Court doesn't even know whether that population is a Wisconsin population, a New York population, a California population .... There's all kinds of information that the court doesn't have, and what we're doing is we're misinforming the court when we put these graphs in front of them and let them use it for sentence." *Loomis*, 881 N.W.2d at 756-57.

[81] *Id.* at 761-62.

given that it was just one of several inputs to the defendant's sentence.[82] Such data, the Court warned, "may not be considered as the determinative factor in deciding whether the offender can be supervised safely and effectively in the community" consistent with due process.[83] Moreover, the Court warned, a machine's prediction "may not be considered as the determinative factor in deciding whether the offender can be supervised safely and effectively in the community" as a matter of due process.[84]

Finally, in April 2019, two Democratic Senators introduced a bill requiring the Federal Trade Commission to enact regulations mandating "automated decision system impact assessments."[85] For certain algorithms, that assessment is required "prior to implementation."[86] The Federal Trade Commission (FTC) must promulgate regulations to enforce this requirement, and both the FTC and Attorney Generals of the several states can enforce it.[87] Although this does not contain an individual right of action, this bill may allow suits against algorithms that may preclude their usage.

American law, in sum, does not create a right to a human decision in so many words. Rather, such a right emerges as an unexpected implication of the Constitution's protections of the jury trial right. To the extent that American regulators follow the lead of the GDPR, the foundations for such a right exist.

C. The Tentative Form of a Novel Right

It is too early to conclude that a robust legal right to a human decision exists as a practical matter in either European or American law. It would be equally premature to deny that such a right is finding some footing in both the criminal and the civil regulator domains. Article 22(1) of the GDPR will likely be the cynosure of such a right. Yet some version of that right has existed in national or supranational European law since 1978, without much practical effect. Technical advances in the capacity of machine learning, and in particular deep learning tools, since the early 2000s,[88] however, may place new strain on this status quo. New fears might spark conflict along unexpected regulatory

---

[82] *Id.* at 765 ("[T]he due process implications compel us to caution circuit courts that because COMPAS risk assessment scores are based on group data, they are able to identify groups of high-risk offenders--not a particular high-risk individual.").

[83] *Id.* at 760. Separately, the Fifth Amendment right against compelled self-incrimination may be triggered by interviews designed to elicit information from a defendant for the purpose of assigning him or her an algorithmic classification associated with a longer sentence. Cassie Deskus, Note, *Fifth Amendment Limitations on Criminal Algorithmic Decision-Making*, 21 N.Y.U. J. LEGIS. & PUB. POL'Y 237, 259-66 (2018). That constitutional question is not well understood as an adjunct to the right to a human decision, and so I leave it to one side here.

[84] State v. Loomis, 881 N.W.2d 749, 760 (2016).

[85] See Algorithmic Accountability Act, S. --, §3(1)(A), https://www.wyden.senate.gov/imo/media/doc/Algorithmic%20Accountability%20Act%20of%202019%20Bill%20Text.pdf.

[86] Id. §3(1)(A)(ii).

[87] Id. §(a)(1) & (d).

[88] *See supra* text accompanying notes 113 to 115.

margins.[89] The rate of ensuing legal change is hard to predict.[90] Prohibitory regulation of machine decisions is likely to undermine the business strategy of business entities that use some form of machine learning.[91] Predictably powerful lobbies favoring expansive use of data unhindered by a right to opt out from machine learning will be powerful adversaries of privacy advocates in both legislative and regulatory debates. At the same time, those lobbies may perceive beneficial compliance-related economies of scale in regulation such as the GDPR and the CCPA. Rather than risk a balkanized regulatory terrain, these interest groups might accept some kind of right to a human decision as a lesser evil. Obviously, a rather different political economy characterizes the criminal justice space, yielding different regulatory pathways with divergent technological equilibria.

Despite these political-economy uncertainties, some tentative conclusions can be drawn in respect to the form and force of a right to a human decision from more recent legal developments. Two merit emphasis here.

First, a right to a human decision can be verbally articulated in quite varied guises, from the explicit commitment contained in GDPR Article 22(1) to the commitment to a (human) jury decision in the Sixth Amendment. Alternatively, exposure to non-human decision-making might be minimized by regulating the sheer volume of data-flows too, though, for example, the use of individual opt-outs of data sharing.[92] This sheer variegation in legal form means that it is unclear how a right to a human decision (should it prove desirable) would best be formulated. The historical desuetude of earlier iterations of the right to a human decision in European law, and the persisting ambiguities of the GDPR's Article 22 all hint at considerable operational difficulty in creating a legally enforceable right of this flavor.

A second inference concerns transatlantic contrasts. American and European law appear to be following divergent priorities in respect to the right to a human decision. In the U.S., that principle has the most influence in criminal justice matters. By contrast, the GDPR carves out crime and security related functions from its purview. Indeed, it concentrates on data processing conducted by nonstate actors. Perhaps this divergence can be glossed as another instance of the European concern with

---

[89] But not such all fears are unwarranted. See, e.g., Edward Geist and Andrew J. Lohn, Rand Institute, *How Many Artificial Intelligence Affect the Risk of Nuclear War?* 2 (2018), https://www.rand.org/pubs/perspectives/PE296.html ("AI has the potential to exacerbate emerging challenges to nuclear strategic stability by the year 2040 even with only modest rates of technical progress.").

[90] This is obviously not because of the difficulty of predicting technological change. *Cf.* Jon ELSTER, EXPLAINING TECHNICAL CHANGE: A CASE STUDY IN THE PHILOSOPHY OF SCIENCE 7-9 (1983) (defining "technical change" as the "manufacture and modification of tools," and discussing possibility different pathways by which such change can occur). The right to a human decision involves a choice to refuse technological change. But there is nothing inexorable about technological adaption and advance, and so no reason that technologies cannot be abandoned.

[91] MIT Technology Review Insights, *Machine Learning: The New Proving Ground for Competitive Advantage*, MIT TECH. REV., March 16, 2017, https://www.technologyreview.com/s/603872/machine-learning-the-new-proving-ground-for-competitive-advantage/ (finding in a survey of businesses, that 60 percent already employed machine learning in some way, while only 5 percent had no interest in doing so in the future).

[92] See, e.g., Assembly Bill No. 375, Ch. 55, "An act to add Title 1.81.5 (commencing with Section 1798.100) to Part 4 of Division 3 of the Civil Code, relating to privacy," June 28, 2018. It remains to be seen whether such an opt-out from processing is effective. The inefficacy of consent-based strategies for vindicating online privacy does not bode well for that kind of an individuated approach.

"respect and personal dignity" working in contradistinction to an American concentration upon "values of liberty."[93] But the obscurity of Article 22's etiology in the 1978 French data protection law and the 1994 European directive tells against confident diagnosis. Whatever its origins, the net effect of this difference in domains is that the right to a human decision will be tested first in very different circumstances in the two continents, perhaps leading to quite divergent legal regimes.

One commonality unites the emerging right to a human decision in its American and European forms. In neither context have legislators or judges developed a robust theoretical account of why the particular technologies, and the distinctive modalities of inferential reasoning they entail, should be deemed objectionable. Nor have they said why a human decision should be preferred. Such an account cannot be mechanically derived from an empirical account of the differences between human and machine decisions. Nor, finally, does the GDPR (or its precursors) or the CCPA contain a robust theoretical account of the right's justification. In consequence, there is at present a theoretical gap between the emergent right and an empirically and normatively persuasive justification. It is that gap that the balance of this Article explores.

II. The Difference Between Machine Decisions and Human Decisions

To make sense of the idea of a right to a human decision, and why it might be normatively appealing, we need to clearly understand what kind of nonhuman (machine) decisions we are concerned about; how they differ from human decision-making processes; and how human and machine decisions in practice can be either distinct or entangled so as to be functionally separable or inseparable. To that end, this Part develops a brief and nontechnical account of the relevant technologies. It focuses in particular on a class of machine-learning tools (also sometimes called artificial intelligence[94]) that hold the most immediate promise for displacing human decisions. I parse the normative implications of these technical differences as a predicate to Part III's more fine-grained normative inquiry. This contextualizing account has an immediate normative payoff: It helps clarify the conditions of technical plausibility of any right to a human decision. I suggest that right is best understood as a demand for an ex post review of machine decisions. This Part's account also sets up the more extended normative inquiry of Part III.

A. Machine Learning as a Substitute for Human Decisions

A machine-learning algorithm in its most general terms solves a "learning problem ... of improving some measure of performance when executing some task through some type of training experience."[95] Such algorithms come in two forms: supervised and unsupervised.[96] Supervised machine learning algorithm are commonly prompted to define a function f(x) which produces an

---

[93] James Q. Whitman, *The Two Western Cultures of Privacy: Dignity Versus Liberty*, 113 YALE L.J. 1151, 1161 (2004).

[94] I avoid this term because it has a potentially wider and more ambiguous scope. *Cf.* STUART RUSSELL & PETER NORVIG, ARTIFICIAL INTELLIGENCE: A MODERN APPROACH 2-14 (3d ed. 2013) (offering a series of alternative definitions of AI that include thinking and acting humanly as well as rationally).

[95] M.I. Jordan & T.M. Mitchell, *Machine learning: Trends, perspectives, and prospects*, 349 SCIENCE 255, 255 (2015).

[96] LeCun et al., *supra* note 33, at 436, 442 (discussing the relatively higher frequency of supervised as opposed to unsupervised instruments).

output y for any given input x. It classifies x in terms of y.[97] Its outputs hence take the form of a sorting of x onto categories of y[98]: for example, images into the classes of "face" and "not face"; suspects into the classes of "dangerous" and "not dangerous"; or shoppers into the classes of "impulse purchasers" or "not impulse purchasers." These classifications are correlational and not causal in nature. An algorithm's performance is measured in terms of how well it captures the strength of the relation of x to y, not by its ability to discern an actual causal relationship of x and y.[99] Its success in that regard is a function of the extent to which it can "optimize a performance criterion using example data or past experience."[100] At bottom, this task is analogous to the function performed by familiar tools such as ordinary least squares and logistic regression analysis.[101] Machine learning tools, however, typically outperform those tools by an order of magnitude in terms of predictive accuracy; they also tend to generate results with lower bias and lower variance than ordinary regression.[102] They do so because their algorithms dynamically update the models used to map relationships within the data as new examples are introduced. They thus "learn[] rules from data" about how to better perform their task in the course of executing it.[103]

An unsupervised machine-learning algorithm uses the same computational tools to a slightly different end. It begins with unlabeled training data, and then develops classifications based on the data's immanent structure rather than any ex ante guidance by the programmer.[104] Provided a set of online images, for instance, an unsupervised algorithm might sort them into any number of categories, none of which have been specified a priori: cats v. not cats; people v. objects, etc. These categories

---

[97] *Id.* This process can also be described in terms of a "classifier" rather than a function, that examines inputs with "feature values" and outputs a class variable. Pedro Domingos, *A Few Useful Things to Know About Machine Learning*, COMM. ACM, Oct. 2012, at 78-80 ("A classifier is a system that inputs (typically) a vector of discrete and/or continuous feature values and outputs a single discrete value, the class.").

[98] The values of y must be identified ex ante by the programmer. Comm. on the Analysis of Massive Data et al., *Frontiers in Massive Data Analysis* 104 (2013), http://www.nap.edu/catalog.php?record_id=18374 (noting that in supervised learning, the analyst must actively specify a variable of interest); PETER FLACH, MACHINE LEARNING: THE ART AND SCIENCE OF ALGORITHMS THAT MAKE SENSE OF DATA 14 (2012) (noting that "multi-class classification" is "a machine learning task in its own right").

[99] Jordan & Mitchell, *supra* note 95, at 255-57 (noting that performance can be defined in terms of accuracy, with false positive and false negative rates being assigned a variety of weights).

[100] ALPAYDIN, *supra* note 3, at 3.

[101] IAN GOODFELLOW, YOSHUA BENGIO & AARON COURVILLE, DEEP LEARNING 34-35 (2016); *see also* Coglianese & Lehr, *supra* note 4, at 1158 (developing the parallel). In one respect, the parallel may be inexact. One of the pioneers of machine learning, Leo Breiman hence contrasts the "data modeling" approach, which starts from the assumption that a stochastic data model describes the data to hand, and then proceeds to estimate its parameters, and an algorithmic modeling approach, which makes no assumption about the structure of the data, and then looks for a function that fits the data. Leo Breiman, *Statistical modeling: The two cultures (with comments and a rejoinder by the author)*, 16 *Stat. Sci.* 199, 199 (2001).

[102] Jon Kleinberg et al., *Prediction policy problems*, 105 AM. ECON. REV. 491, 493-94 (2015). For a discussion of technical tools by which machine learning instruments achieve this advance beyond ordinary regression, see Esteban Alfaro, Matias Gamez, and Noelia Garcia, *Adabag: An R package for classification with boosting and bagging*, 54 J. STAT. SOFTWARE 1, 1-2 (2013) (describing the operation of 'boosting' and 'bagging' techniques).

[103] Ziad Obermeyer & Ezekiel J. Emanuel, *Predicting the future—big data, machine learning, and clinical medicine*, 13 NEW ENGLAND J. MED. 1216, 1217 (2016); PEDRO DOMINGUES, THE MASTER ALGORITHM: HOW THE QUEST FOR THE ULTIMATE LEARNING MACHINE WILL REMAKE OUR WORLD 6-7, 23 (2015).

[104] FLACH, *supra* note 3, at 14-15.

can be imagined as clusters of instances in the data that are more similar to each other than to other instances. The algorithm identifies these clusters by constructing multiple layers of representation, with each layer having a different level of abstraction.[105] By iteratively updating the boundaries of different clusters identified within the data at a given layer of representation, the algorithm increases within-cluster similarity and between-cluster divergence.[106] The aim of unsupervised machine learning, in other words, is "to see what generally happens and what does not."[107]

Supervised and unsupervised machine-learning tools generally rest on distinct computational architectures.[108] Among the prominent forms of unsupervised learning methods are associational learning, cluster analysis, principal component analysis, and multidimensional scaling.[109] Supervised learning relies upon a number of different computational strategies, and two common ones are worth mentioning. First, some supervised learning employs what are called 'neural networks.'[110] The latter employ a series of layers of 'neurons,' or nodes. Inputs are received by the first layer of neurons, which apply a function that transforms those inputs. The resulting outputs are then transmitted to other layers in the network, where they are subject to different transformations, until they reach an output layer.[111] Relations between the neurons are recalibrated constantly by a learning algorithm called a Hebbian learning rule, which reinforces connections between neurons that are activated at the same time.[112] A second approach is the "random forests" school of algorithms, which generate predictions by producing thousands of decision trees mapping the data.[113] Each "tree" is trained on a random sample of the training data, and the model returns a prediction that is the majority prediction of the

---

[105] Geoffrey E. Hinton, *Learning Multiple Layers of Representation,* 11 TRENDS COGNITIVE SCI. 428 (2007).

[106] JOHN D. KELLEHER & BRENDAN TIERNEY, DATA SCIENCE 101-02 (2018). The algorithm can be understood as maximizing these parameters.

[107] ALPAYDIN, supra note 3, at 11.

[108] Some tools are used for both. ALPAYDIN, *supra* note 3, at 114 (noting the use of support vector machines for both structured and unstructured learning). A support vector machine (SVM) is a way of identifying relationships among variables that would not be apparent from human inspection of the graphical representation of such data. See Isabelle Guyon, *Data Mining History: The Invention of Support Vector Machines*, KDNuggets (July 2016), http://www.kdnuggets.com/2016/07/guyon-data-mining-history-svm-support-vector-machines.html [https://perma.cc/N459-CRUY] (describing the history of the SVM by one of the scientists that modified the algorithm in the 1990s).

[109] Trevor Hastie, Robert Tibshirani & Jerome Friedman, *Unsupervised Learning*, *in* THE ELEMENTS OF STATISTICAL LEARNING 485 (2009). In clustering, for example, "we generate a tree structure with clusters at different levels of granularity and clusters higher up in the tree are subdivided into smaller clusters" to "find structure in the data" that was previously unknown. ALPAYDIN, *supra* note 3, at 117.

[110] The following draws on the lucid account in ALPAYDIN, *supra* note 3, at 88-103, and Bengio, *supra* note 33, at 48-49.

[111] For a useful graphical representation, see Bengio, *supra* note 33, at 49.

[112] KELLEHER & TIERNEY, *supra* note 106, at 127 ("Training a neural network involves finding the correct weights for the connections in the network."). The standard means to train a neutral network is with an algorithm called a backpropagation algorithm. This works by assigning random weights to connections in the network, and then updating those weights each time a training instance is encountered. It is so named because the algorithm passes (or backpropagates) errors from the output layer to the input layer. *Id.* at 129-30; James Somers, *Is AI Riding a One –Trick Pony?*, 120 MIT TECH. REV. 29, 31 (2017) (offering a nontechnical account of backpropagation that emphasizes its centrality to machine learning). For the seminal technical account, see David E. Rumelhart, Geoffrey E. Hinton & Ronald J. Williams, *Learning representations by back-propagating errors*, 323 NATURE 533 (1986).

[113] Leo Breiman, *Random forests*, 45 MACHINE LEARNING 5, 5 (2001).

trees in the forest.[114] Random forests, and the wider category of decision-tree models in which they fall, are particularly useful for nominal or ordinal data; in contrast, neural networks work well with numerical data.[115]

A more recent development is the technology of deep learning. In deep learning, an algorithm is constructed with multiple levels of representation, each of an increasing degree of complexity. The "key aspect" of deep learning is that "layers of features are not designed by human engineers: they are learned from data using a general-purpose learning procedure."[116] Deep-learning instruments are especially apt for unsupervised tasks, with no specification of features. They require little "manual interference," such that designers "just wait and let the learning algorithm discover all that is necessary by itself."[117]

The ends to which machine learning tools can be put are constrained.[118] Most are used for the (non-causal) prediction of outcomes in the case of supervisory data, or the identification of clusters or associations (in the case of recommendation systems such as those employed by Netflix and Amazon). Kelleher and Tierney, indeed, usefully reduce the "real-world" problems amenable to machine learning to four broad categories: the identification of clusters, or association within a population; the identification of outliers within a population; the development of associational rules; and prediction problems of classification and regression.[119] In contrast, machine learning tools are of more limited uses when a problem also requires "estimating the casual effect of an intervention."[120] An example is a recent study of the allocation of hip replacement surgery among otherwise eligible patients.[121] The study used machine-learning tools to identify which patients would live long enough to benefit from the surgery. But those tools could not estimate the causal effect of the surgery on patient welfare so as to facilitate a prioritization among those likely to survive long enough to benefit from the surgery.[122] More generally, in the use of machine learning, "causal inference is only possible when the analyst makes assumptions beyond those required for prediction methods."[123]

---

114 KELLEHER & TIERNEY, *supra* note 106, at 141-42.

115 *Id.* at 136.

116 LeCun et al., *supra* note 33, at 436; *id.* at 438 ("A deep learning architecture is a multilayer stack of simple modules, all (or most) of which are subject to learning, and many of which compute non-linear input-output mappings."); KELLEHER & TIERNEY, *supra* note 106, at 131-32 ("Deep learning networks are simply neural networks that have multiple layers of hidden units." (emphasis and footnote omitted)).

117 ALPAYDIN, *supra* note 3, at 107-08.

118 Judea Pearl, *Theoretical impediments to machine learning with seven sparks from the causal revolution* PREPRINT ARXIV:1801.04016, AT 1-2 (2018) (arguing that inability of machine learning to analyze counterfactuals to infer causation is a major impediment).

119 KELLEHER & TIERNEY, *supra* note 106, at 151-80 (providing examples of these different tasks). Another typology of uses identifies four uses of machine learning: prioritization, classification, association, and filtering. Bruno Lepri et al., *The tyranny of data? The bright and dark sides of data-driven decision-making for social good*, *in* TRANSPARENT DATA MINING FOR BIG AND SMALL DATA 3, 4 (2017).

120 Athey, *supra* note 3, at 483.

121 Jon Kleinberg et al., *supra* note 102.

122 *Id.* at 493.

123 Athey, *supra* note 3, at 484-85 (also flagging limitations due to "incentives and manipulability").

B. Distinguishing Machine from Human Decisions

Machine learning is obviously dissimilar from human decision-making. But how? And why might that difference matter normatively? I consider here three perspectives on those differences. First, probably the most obvious discontinuities are in relation to the scale, capacity, and the underlying mechanisms of machine learning. Second, it is also sometimes said that machine decisions are more opaque than human ones. I accept the efficiency gap between machines and human, but query its significance for an inquiry into rights. I doubt, though, whether a transparency gap exists. Rather, even when the opacity of an algorithm is a function of its complexity and operational unpredictability, it is not clear that complaints about the *greater* impenetrability of machines than humans are well founded, at least when framed as a generalization rather than case-to-case comparisons.[124] Finally, I consider the necessary degree of entanglement between human and machine action.

1. How Machine and Human Decisions Diverge in Operation

In three ways, machine and human decisions diverge in basic operation: in terms of the basic architecture of reasoning; in the propensity to err; and in respect to the sheer capacity to complete identification and prediction tasks. Both the second and the third difference have normative salience—but not necessarily in ways that bear on a right to a human decision.

At a very elementary level, the architecture of machine learning diverges from that of human decision-making. It is tempting to think that some forms of computational architecture, in particular, neural networks, track in some fashion the human brain's cognitive process.[125] But, except at a very high level of generality, this analogy should be resisted. It is true that the science of human cognition "influenced the emergence of artificial neural networks."[126] Despite the verbal resonance, "a neural network is inspired by the brain in the same way that the Olympic stadium in Beijing is inspired by a bird's nest," and is in practice more akin to regression analysis than to the operation of human neurons.[127] Among the first neural networks was the "perceptron," comprising a single 'neuron' or node, that executed a single non-linear function.[128] But it obviously bears little resemblance to the human brain. And today, whereas the study of human cognition has aimed at taxonomizing "cell types, molecules, cellular states, and mechanisms for computation and information storage," the machine learning field "has largely focused on instantiations of a single principle: function optimization."[129] It

---

[124] For an example of this species of complaint, see W. Nicholson Price II, *Black-Box Medicine*, 28 HARV. J.L. & TECH. 419, 421 (2015) (identifying a "type of medicine [that] is "black-box" to everyone by nature of its development… not … because its workings are deliberately hidden from view").

[125] Davide Castelvecchi, *Can We Open the Black Box of AI?*, NATURE, Oct. 5, 2016, at 19 (asserting that neural networks are modeled on the brain).

[126] Bengio, *supra* note 33, at 49.

[127] Jayesh Bapu Ahire, *Artificial Neural Network: Some Misconceptions*, THE MEDIUM, Jan. 27, 2018, https://medium.com/swlh/artificial-neural-network-some-misconceptions-cb93e80b34bb; Adam H. Marblestone, Greg Wayne & Konrad P. Kording, *Toward an integration of deep learning and neuroscience*, 10 FRONTIERS IN COMPUTATIONAL NEUROSCIENCE 94, 94 (2016) (noting the importance of mathematical advances, and not neuroscience breakthroughs, in machine learning).

[128] The seminal paper is Frank Rosenblatt, *The perceptron: a probabilistic model for information storage and organization in the brain*, 65 PSYCH. REV. 386 (1958).

[129] Marblestone et al., *supra* note 127, at 94.

is thus a mistake to think that machine learning is simply a silicon-based version of a familiar carbon-based process, even if much of the machine-learning agenda is in some sense calibrated in terms of familiar human capabilities.[130]

A second margin along which human and machine decisions vary is in terms of quality. Even in putatively high-quality settings, human decision-making is plagued by the distortive effects of heuristics,[131] implicit bias,[132] and sheer noise.[133] Machine learning, to be sure, remains vulnerable to polluted training data and requires ineluctably normative choices in crafting classification rules.[134] For example, text analysis of large corpora, such as can be scraped from the world wide web, are influenced by biases (race- and gender-related, for instance) that influenced the corpora's creation.[135] Incautiously deployed, such tools can exacerbate morally suspect social stratification. Yet at the same time, they can also generate accurate predictions that in practice no human can match.[136] If there is a limit to the predictability of human behavior, it has not yet been identified.[137] Even when there is a human substitute for a decision, studies in a variety of fields suggest that large gains in human wellbeing can be attained by using a machine learning tool rather than a person.[138] For instance, early deployment of driverless cars will almost certainly reduce dramatically the social toll of traffic accidents.[139] As I shall argue below,[140] just as it would be a mistake to ignore the risks involved in the design of machine

---

[130] Consider for instance the problem of reinforcement learning, in which humans are confronted with a difficult task, and must derive efficient representations of the environment from high-dimensional sensory inputs to solve new challenges. This is a task that machine learning is just starting to address. Volodymyr Mnih et al., *Human-Level Control Through Deep Reinforcement Learning*, 518 NATURE 529, 529 (2015).

[131] *See, e.g.,* Jeffrey J. Rachlinski, *A Positive Psychological Theory of Judging in Hindsight*, 65 U. CHI. L. REV. 570, 571 (1998) (finding anchoring bias).

[132] Jeffrey J. Rachlinski et al., *Does Unconscious Racial Bias Affect Trial Judges?* 84 NOTRE DAME L. REV. 1195, 1195 (2009) (documenting implicit racial bias in the decisions of trial judges).

[133] Daniel L. Chen & Jess Eagel, *Can Machine Learning Help Predict the Outcome of Asylum Adjudications?*, PROCEEDINGS OF THE 16TH EDITION OF THE INTERNATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE AND LAW 237, 237 (2017) (finding that case-relevant factors explain only about one-third of the outcomes in asylum decisions).

[134] *See* Huq, *supra* note 45, at – (discussing flawed training data and moral choices created by background racial inequalities).

[135] Aylin Caliskan, Joanna J. Bryson, and Arvind Narayanan, *Semantics derived automatically from language corpora contain human-like biases*, 356 SCIENCE 183, 183 (2017).

[136] For example, recent applications include better programming of traffic signals to minimize aggregate traffic. Rose Yu et al., *Deep learning: A generic approach for extreme condition traffic forecasting*, Proceedings of the 2017 SIAM International Conference on Data Mining 777-785 (2017).

[137] Jake M. Hofman et al., *Prediction and explanation in social systems*, 355 SCIENCE 486, 487-88 (2017) (describing these limits as an "open" question).

[138] Kleinberg et al., *supra* note 102, at 494 (medical decisions); Jon Kleinberg et al., *Human decisions and machine predictions*, 133 Q. J. ECON. 237 (2017) (bail). In my view, the bail example is ambiguous because welfare gains depend on how other actors in the criminal justice system respond to the prospect of more targeted instruments of pretrial coercion. It is not at all clear their response will be to use coercion less often or more wisely.

[139] For an argument in favor of accelerated implementation of driverless cars on social welfare grounds, see Nidhi Kalra and David G. Groves, *The Enemy of Good: Estimating the Cost of Waiting for Nearly Perfect Automated Vehicles* (2017) https://www.rand.org/pubs/research_reports/RR2150.html (recommending rapid introduction of driverless vehicles on safety grounds).

[140] *See infra* Part III.C.

learning tools, then, so too it would be error to ignore the range of social goods that can be generated by their employ.

A third and related difference between machine and human decision-making is sheer capacity. An algorithmic instrument can "sift through vast numbers of variables, looking for combinations that reliably predict outcomes" to generate "enormous numbers of predictors—sometimes, remarkably, more predictors than observations—and combining them in nonlinear and highly interactive ways."[141] There are hence instances in which machine learning can detect patterns and offer predictions that would necessarily escape human cognition.[142] Note, though, that an absolute difference in computational capacity, however, does not distinguish machine-learning cleanly from other transformational technologies. No human can run as fast as a readily available passenger car can drive. No human can execute mathematical operations as quickly as a common pocket calculator can. And no human can see as far as a modern telescope, or as deep into materials as a modern microscope. Changes in the scale of capability, that is, are endemic to technological evolution. Moreover, just like other transformative technologies, machine-learning tools are likely to remain outperformed by humans in respect to some tasks. Indeed, it is striking that the most sophisticated forms of deep learning are used today to execute functions such as speech or handwriting recognition—functions that most children manage with ease.[143] Most importantly, as I have already emphasized, machine-learning algorithms are well-designed for prediction problems, but not honed for the sort of causal inference problems with which econometrics has traditionally been concerned.[144]

At least in theory, therefore, machine learning is capable of better (in the sense of more accurate) predictions than humans. This capacity is not always realized. But legal design, including the design of rights, is best thought of dynamic rather than static. That is, it should be aimed at eliciting improvements in state action. From a dynamic perspective, the space for improvement in machine decisions provides a reason to think that a right to a human decision risks stymying beneficial institutional changes.

2. *The Opacity of Other (Human and Machine) Minds*

It is commonly asserted that algorithmic decisions derived from machine-learning instruments are more opaque, and hence more resistant to explanation, than human decisions. Machine learning is said to involve processes "which [are] not explainable in human language."[145] It rests instead on "the

---

141 Obermeyer & Emanuel, *supra* note 103, at 1216-17.

142 Coglianese & Lehr, *supra* note 4, at 1159 ("By eschewing [a] dependency on existing knowledge and the need to identify the functional form of any relationships, machine learning can apply to a wider range of problems and yield vastly enhanced accuracy over its alternatives, whether human intuition, expert judgment, or traditional statistical techniques").

143 *See, e.g.*, Cheng-Lin Liu et al., *Handwritten Digit Recognition: Benchmarking of State-of-the-Art Techniques*, 36 PATTERN RECOGNITION 2271 (2003).

144 Sendhil Mullainathan & Jann Spiess, *Machine learning: an applied econometric approach*, 31 J. ECON. PERSP. 87, 87-88 (2017).

145 Tal Z. Zarsky, *Transparent Predictions*, 2013 U. ILL. L. REV. 1503, 1568 (2013) (acknowledging "the important strengths of transparency," but also flagging its limits in reference to predictive tools); *accord* Knight, *supra* note 9 (advocating transparency); *see also* Alyssa Carlson, *The Need for Transparency in the Age of Predictive Sentencing Algorithms*, 103 IOWA L. REV. 303, 344-46 (2017) (advocating that freedom-of-information laws extend to nonstate providers of

high dimensionality of data, complex code, and changeable decision-making logic."[146] This concern with transparency seems to motivate in part the demand for a right to a human decision.[147] The empirical predicates of this claim hence warrant separate treatment.

To begin with, we should rule out a common complaint about algorithmic transparency. This hinges on the unwillingness of the corporate entities that own the machine-learning instrument to disclose their details for fear of economic harm.[148] Such secrecy does not plainly distinguish machine from human decisions. There is not much ultimate difference between the use of trade secrets law, or other forms of intellectual property protection for algorithms, and the use of contractual clauses such as do-not-compete clauses, to prevent the diffusion of technical information.[149] There is nothing distinctive, that is, in the fact that it is a machine rather than a person who is being shielded by law from examination.

More plausibly, the idea of distinctive machine opacity hinges on the complex, recursive, and unprogrammed way in which computational algorithms operate. Machine learning is typically applied to "problems for which encoding an explicit logic of decision-making functions very poorly."[150] The classification rule identified by a machine-learning tool can be a dynamic function of (for example) a neural network rather the result of one sequence of calculations.[151] But replicating or understanding the network's emergent properties strains human imagination. Mere examination of "complicated or obfuscated" source code reveals little about how the program operates in the real world.[152] From an ex post perspective, therefore, there is a sense that algorithms may not be transparent because it is impossible to reconstruct the grounds upon which a given decision was reached.

But this kind of preclusive ex post complexity need not be taken as a technological given. Computer scientists have suggested that it is possible to guarantee ex ante "a tamper-evident record that provides nonrepudiable evidence of all nodes' actions."[153] From this record, "faulty" nodes in a

---

algorithmic tools); Danielle Keats Citron & Frank Pasquale, *The Scored Society: Due Process for Automated Predictions*, 89 WASH. L. REV. 1, 25 (2014) (arguing for oversight and transparency).

146 Brent Daniel Mittelstadt et al., *The ethics of algorithms: Mapping the debate* 3 BIG DATA & SOC. 1, 3 (2016).

147 *See supra* text accompanying notes 9 to 13.

148 For examples of this concern, see Rebecca Wexler*, Life, Liberty, and Trade Secrets: Intellectual Property in the Criminal Justice System*, 70 STAN. L. REV. 1343 , 1350 (2018) (documenting "the introduction of trade secret evidence into criminal cases"); FRANK PASQUALE, THE BLACK BOX SOCIETY 12-15 (2013).

149 *See, e.g.,* Orly Lobel, *Enforceability Tbd: From Status to Contract in Intellectual Property Law*, 96 B.U. L. Rev. 869, 871 (2016) (arguing that employment "contracts serve firms as means to enclose information beyond traditional intellectual property boundaries without adequate notice or debate").

150 Jenna Burrell, *How the machine 'thinks': Understanding opacity in machine learning algorithms*, 3 J. BIG DATA & SOC. 1, 6 (2016).

151 *Id.* at 8.

152 Kroll et al., *supra* note 2, at 647; Joshua A. Kroll, *The fallacy of inscrutability*, PHIL TRANS. SOC. A376 (2018) (describing such disclosure as "neither necessary nor sufficient" for improving understanding).

153 Andreas Haeberlen, Petr Kuznetsov & Peter Druschel, *PeerReview: Practical Accountability for Distributed Systems*, 41 ACM SIGOPS Operating Sys. Rev. 175, 175 (2007); Kroll et al., *supra* note 2, at 662-65 (same); Deven R. Desai & Joshua A. Kroll, *Trust but Verify: A Guide to Algorithms and the Law*, 31 Harv. J.L. & Tech. 1, 10-11(2017) (same).

machine-learning system can be detected.[154] In effect, the consequences of pivotal elements of an algorithm's architecture can be isolated and analyzed. Alternatively, it may be possible to offer "multiple diverse counterfactual[s]" to an algorithm, testing thereby the effect of incremental changes to input outcomes even after the fact.[155] This has a human analog of sorts in experimental tests of the trolley problem.[156] Finally, some kinds of algorithmic design may be more amenable to interpretation than others.[157] Models that are simply to learn tend to be simpler to represent.[158] A designer thus can, within limits, select for a computational architecture's amenability to explanation. Alternatively, it is possible to "approximate the model in simpler form" even after it has been created and applied in the wild.[159] The absence of a capacity to generate an explanation of the requisite sort that enables relevant evaluation of a machine learning tool's effects on the world is "merely a design choice, not an inevitability of the complexity of large system."[160]

Given the availability of mechanisms for investigating machine learning decisions—some of which parallel methods for understanding human decision-making—it cannot be said a priori that the former are any more opaque than humans.[161] True, specialized tools are necessary for interrogating algorithmic results. But the elaborate evidentiary rules that courts have developed for evaluating human testimony suggests that experts are just as needful to the task of understanding human testimony. The simple fact that the diagnostic tools for understanding machine decisions are more alien than those for human decisions does not make them either more or less penetrating.

If there are not systemic reasons to think that machine decisions are always more opaque than human decisions, consider the possibility of a rough equality between human and machine transparency, at least across the mine run of cases. To begin with, notice that other minds are just as much black boxes as machine-learning instruments. There simply is "no generally accepted theory of

---

154 Haeberlen et al., *supra* note 153, at 175.

155 Sandra Wachter, Brent Mittelstadt & Chris Russell, *Counterfactual explanations without opening the black box: Automated decisions and the GDPR* 13-14 (2017), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3063289. ("Finding a closest possible world to x such that the classification changes is, under the right choice of distance function, the same as finding the smallest change to x"). Studies of deep-learning visual recognition tools reveal that even small perturbations in inputs can generate categorical classification changes. *See* S. Huang, N. Papernot, I. Goodfellow, Y. Duan, & P. Abbeel, *Adversarial attacks on neural network policies*, arXiv preprint arXiv:1702.02284 (2017). These are both "subject-centered" rather than "model centered" approaches. Lilian Edwards & Michael Veale, *Slave to the Algorithm? Why A "Right to an Explanation' Is Probably Not the Remedy You Are Looking for*, DUKE L. & TECH. REV., December 4 2017, at 18, 56.

156 *See, e.g.*, Alessandro Lanteri, Chiara Chelini & Salvatore Rizzello, *An experimental investigation of emotions and reasoning in the trolley problem*, 83 J. BUS. ETHICS 789, 789-90 (2008).

157 Desai & Kroll, *supra* note 153, at 11 n.61 (suggesting that decision tree, naïve Bayes, and rule learners are easier to interpret than neural networks or support vector machines).

158 Michael Gleicher, *A Framework for Considering Comprehensibility Modeling*, 4 BIG DATA 72, 82 (2016).

159 Andrew D. Selbst & Solon Barocas, *The Intuitive Appeal of Explainable Machines*, 87 FORDHAM L. REV. 1085, 1110-12 (2018).

160 Kroll, *supra* note 152, at 3. For a useful discussion of both why different kinds of explanation differ, and how to craft effective responses, see Menaka Narayanan et al., *How do Humans Understand Explanations from Machine Learning Systems? An Evaluation of the Human-Interpretability of Explanation*, arXiv preprint arXiv:1802.00682 (2018).

161 Anthony J. Casey & Anthony Niblett, *A Framework for the New Personalization of Law*, 86 U. CHI. L. REV. 333, 355-56 (2019).

how cognitive phenomena arise from computations in cortex,"[162] or of whether consciousness serves any "significant function" for an organism.[163] Nor is there even a generally held folk theory that fills the gap. As one former biophysics professor observed, "we make decisions in areas that we don't fully understand every day," and "can't explain the complex underlying basis for how we arrived at a particular conclusion."[164] The bigger the decision, moreover, the less amenable it can be to reasoned resolution.[165]

The problem of human decisions' opacity is acute when it comes to understanding other minds. We do not hold in our minds clear and distinct (let alone accurate) explanations of how other people think.. Nor do we have direct access to their cognitive processes. Nevertheless, we are able to interact with them successfully. We are even able to impute meaningfully to them beliefs and other mental states in ways that are not obviously fallacious. Further, another person's beliefs and mental states can be interrogated after the fact (as can our beliefs about their beliefs and mental states). Indeed, in many instances, humans "generate and store the information needed to explain [a] decision" and can be asked to produce that information after the fact.[166] Problems of sincerity and candor abound. But these are not treated as insoluble by the law. Transparency when it comes to other minds, in other words, may not always be possible as a theoretical or practical matter. But it may also not be always needful.

Social action based on the understanding that other people have beliefs and mental states requires a skill known as mentalizing. It is a capability that precedes sophisticated cognition. Most people (at least over the age of four who are not law professors) navigate the social world on the working assumption that other people have minds.[167] Some reasonable proportion of the time, our beliefs about other minds are even close enough to the mark to permit effectual social interactions.[168] Further, the absence of generally accessible or widely understood explanations of the biochemical processes through which cognition occurs does not appear to undermine the ability to mentalize or to make judgments about others' beliefs and mental states. The absence of effectual transparency

---

162 Leslie G. Valiant, *What must a global theory of cortex explain?*, 25 CURRENT OP. IN NEUROBIOLOGY 15, 15 (2014).

163 David M. Rosenthal, *Consciousness and its function*, 46 NEUROPSYCHOLOGIA 829, 831, 839 (2008) (arguing that it does facilitate "rational thinking intentional actoon, executive function, or complex reasoning").

164 Vijay Pande, *Artificial Intelligence's 'Black Box' Problem is Nothing to Fear*, N.Y. TIMES, Jan. 25, 2018; *see also* Saul Levmore & Frank Fagan, *The Impact of Artificial Intelligence on Rules and Standards* 7 (2018) (draft on file with author) ("The best judges, like the best athletes and teachers, are often unable to identify the reasons for their success.").

165 Edna Ullmann-Margalit, *Big Decisions: Opting, Converting, Drifting*, 58 ROYAL INST. PHIL. SUPP. 157, 158-59 (2006) (describing a category of "big" decisions that cannot be resolved by standard cost-benefit analysis).

166 Finale Doshi-Velez et al., *Accountability of AI under the law: The role of explanation*, arXiv preprint arXiv:1711.01134 at 9 (2017).

167 Rebecca Saxe, Susan Carey & Nancy Kanwisher, *Understanding other minds: linking developmental psychology and functional neuroimaging*, 55 ANN. REV. PSYCHOL. 87, 94-95 (2004). There appear to be particular regions of the brain responsible for this capacity. *Id.* at 99-100.

168 Chris D. Frith & Uta Frith, *Interacting minds--a biological basis*, 286 SCIENCE 1692, 1692 (1999) (characterizing "the capacity to understand and manipulate the mental states of other people and thereby to alter their behavior" as an important component of "social intelligence").

when it comes to other people's mental processes, in short, is a problem for philosophers.[169] It is not necessarily a problem for the rest of us.[170]

So it is far from clear that transparency is systematically more inaccessible for machine rather than human interlocutors.[171] In both domains, ex ante regulation can elicit better rather than worse contemporaneous records ex post. Technical skills are required to interpret evidentiary records in both domains. And with machines and humans alike, there are some reasons for thinking that transparency falls short at least in some class of cases.[172] A difference of a predictable and stable sort between the two domains, as a general matter, is hard to discern.[173] A healthy dose of skepticism about the putative opacity gap between machines and humans takes off the table on empirical grounds one potential justification for the right to a human decision. If human and machine decisions are similarly opaque, albeit in different ways, that is, a right to the former cannot be explained in terms of mere legibility. The arguments explored in Part III, therefore, must be carefully framed to avoid any assumption about necessary differentials in opacity.

C. The Entanglement of Human and Machine Decisions

There is yet one other perplexing preliminary of an empirical rather than a legal cast: to what extent it is even empirically plausible to imagine that there can be a 'machine decision' that is acoustically separate from a human decision? Article 22 of the GDPR envisages the possibility of decisions "based *solely* on automated processing."[174] But it there really such a beast?

Perhaps not—for three reasons. First. all machine learning tools are at their origin the fruit of specific human design and engineering choices. There is no such thing as a wholly endogenous algorithm.[175] And the design of a machine-learning tool is not a mechanical task. It is freighted with

---

[169] The classic statement of the problem is Norman Malcolm, *Knowledge of Other Minds*, 55 J. PHIL. 969, 969-70 (1958) (invoking Mill for the question how one can know that there are indeed other minds). There is an equally old tradition that such thoroughgoing skepticism is self-refuting. *See* ANITA AVRAMIDES, OTHER MINDS 5-6 (2001) (tracing this skepticism to the common sense philosopher Thomas Reid).

[170] Recall here Samuel Johnson's famous refutation of Bishop Berkeley.

[171] In other words, knowing that one is dealing with a machine or a human may not tell you much about how transparent a decision is going to be. Doshi-Velez et al., *supra* note 166, at 9 ("[T]here may be situations in which it is possible to demand more from humans, and other situations in which it might be possible to hold AI systems to a higher standard of explanation.").

[172] Moreover, some have argued that transparency is a flawed solution because it "may occlude the true problems which rest in societal power relations and institutions as much as the software tools employed." Edwards & Veale, *supra* note 155, at 68; *accord* Mike Ananny & Kate Crawford, *Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability*, 20 NEW MEDIA & SOCIETY 973 (2018) (offering a critique of transparency as a "neoliberal" solution).

[173] Paradoxically, there may be more ways of checking algorithmic decisions than human decisions. Consider, for example, the robustness of encryption algorithms, which rest not on 'explanation' but on formal mathematical proofs. Doshi-Velez et al., *supra* note 166, at 11. It is hard to think of a parallel with humans.

[174] GDPR § 22(1) (emphasis added).

[175] In late 2017, Google announced its AutoML project, which aimed to create "a machine-learning algorithm that learns to build other machine-learning algorithms." Cade Metz, *Building A.I. That Can Build A.I.*, N.Y. TIMES, Nov. 5, 2017, at B1. Similar research is ongoing at Carnegie Mellon University. *See* Renato Negrinho & Geoff Gordon,

normative choices. A designer must select a certain kind of algorithmic architecture—a neural network, a decision-tree-based model such as random forests, or something else. The choice is a difficult one and necessarily evaluative in character. All forms of machine learning, moreover, have distinctive learning biases—that is particular functions that they are more likely to employ during analysis. Finding the "best match" between an algorithm's learning bias and a data set impels an exercise of human judgment.[176]

The call for human judgment does not end there. Consider the process of a deep learning instrument's start-up. In this form of machine learning, a multilayered neural network must be created. The instrument's designer must determine how many layers to build in, and how many neutrons to include in each layer.[177] The designer must then decide how to connect the network's different elements. She might chose to create a recurrent neural network, in which the network's topology is looped. Each neuron processes inputs in the context of the previous inputs processed, creating a sort of "memory."[178] Alternatively, she might craft a convolutional neural network, in which localized groups of neurons are trained to recognize particular patterns regardless of where they appear in the data (for example, an eye or a nose in a visual recognition system).[179] The choice of network topology, again, is a human decision grounded in the quiddities of human conduct. Nothing intrinsic to the actual algorithm can answer that question.

Second, the human role in machine learning is not limited to the initial design of an algorithm. A designer must also select the data upon which the machine learning instrument is initially trained. This training data, moreover, is generally not produced by an algorithm. It is a function of human action. As a result, it can replicate the biases and blind-spots of the individuals who created it.[180] For example, in the policing context, there is a concern that historical arrest data, if used to drive future deployment decisions, will reflect and reproduce the troubling assumptions about racial proclivities toward crime that have been prevalent among police officers in the past.[181] Having collected data, a designer needs to preprocess and transform that data, adjust the algorithm's parameters in light of the data, and fine-tune the algorithm based on the quality of the results.[182] Saul Levmore and Frank Fagan, arguing for the inevitability of human-machine partnerships, additionally suggest that missing data problems mean that human "thinking about data-selection and goal-setting" will inevitably be needed

---

DeepArchitect: Automatically Designing and Training Deep Architectures, arXiv preprint arXiv:1704.08792 (2017). A human engineer is still needed to guide some of the search processes over different potential architectures. *Id.* at 1.

176 KELLEHER & TIERNEY, *supra* note 106, at 99-100; *see also* Casey & Niblett, *supra* note 161, at 354 ("[H]umans are involved in all stages of setting up, training, coding, and assessing the merits of the algorithm.").

177 Bengio, *supra* note 33, at 50.

178 KELLEHER & TIERNEY, *supra* note 106, at 133.

179 *Id.*; *see also* LeCun et al., *supra* note 33, at 439 (discussing convolutional networks in general terms).

180 On biases, see Kroll et al., *supra* note 2, at 680 ("[A]lgorithms that include some type of machine learning can lead to discriminatory results if the algorithms are trained on historical examples that reflect past prejudice or implicit bias"). On blind-spots, see Kate Crawford, *Artificial Intelligence's White Guy Problem*, N.Y. TIMES, June 25, 2016, https://www.nytimes.com/2016/06/26/opinion/sunday/artificial-intelligences-white-guy-problem.html (noting problems that can arise from missing data).

181 Huq, *supra* note 45, at --.

182 Saleema Amershi et al., *Power to the people: The role of humans in interactive machine learning*, 35 AI MAGAZINE 105, 105 (2014).

even after a system is up-and-running.[183] As they intimate, the implementation as well as the design process for machine learning is punctuated by "lengthy and asynchronous iterations" of human-machine interaction that ensure that machine-learning is always critically molded by a human hand.[184]

Third, once up and running, machine learning tools still need "human caretakers," tasked with everything from moderating the results of deep-learning algorithms used to simulate vision to maintenance workers who clean and repair the data centers used to house large pools of information necessary for algorithms' operation.[185] The process of learning within neural networks also "require[s] close involvement by a human," who must craft labels for training data, and then generate hypotheses that guide the process of optimization.[186]

The idea of a machine that will run of its own accord, in short, appears far from accurate. To be sure, this could change. Some deep learning instruments already require "very little learning by hand" at the third stage (although their design still presents considerable challenges).[187] But at least for the time being, a machine learning instrument rests on a foundation of human design and performance-specification decisions. The possibility of a purely machine decision, again for now,[188] lies beyond the technological event horizon.

That empirical fact raises a question about how precisely the notion of a decision "based solely on automated processing"[189] should be construed. This is so even if no human designer could either replicate the cognitive labor of the algorithm or alternatively predict every classification choice the algorithm would ultimately make. A literal understanding of the right to a human decision, it follows, is not plausible. But that need not mean the idea of a solely autonomous machine decisions should be jettisoned. Another possible interpretation of the intuition behind a right to a human decision would focus not on the absolute extent of human involvement, but on the timing and quality of such involvement. The design and training of machine learning that I have stressed here occurs largely ex ante. Human involvement contemporaneous to an algorithm's operation comprises trivial or largely ministerial action. The demand for a human decision might be construed as hinged more narrowly on the slice of time in which a machine acts in respect to a specific case. It is a demand, that is, that in

183 Levmore & Fagan, *supra* note 164, at 8.

184 Amershi et al., *supra* note 182, at 105.

185 AINow, AINOW 2017 REPORT 12 (2017), https://assets.ctfassets.net/8wprhhvnpfc0/1A9c3ZTCZa2KEYM64Wsc2a/8636557c5fb14f2b74b2be64c3ce0c78/_AI_Now_Institute_2017_Report_.pdf.

186 Bengio, *supra* note 33, at 50-51 (flagging the risk that a "learning algorithm can get stuck in what is called a local minimum, in which it is unable to reduce the prediction error of the neural network by adjusting parameters slightly"); *see also* Amershi et al., *supra* note 182, at 106 (describing "iterative exploration of the model space" by practitioners once a learning algorithm is up and running) .

187 LeCun et al., *supra* note 33, at 436.

188 A recent survey finds a one in two chance that high-level machine intelligence will be developed around 2040–2050, rising to a nine in ten chance by 2075. V.C. Müller & N. Bostrom, *Future progress in artificial intelligence: A survey of expert opinion*, in FUNDAMENTAL ISSUES OF ARTIFICIAL INTELLIGENCE 555, 555-56 (2016).

189 GDPR § 22(1).

the immediate "transactional frame"[190] of human/machine interaction, humans should not lie on one temporal side of the interaction alone. Viewed through temporal blinders, this assumes that the fact a human engineer once calibrated the machine with which one is enacting is cold comfort, or no comfort at all.

On this interpretation, the demand for a human decision bears a family resemblance to a "standard," as opposed to a rule, as defined in the law-and-economics literature: a norm that is given content after regulated subjects act rather than beforehand.[191] Just as the rules/standards distinction hinges on the timing (and to some extent the degree of specification[192]) of a norm, so the idea of a right to a human decision introduces an assumption that ultimately the normative grounds for an action can be supplied only by a human rather than a machine actor intervening after the computational processes of an algorithm have come to a close.[193] This suggests that to the extent that the Working Party Guidelines to Art. 22(1) of the GDPR are ambiguous, they should be interpreted as mandating "meaningful" and "ex post" review.[194] This suggests that Article 22(1) should be read to be violated only when there is an absence of ex post, as opposed to ex ante, human involvement. The GDPR also does not explain how the quality of such review is to be assessed.[195]

To summarize, the right to a human decision is not plausibly understood to mean simply the right to a decision with some human element, however timed and however substantive: To define the right that way would rob it of effectual content given present technological constraints. But if it is not meaningful to speak of machine decisions that do not have a human in the loop, there is a question as to why the timing of necessary human involvement makes a practical difference. If the right to a human decision is not to boil down to something like an aesthetic preference, the normative grounds for a right *timed in this fashion* must be supplied. This is the question taken up in Part III.

---

[190] *Cf.* Aziz Z. Huq, *The Consequences of Disparate Policing: Evaluating Stop and Frisk As A Modality of Urban Policing*, 101 MINN. L. REV. 2397, 2402 (2017) (noting that a "narrow, individualist 'transactional frame' … currently dominates both law and economics).

[191] Louis Kaplow, *Rules Versus Standards: An Economic Analysis*, 42 DUKE L.J. 557, 559-63 (1992).

[192] *See, e.g.,* Anthony J. Casey & Anthony Niblett, *The Death of Rules and Standards*, 92 IND. L.J. 1401, 1407 (2017) ("Rules are precise and ex ante in nature…. Standards, on the other hand, are imprecise when they are enacted.").

[193] The draft Universal Guidelines on Artificial Intelligence awkwardly recognizes the relevance of temporality by stating that when it is not "possible or practical to insert a human decision prior to an automated decision," then there should be a human focus on outcomes. The Public Voice, *Draft Universal Guidelines on Artificial Intelligence*, at 1, https://epic.org/international/AIGuidleinesDRAFT20180910.pdf.

[194] Working Party Guidelines, *supra* note 59, at 21-22 (excluding from Article 22(1) instances in which a human "reviews and takes account of other factors in making a final decision"); *see also supra* text accompanying note 195.

[195] It is not clear what "meaningful" supervision entails. *See* Michael Veale & Lilian Edwards, *Clarity, surprises, and further questions in the Article 29 Working Party draft guidance on automated decision-making and profiling*, 34 COMP. L. & SEC. REV. 398, 401(2018) ("How this expanded notion of 'solely' could practically be assessed from the point of view of the data controller or the data subject is one of the significant grey areas th[e] guidance leaves in its wake."). In addition, if an automated process does not change "legal rights" or have an "equivalent or similarly significant" effect, the Working Party suggests that it is not covered by Article 22(1). Working Party Guidelines, supra note 59, at 21-22 (noting that while targeted advertising is not typically covered, the "intrusiveness" of the targeting, an individual's expectations, and the operator's knowledge of the "vulnerabilities" of the person might render it covered by Article 22(1)).

D. Conclusion

Machine learning denotes a large field of heterogenous and evolving computational forms. What is mapped here barely scratches its surface obviously. I have stressed some of the central taxonomical lines sufficiently to allow intelligent discussion of the right to a human decision. I have further explored dimensions along which a machine decision might be deemed categorically distinct from its human counterpart. This discussion should clarify the stakes of claiming human as distinct from machine determinations.

At a very general level, scale (capacity) and mechanisms do vary systematically between humans and machines, but one must be careful about leaping from that conclusion to a claim about right. Machines have the capacity to classify and predict with fewer errors than humans. At least from a dynamic perspective, this suggests that legal rules should incentivize the creation of better machines, rather than their substitution with humans. In contrast, transparency, which is a focal point for much legal scholarship, does not provide a meaningful point of distinction between humans and machines. Finally, I have suggested that most algorithms in operation now (and arguably for the next twenty years) will be shaped and orientated by human action. If a right to a human decision is to have meaningful content now, therefore, it must be understood to require human judgment at a particular moment: after a machine learning instrument in the wild encounters and classifies a human actor. Any other definition is either technologically infeasible or too easily satisfied to allow the putative right to stand as a coherent, independent entity.

III. The Normative Grounds of the Right to a Human Decision

With this technological context in hand, this Part explores the best available normative account of a right to a human decision. I do not start from the presumption that a right to a human decision is warranted. To the contrary, by airing the best normative grounds for recognizing this right, I hope to make a judgment as to whether the right is worth embedding into law, and if so for what reasons. I focus on state applications of machine-learning tools. As I have noted, criminal justice and welfare administration have to date been characterized by the most rapid uptake of such tools in helping determine who is stopped by police, who is imprisoned before (or after) trial, and who receives public benefits.[196] The case of the person who is subject to coercion, or denied a benefit, because of a mistake machine decision is the most compelling instantiation of why a right to a human decision might be warranted. I thus assume that something material is at stake when the machine (or human) decides, an assumption that is not always satisfied in the context of private uses of machine learning.

For the sake of exposition,[197] I delineate three kinds of reasons for a right to a human decision (understood, per Part II, as a right to ex post human review). The first is *subject-focused*, in that it trains on the actions, or potential actions, of the subject of the algorithmic classification rule. This class of reasons builds on the intuition that a nonhuman instrument forecloses certain opportunities for the exercise of meaningful human agency. The second cluster of reasons is *classifier-focused.* These build on the intuition that the state in particular owes to individuals a certain kind of decision, without regard to whether the individual can, or would be likely to, respond or seek to alter that decision. The third group of reasons focuses on *systemic effects*. Whereas the first and the second kinds of reasons narrowly

[196] *See supra* text accompanying note 31.

[197] That is, my categories are instruments of convenience, and not mean to capture any underlying moral facts, if such things exist.

train upon the interaction between the individual and the classifier, this class of reasons dilates the lens, capturing the possibility of spillover consequences that unfold only dynamically and cumulatively. I shall argue that none of these clusters of reasons provide secure normative ground for a right to a human decision. Although systemic effects in particular have some force, they are poorly suited to work as foundations for an individual right.

My approach is not to lean upon the jurisprudence of any specific legal right, but to consider instead the widest possible range of normative theories. It is hence possible (and even likely) that the Sixth Amendment right to a jury trial would, as interpreted by the Supreme Court mechanically foreclose a machine decision on facts related to guilt or innocence.[198] My project here, however, is not focused on mechanical application of the law: I am interested in whether there is a plausible normative theory upon which that application can rest. (And if the answer is no, so much the worse for Sixth Amendment jurisprudence). Moreover, the taxonomy of reasons offered here does not rely solely on ambiguous, portmanteau concepts such as 'autonomy' and 'dignity.'[199] These concepts, to be sure, have played a leading role in normative theorizing about the state's obligations in respect to various decision-making processes under the rubric of "due process."[200] But terms such as "autonomy" and "dignity" require specification. They have wide and variegated intellectual histories within and beyond the law. For example, a 2011 survey of the manner in which the Supreme Court used the word "dignity" found five different usages by the Justices alone: "status as dignity, equality as dignity, liberty as dignity, personal integrity as dignity, and collective virtue as dignity."[201] And in the philosophical tradition, Schopenhauer once described dignity as "the shibboleth of all the perplexed and empty-headed moralists who concealed behind that imposing expression their lack of any real basis in morals."[202] Abstract and vague normative terms such as 'autonomy' and 'dignity' are generally useful only after they have been colored and bounded through the invocation of a full blooded normative theory.[203]

---

[198] *See supra* text accompanying notes 71 to 72.

[199] These ideas have been invoked in other scholarship on algorithmic tools as decisive normative grounds—not always in the clearest of fashions. *See, e.g.,* Tal Z. Zarsky, *The trouble with algorithmic decisions: An analytic road map to examine efficiency and fairness in automated and opaque decision making*, 41 SCI. TECH. & HUMAN VALUES 118, 118-19 (2016) (invoking "autonomy-related concerns that involve harms to individual dignity"); *see also* Margaret Hu, *Algorithmic Jim Crow*, 86 FORDHAM L. REV. 633, 696 (2017) (offering a similar argument).

[200] *See, e.g.,* Obergefell v. Hodges, 135 S. Ct. 2584, 2597 (2015) (identifying "individual dignity and autonomy" as among "the fundamental liberties" protected by the Due Process Clause"); *accord* Jerry L. Mashaw, *Administrative Due Process: The Quest for a Dignitary Theory*, 61 B.U. L. REV. 885, 901-04 (1981) [hereinafter 'Mashaw, *Administrative Due Process*'] (discussing dignitary definition of due process).

[201] Leslie Meltzer Henry, *The Jurisprudence of Dignity*, 160 U. PA. L. REV. 169, 190 (2011); *see also* Vicki C. Jackson, *Constitutional Dialogue and Human Dignity: States and Transnational Constitutional Discourse*, 65 MONT. L. REV. 16 (2004) (canvassing the use of the term "dignity" in state law and international law). For a similar taxonomy of different usages of "autonomy," see John Christman, *Constructing the Inner Citadel: Recent Work on the Concept of Autonomy,* 99 ETHICS 109 (1988) (reviewing the conceptions of autonomy employed in recent philosophical literature); *see also* Richard H. Fallon, Jr., *Two Senses of Autonomy*, 46 STAN. L. REV. 875, 876-77 (1994) (distinguishing a "descriptive" and an "ascriptive" sense of autonomy in reference to First Amendment debates).

[202] MICHAEL ROSEN, DIGNITY: ITS HISTORY AND MEANING 1 (2012) (citation omitted).

[203] For example, Martha Nussbaum has employed the capabilities approach to human wellbeing to give the idea of dignity meaningful content. Martha C. Nussbaum, *Human Dignity and Political Entitlements*, *in* HUMAN DIGNITY AND BIOETHICS: ESSAYS COMMISSIONED BY THE PRESIDENT'S COUNCIL ON BIOETHICS 351 (Adam Schulman & Martha C. Nussbaum, eds., 2008). Jeremy Waldron uses the history of human rights law to discern a conception of dignity that turns on a repudiation of certain kinds of social ranking. JEREMY WALDRON, DIGNITY, RANK, AND RIGHTS (2012). Neither of

To avoid question-begging arguments, I think it is useful to begin with quite specific ways in which machine decisions might implicate particular normative values. I also think it is useful to assume that the algorithm in question does not rely on tainted training data, has not been malignantly manipulated, and has been competently designed. That is, I will not stack the deck by assuming the machine decision to be flawed ab initio. Further, I return in my conclusion to the law of law in catalyzing improvements in the accuracy of state classification.

Before turning to the three potentially positive grounds for a right to a human decision, I consider first an appealing ground that turns out on closer inspection to have little load-bearing capacity. This is the notion of accuracy. Having gotten this out of the way, I then analyze whether any of the three clusters of normative values that might be used to support a right to a human decision do so. I conclude that the foundations for this right are far from secure.

A. Accuracy in Decision-Making

To begin with, I think it is useful to clarify one normative strand that is not plausibly braided into the right to a human decision's foundation. The Supreme Court has at times suggested that the Due Process Clause of the Fourteenth Amendment creates an entitlement to an accurate decision.[204] Typical of the Court's pronouncements in this regard is its assertion that "[t]he function of legal process, as that concept is embodied in the Constitution, and in the realm of fact-finding, is to minimize the risk of erroneous decisions."[205] It is tempting to define this as a right to an accurate and true decision. If a decision deviates from ground truth, then I am wronged. But even in high-stakes contexts such as criminal cases or postconviction review of capital punishment, the Supreme Court has shied away from a personal right to a *true* determination.[206] A right to accuracy is instead understood in probabilistic terms, such as "beyond a reasonable doubt."[207] It is further viewed as an attribute of an adjudicative process as a whole. A system that can "reduce the risk of error over the aggregate of cases to an acceptable level"[208] is sufficient for constitutional purposes.

---

these conceptions of dignity is obviously to the fore in the algorithmic context. Writing in a Hegelian vein, Margaret Radin has suggested that "autonomy is [best] understood as abstract rationality and responsibility attributed to an individual." Margaret Jane Radin, *Property and Personhood*, 34 STAN. L. REV. 957, 960 (1982).

[204] *See* Honda Motor Co. v. Oberg, 512 U.S. 415, 430 (1994) (noting that "arbitrary and inaccurate adjudication" can violate Due Process); Martin H. Redish & Lawrence C. Marshall, *Adjudicatory Independence and the Values of Procedural Due Process*, 95 YALE L.J. 455, 476 (1986) ("The due process protections such as notice, hearing, and right to counsel are valuable because they contribute to the goal of accuracy.").

[205] Greenholtz v. Inmates of the Neb. Penal & Correctional Complex, 442 U.S. 1, 13 (1979); *accord* Heller v. Doe *ex rel.* Doe, 509 U.S. 312, 332 (1993) (defining Due Process in terms of an interest in an "accurate determination of the matters before the court"); Jerry L. Mashaw, *The Supreme Court's Due Process Calculus for Administrative Adjudication in* Mathews v. Eldridge*: Three Factors in Search of a Theory of Value*, 44 U. CHI. L. REV. 28, 48 (1976) 'hereinafter Mashaw, *Due Process Calculus*'] ("The *Eldridge* Court ... views the sole purpose of procedural protections as enhancing accuracy, and thus limits its calculus to the benefits or costs that flow from correct or incorrect decisions.").

[206] *See* Dist. Attorney's Office v. Osborne, 129 S. Ct. 2308, 2321 (2009) (explaining that whether "actual innocence" exists as a federal right remains an "open question"); Herrera v. Collins, 506 U.S. 390, 404 (1993) (explaining that "actual innocence" has never been held to be an independent constitutional claim).

[207] *See, e.g.*, *In re* Winship, 397 U.S. 358, 372 (1970) (Harlan, J., concurring) ("[I]t is far worse to convict an innocent man than to let a guilty man go free.").

[208] Patrick Woolley, *Rethinking the Adequacy of Adequate Representation*, 75 TEX. L. REV. 571, 630 (1997)

Can a due process right to an accurate decision, understood in these systemic and probabilistic terms, provide a normative foundation for a right to a human decision? Let's bracket the idea that a right to a human decision can be grounded on the idea that a machine is incapable of taking new evidence from a regulated subject,[209] and focus on the pure idea that humans are *eo ipso* better than machines. The evidence collated in Part II suggests that the answer will generally be no. As Part II explained, machine learning performs a set of tasks that overlaps with a set of human decisions. Because the current crop of algorithmic tools identify correlational rather than causal relationships, there is a cluster of empirical questions that they are not well designed to answer.[210] Although an individual might have a legitimate complaint if subject to an algorithmic decision on a matter of causal inference, her claim is not really about accuracy so much as the inaptness of the method employed.

For the class of tasks that can be performed by either a human or a machine-learning tool, available evidence suggests that the latter will often generate fewer positives and negatives in the aggregate, than most human decision making.[211] This is true in contexts such as pretrial bail[212] and domestic-violence related arraignments.[213] So it is not empirically plausible to say that a right to an accurate decision-maker (in this sense) entails a right to a human decision-maker in these cases.[214] Indeed, as the conclusion explores, perhaps the opposite is true such that legal rules might be designed with an eye to improving, rather than substituting, extant machine decisions.

Consider two possible objections to this position. The first runs as follows: Algorithmic classification is inappropriate because the state wholly lacks the necessary training data to measure the variable of interest, and the available data employed to that end is misleading. One can imagine such an objection being levelled (for instance) against teacher evaluation tools that relied on testing modalities that failed to elicit useful measures of children's progress. The problem here, however, is not the choice to use a machine decision rather than a human decision. Rather, it is the prior (human) determination to rely on irrelevant data to measure an unquestionably relevant parameter. It is not a priori obvious that the solution in such cases should be a right to a *human* decision, as opposed to a right to a machine decision well-tailored to the task at hand.

Second, where the population being classified by an algorithm is socially stratified (say, by race or by gender), and where the distribution of errors tracks and reinforces hierarchical fault lines, I think there are serious normative concerns that warrant careful scrutiny. But their resolution turns out to be

---

[209] *See infra* Part III.B.2.

[210] *See* Mullainathan & Spiess, *supra* note 144, at 88 (distinguishing between prediction—the uncovering of generalizable patterns—and parameter estimation, and noting that machine learning does well the former but not the latter). The virtue of machine learning tools is their use of "flexible functional forms [that] allow us to fit varied structures of the data." *Id.* at 91-92

[211] *See supra* text accompanying notes 136 to 138.

[212] Kleinberg et al., *supra* note 138, at 237-38.

[213] Richard A. Berk, Susan B. Sorenson & Geoffrey Barnes, *Forecasting domestic violence: A machine learning approach to help inform arraignment decisions*, 13 J. EMP. L. STUD. 94, 110 (2016) (finding that the release rate of 20 percent repeat offenders in a pool of domestic violent defendants could be dropped to a 10 percent rate through a move from judicial to machine-led determinations).

[214] It may be that an algorithmic tool is implemented in such a way that the rate of false positives increase. But then the objection is the faulty human implementation, and not the algorithm itself.

quite difficult. Studies of algorithmic bail tools demonstrate that the most common population-wide measures of false-positive rates cannot simultaneously be equalized.[215] Instead, equalizing one measure of false positives inevitably leads to an inequality in another measure of false positives. This puzzle, I have suggested elsewhere, is better characterized as a problem of equity rather than accuracy.[216] It is not solved, in any case, by reverting to a more error-prone human decision-making protocol. The same maldistributions of error can arise, just with higher numbers of false positives, with human decisions. As a result, this problem should be considered separately from any right to a human decision.

B. Subject-Facing Grounds

A more plausible set of normative foundations for a right to a human decision can be generated by focusing on the person subject to a machine decision. She may feel disempowered and enervated by her exclusion from any effectual role in the process in which they are engaged. That is, she experience hedonic loss because of an absence of effectual *participation*. Second, an automated decision is seemingly impermeable to human-offered reasons relevant to the decision being taken. The machine, in other words, extinguishes any *opportunity to give reasons* that the individual may seek. An individual may want to give reasons directly pertaining to their treatment and the accuracy of a machine judgment, or alternatively seek to offer an explanation that runs beyond the formal scope of the governing rule. That is, they may want to vindicate an accuracy interest with bespoke information or else see an exception from the rule even if it would normally apply to their case.

I explore here whether a right to a human decision might be predicated on the interest in bare participation, or alternatively the interest in giving (broadly understood) reasons.[217] I conclude that there are more reasons for skepticism than hope on either ground about the right to a human decision.

1. Participation

A bare right to be involved in an important decision is often treated as meaningful even when it cannot be justified or explained in instrumental or accuracy-related terms. That right might be said to flow from a "deep-rooted historic tradition that everyone should have his own day in court,"[218] even if that participatory entitlement will not necessarily alter the outcome of a proceeding. This interest might be founded on the felt human "need to explain and justify our actions," such that "the

---

[215] Roughly speaking, there are different ways of measuring the rate of false positives and false negatives, and the various metrics almost inevitably point in different directions. For mathematical proof of this point, see Sam Corbett-Davies et al., *Algorithmic decision making and the cost of fairness*, *in* PROCEEDINGS OF THE 23RD ACM SIGKDD INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING 797, 798 (2017); Jon Kleinberg, Sendhil Mullainathan & Manish Raghavan, *Inherent trade-offs in the fair determination of risk scores* 4, 9 (2016), https://arxiv.org/abs/1609.05807.

[216] *See, e.g.,* Huq, *supra* note 45, at – (analyzing how different notions of accuracy within a population can be applied in a criminal justice context, and suggesting that racial equity is advanced by minimizing the aggregate cost to racial minorities).

[217] I use this term to cover both reasons that count as justifications and reasons that count as excuses. Nothing turns on the difference between them in this context.

[218] Martin v. Wilks, 490 U.S. 755, 762 (1989).

loss of the opportunity to do so denies our self-worth."[219] Participation, on this view, creates immediate hedonic gain.[220] Alternatively, participation can be framed as the exercise of "moral responsibility as [an] equal citizen[]" upon which rests the law's "moral claim to the citizen's allegiance."[221] In this noninstrumental vein, participation has been glossed as a morally important manifestation of autonomy,[222] or as a manifestation of dignity.[223] The overlap in the scope of those two general terms is suggestive of their joint and overlapping ambiguity.[224]

To evaluate the strength of a noninstrumental participation rationale, I consider when and how that interest is recognized in present law. This is not to assume contemporary law is a precise measure of moral value. It is not. But it provides a rough guide to the weight that an interest is generally accorded. If the law now does not recognize a participation interest in situations akin to those in which a right to a human decision would operate, and if there is no serious objection to that failure, that is some reason to be skeptical of that right. Conversely, if a right to participate is recognized, or strong arguments for its recognition exist, this is some reason for thinking the right should be taken more seriously.

One place where a bare right to participation, even when it is inconsistent with a defendant's interest, is embodied in jurisprudence is the Sixth Amendment right to counsel.[225] The latter constitutional provision extends to create rights to self-representation and to counsel-of choice in the criminal adjudication context, notwithstanding the absence of reasons to think the former in particular yield better outcomes for defendants.[226] The strength of the autonomy-related justification for these

---

219 Mashaw, A*dministrative Due Process*, *supra* note 200, at 903.

220 *Cf.* Amershi et al., *supra* note 182, at 111 (noting that transparency about an algorithm's operation is related to greater user satisfaction).

221 T.R.S. Allan, *Procedural Fairness and the Duty of Respect*, 18 OXFORD J. LEG. STUD. 497, 509 (1998). I am not sure I have a complete response in what follows to Allan's claim. If that claim is the brute assertion that the state's moral legitimacy rests on a particular form of personal participation in legal processes, then it is probably not subject to refutation by the kind of legal and doctrinal examples I develop below. Because I offer no full-blown theory of the state's legitimacy here, I cannot fully respond to this version of the claim. It must suffice here for me to say that I do not find the notion that the state to be legitimate must generate specific, personal involvement in its deliberation compelling as a general matter.

222 Jane Rutherford, *The Myth of Due Process*, 72 B.U. L. REV. 1, 57 (1992) ("When an individual participates in government decisionmaking she has an opportunity not only to influence the accuracy and enhance the legitimacy of the decision, but also to exercise autonomy").

223 Mashaw, *Due Process Calculus*, supra note 205, at 49-52 (identifying dignity as an important consideration for due process); Sanford H. Kadish, *Methodology and Criteria in Due Process Adjudication--A Survey and Criticism*, 66 YALE L. J. 319, 347 (1957) (identifying dignity as a basic due process value).

224 In a different formulation of the right to participate, Lawrence Solum has focused on the effect of participation on the legitimacy of an adjudicative system. Lawrence B. Solum, *Procedural Justice*, 78 S. CAL. L. REV. 181, 191, 274 (2004) (arguing that "a right of participation can be justified for reasons that are not reducible to either participation's effect on accuracy or its effect on the cost of adjudication").

225 *See* U.S. Const. amd VI ("In all criminal prosecutions, the accused shall… have the Assistance of Counsel for his defence.").

226 Faretta v. California, 422 U.S. 806, 819 (1975) ("The Sixth Amendment does not provide merely that a defense shall be made for the accused; it grants to the accused personally the right to make his defense."). For counsel of choice doctrine, see United States v. Gonzalez-Lopez, 548 U.S. 140, 146 (2006) ("[The Sixth Amendment] commands, not that a trial be fair, but that a particular guarantee of fairness be provided—to wit, that the accused be defended by the counsel he believes to be best."); *accord* Flanagan v United States*,* 465 US 259, 268 (1984) (reasoning that this right "reflects

rights remains contested among academics.[227] In a recent choice-of-counsel case, nevertheless, a plurality of the Court described the defendant's right to elect counsel as "fundamental."[228]

The participation interest embodied in these decisions, though, is not necessarily one of *personal* involvement, and certainly not a matter of participation in whatever form the defendant thinks appropriate. The Sixth Amendment extends invariantly to both the right to self-representation and also the right to choose one's own counsel. When the Sixth Amendment is manifest through the latter form (as is more often the case), then there is no interest in personal participation at stake. What is at stake instead is the free choice of agents (lawyers) who will act as an individual's representation in a given proceeding. This can be described as a participation right. But doing so elides the agency relationship between lawyers and clients, as well as the extent of control exercised by professionals in the adjudicative context. It is reasonable to think that the relevant autonomy protected by the Sixth Amendment is that to elect one's counsel; the right of self-representation on this view, is a byproduct of this more general right to choose one's own counsel. It is not an instantiation of a specifically protected participation interest.

Even if a criminal defendant exercises a right to self-representation, the right is circumscribed in highly formalized ways that establish "a certain distance" between judges and parties.[229] A criminal defendant who exercises her right of self-representation is not exercising a right to represent herself in whatever fashion she wishes. To the contrary, she is invoking a highly constrained entitlement under conditions in which she likely will lack the epistemic competences to navigate. The Sixth Amendment's right to participate in one's own criminal proceeding, therefore, is a highly limited one. It is not clear on its face whether it should be understood as the normative core of that right, as opposed to an unanticipated side-effect of creating a general counsel-of-choice regime, one that cannot be narrowed to exclude self-representation without some elaborate textual or doctrinal legwork. On this view, the right to participate in a criminal proceeding by invoking the Sixth Amendment is not well understood as an interest in participation per se, as opposed to a peculiar (and marginal) instantiation of the defendant's free election of counsel.[230] What is the case in the criminal context spills over also into the civil context. Any right to participate in a civil proceeding is probably best understood as a right to participate through an agent such as a lawyer. (We will turn in a moment to the due process interests

constitutional protection of the defendant's free choice independent of concern for the objective fairness of the proceeding"). Notice that the same might be said for elections more generally.

[227] *Compare* John Rappaport, *The Structural Function of the Sixth Amendment Right to Counsel of Choice*, 2016 SUP. CT. REV. 117, 118 (2016) (concluding that "majestic-sounding notions of fairness and autonomy, respectively--struggle to explain counsel-of-choice doctrine"), *with* Erica J. Hashimoto, *Resurrecting Autonomy: The Criminal Defendant's Right to Control the Case,* 90 B.U. L. REV. 1147 (2010) (defending the doctrine in autonomy terms).

[228] Luis v. United States, 136 S. Ct. 1083, 1089 (2016) (plurality op.).

[229] Emily Buss, *The Missed Opportunity in* Gault, 70 U. CHI. L. REV. 39, 47 (2003) ("The formality of the procedure and the qualification of the decisionmaker as a neutral arbiter of the law ensure a certain distance between decisionmaker and parties designed to increase the reliability of decisions made, even while it dignifies the parties and the interests at stake.").

[230] *Cf.* Rappaport, *supra* note 227, at 118 (characterizing the choice of counsel rule as "a weak, system-level safeguard against socialization of the criminal defense bar").

in notice and a chance to give reasons).[231] Perhaps this gap is undesirable.[232] But it hardly supports the intuition that a bare right to participate is sufficiently widely regarded to justify a right to a human decision.

Most forms of adjudication, moreover, entail a sufficiently complex and formal enterprise that a right to personal participation in decisions with direct, personal effects is not plausible. As the recent recognition of Sixth Amendment rights in the plea bargaining context suggests,[233] even informal dispute resolution in the judicial context requires a modicum of technical expertise that few nonlawyer individual litigants possess. A right to *personal* participation in adjudication is inconsistent with this complexity.

Against this background, it is hard to see the force of the noninstrumental, participation-based argument in favor of a right to a human decision even when life or liberty are at stake. Machine learning instruments, recall, may not be much more opaque than human decision-makers.[234] Similarly, there is not necessarily much phenomenological distance between the bafflement an unschooled criminal defendant may reasonably feel when faced with the reticulate and complex forms of the criminal justice system, and the confusion elicited by an algorithmically derived outcome. For all practical purposes, both are black boxes. Moreover, there is no real social movement so far as I can tell to make the criminal adjudicative process simpler—as distinct from fairer or more favorable to defendants—for its own sake. If the interest in noninstrumental participation gets short shrift under circumstances in which human liberty will often be directly imperiled, it would be surprising to learn that there was a powerful human interest in personal participation that carried over into the machine learning context.

A final reason for resisting a participation-based right to a human decision turns on the inadequacy of this assertion. An ex post human decision will often not be the optimal response to felt psychological loss. Often, the best way of 'explaining' a discrete decision will not be through human review, which may cast limited light on the operation of a complex algorithm. At least one algorithmic tool has been developed instead as a means of "explain[ing] the predictions of any classifier or regressor in a faithful way."[235] When the instrument most likely to yield a comprehensible account of machine-learning outputs is itself a machine, the need for a human to diagnose or "soothe"[236] the grievances of affected individuals must rest not on a demand for explanation, but rather on a raw and unreasoned need for a human interlocutor. But this need for human interaction may be a contingent feature of social experience, and that which strikes us today as dehumanizing or insensitive will appear

---

[231] For example, "in absentia" removal proceedings are envisaged, and commonly employed, under Title 8. 8 U.S.C.A. § 1229a(5)(A) (2018).

[232] That said, I am inclined to think in absentia removal is per se objectionable.

[233] *See* Padilla v. Kentucky, 559 U.S. 356, 373 (2010) (applying Sixth Amendment right to counsel to plea bargaining that has immigration consequences). For Padilla's extension beyond the immigration context, see Missouri v. Frye, 132 S. Ct. 1399, 1408 (2012); Lafler v. Cooper, 132 S. Ct. 1376, 1388 (2012).

[234] *See supra* text accompanying notes 145 to 173.

[235] Marco Tulio Ribeiro, Sameer Singh & Carlos Guestrin, *Why should I trust you?: Explaining the predictions of any classifier*, PROCEEDINGS OF THE 22ND ACM SIGKDD INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING 1135, 1135 (2016).

[236] Matthew J.B. Lawrence, *Mandatory Process*, 90 IND. L.J. 1429, 1432 (2015).

to our children as merely sensible and mundane.[237] Right now, the demand for human review in the teeth of its likely costs and available alternative responses, might seem little more than an aesthetic preference about the manner in which one interacts with state actors. I am not sure that is enough to get a right to a human decision off the ground.

Perhaps there is a way of rehabilitating this psychological argument. Perhaps an individual subject to a machine decision might seek ex post human review not because she thinks the decision incorrect, but because she hopes that exogenous factors will prompt some mitigation of the decision's consequences. She seeks, in other words, mercy—or the official exercise of discretion to mitigate a legal consequence that is otherwise a person's lawful fate.[238] But a hope for mercy is a poor foundation for any 'right' to a human decision. Mercy is generally understood as a *discretionary* act to which one has no entitlement. As such, mercy is an increasingly elusive quality even in the criminal justice context where it has probably the most plausible berth.[239] Its presently penurious titration may be unwarranted, but still it is hard to see why the case for reviving mercy would begin in the algorithmic domain.

2. *Reason Giving*

But perhaps it is a mistake to think of participation in noninstrumental terms. Instead, it may be better to focus on the tangible ways in which participation can alter outcomes. There are several possible arguments to this end. Most obviously, participation matters because of the factual contributions an individual can make to the consideration of her case. The most important form such a contribution might take turns on the individual's ability to supply information, and more generally to offer reasons, for a decision to be made in her favor.[240] Recall that this is the force of Cathy O'Neill's anecdote about the welfare applicant Catherine Taylor, which I recounted earlier[241]: Only Taylor, we are to presume, had the information needful to a right resolution of her own case, and only human review could have elicited that information. This argument from reason-giving closely complements the most plausible technological understanding of a right to a human decision. As Part II explored, human decisions permeate and structure machine-learning tools.[242] The most plausible gloss of a claim to additional human input needs to hinge on an ex post human role after a machine-learning decision has been delivered. That would mean an individual subject to a machine decision could speak directly

---

[237] Robot companions capable of recognizing and dynamically responding to their charges' mutative emotional states lie within the technological event horizon. For a review of the current science, see Giulio Sandini et al., *Social Cognition for Human-Robot Symbiosis-Challenges and Building Blocks*, 12 FRONTIERS IN NEUROROBOTICS 34 (2018).

[238] Aziz Z. Huq, *The Difficulties of Democratic Mercy*, 103 CAL. L. REV. 1679, 1681 (2015). Mercy generally involves a "remission of deserved punishment, in part or in whole, to criminal offenders on the basis of characteristics that evoke compassion or sympathy but that are morally unrelated to the offender's competence and ability to choose to engage in criminal conduct." Dan Markel, *Against Mercy*, 88 MINN. L. REV. 1421, 1436 (2004). There will be some instances in which mercy so defined can be exercised even after an algorithmic classifer has been used to impose a decision.

[239] On the decline of the pardon, see Rachel E. Barkow, *The Ascent of the Administrative State and the Demise of Mercy*, 121 HARV. L. REV. 1332, 1348–49 (2008). On the decline of the jury as a mitigating institution, see Nancy J. King, *Silencing Nullification Advocacy Inside the Jury Room and Outside the Courtroom*, 65 U. CHI. L. REV. 433, 492–94 (1998).

[240] In discussing the right to give reasons in this subsection, I will not distinguish between an individual's ability to offer empirical evidence and her ability to proffer legal or normative claims that do not rest on information about the world. I suspect that in practice reason-giving will entail some blend of factual and normative assertion. To distinguish them here serves no useful purpose.

[241] *See supra* text accompanying notes 17 to 18.

[242] *See supra* text accompanying notes 175 to 188.

to that decision by drawing it to a human's attention. Such a right would respond to the possibility, for instance, that the "feature values" used to train an algorithm excluded some parameter of relevance to a subset of individuals, but not the general population.[243]

Separately, the right to give reason might be defended on outcome-independent, yet instrumental, grounds. In the longer term, for example, participation might work as a balm to those whose causes falter.[244] Consistent with this view, the literature on procedural justice associated with Tom Tyler has pressed the empirical claim that the opportunity to be heard by an official is associated with higher rates of legal compliance after an interaction has passed.[245] Systemic legitimacy, on the procedural justice view, flows from an embedding of the opportunity to give reasons in the texture of citizen interactions with the legal system.[246]

Arguments of either form might seek doctrinal footing in the Due Process Clause.[247] The requirement of a hearing, to be sure, does not extend to *all* state decisions that directly and immediately affect individuals. Legislation and law-like regulations promulgated by agencies, for example, can dramatically and immediately change an individual's rights, obligations, and exposure to coercive risk. Individuals have no entitlement to individualized participation in respect to them.[248] The boundary line between permissible legislative fiat (where due process does not apply) and forms of government action to which it does attach is not wholly clear.[249] But it does exist. Some commentators suggest, for example, that "due process requires an oral hearing where particularized deprivations affecting a small number of people based on adjudicative facts are concerned,"[250] regardless of whether the outcome is denominated as legislation. At least some of the criminal justice and social welfare decisions for which algorithmic tools are presently used plainly fall within this domain. This individual interest in giving reasons is, moreover, distinct from the more general interest in accuracy addressed above[251]:

---

[243] Domingos, *supra* note 97, at 78; *see also* LeCun et al., *supra* note 33, at 436 (discussing feature selection in algorithmic design).

[244] For an instrumental argument to this effect sounding in behavioral economic terms, see Matthew J.B. Lawrence, *Mandatory Process*, 90 IND. L.J. 1429, 1432 (2015) ("The inherent value of participating in a dispute resolution process comes in part from its power to soothe such a grievance when it does occur, win or lose.").

[245] TOM R. TYLER & YUEN J. HUO, TRUST IN THE LAW 7 (2002) ("There is considerable evidence that when people regard the particular agents of the legal system whom they personally encounter as acting in a way they perceive to be fair and guided by motives that they infer to be trustworthy, they are more willing to defer to their directives ...."). The evidence for this effect has been recently challenged. *See* Daniel S. Nagin & Cody W. Telep, *Procedural Justice and Legal Compliance*, 13 ANN. REV. L. & SOC. SCI. 5 (2017).

[246] Stephen J. Schulhofer et. al., *American Policing at A Crossroads: Unsustainable Policies and the Procedural Justice Alternative*, 101 J. CRIM. L. & CRIMINOLOGY 335, 345 (2011) ("Empirical research indicates that this sort of legitimacy is sustained not by an aggressive style that subordinates individual rights but rather by something closer to its opposite--practices that can be grouped under the heading of procedural justice.").

[247] The seminal cases are *Goldberg v. Kelly*, 397 U.S. 254 (1970); *Bell v. Burson*, 402 U.S. 535 (1971); and *Arnett v. Kennedy*, 416 U.S. 134 (1974).

[248] *See* Bi-Metallic Inv. Co. v. State Bd. of Equalization, 239 U.S. 441 (1915); Londoner v. City of Denver, 210 U.S. 373 (1908).

[249] Henry Friendly, *"Some Kind of Hearing"*, 123 U. PA. L. REV. 1267, 1276-77 (1975) ("[I]t seems impossible at the moment to predict at what level, if any, the Court will set the floor below which no hearing is needed.").

[250] Adrian Vermeule, *Conventions of Agency Independence*, 113 COLUM. L. REV. 1163, 1213 (2013).

[251] *See supra* Part III.A.

Its claim is not quite that the machine itself is inaccurate overall, but rather that the addition of human review can eliminate a class of false negatives (positives) by leveraging private information held by regulated subjects.

In my view, this is the strongest argument for a right to a human decision. The fact that state agencies may have incentives to precipitously adopt deeply flawed machine-learning systems only adds to its appeal. Yet without rejecting the normative concerns raised by cases such as Catherine Taylor's, I want to push back on the idea that that a right to a human decision is a compelling response. In particular, I want to resist the temptation to conflate the short-term gain from human review in cases such as Taylor's with the question of dynamic optimality: That is, how adoption of such a right will, in the long term, shape both desirable and undesirable outcomes, and (implicitly) whether there is some better alternative vehicle for addressing the normative concerns.

A first, concededly tenuous, reason for hesitation is doctrinal. It is not clear that the Due Process Clause requires a supplemental human action to ensure a sufficiently accurate machine decisions The adjudicative forms that due process can take are often desultory.[252] It is not at all obvious that algorithmic tools cannot, outside of the specific context of the Sixth Amendment's jury trial right, supply whatever due process is constitutionally needful (unless their results are entirely orthogonal to the quality being measured). This legal conclusion, however, arguably warrants relatively little weight. Due process jurisprudence, after all, might simply be wrong, and in need of updating in light of technological change.

The second reason has more heft. Installation of a human decision as a backstop to a machine decision might have perverse and undesirable consequences for the regulated population that outweigh any participation-related benefits.[253] Rather, depending on the empirics of the situation, the addition of a human decision may become a form of what Adam Samaha called "undue process," that traduces "constitutionally mandated ceilings on government process."[254] This second argument turns on the premise that not all process is "due." As Judge Henry Friendly observed in his canonical reflection on the hearing requirement, every additional increment of process comes at a cost, since "procedural requirements entail the expenditure of limited resources, [so] that at some point the benefit to individuals from an additional safeguard is substantially outweighed by the cost of providing such protection."[255] Where the addition of a procedural step has the expected systemic effect of increasing the overall frequency of error rates, or generating some other cost, there would be reason to pause and reconsider the mandate as a matter of law and as a matter of public policy.

---

[252] Goss v. Lopez, 419 U.S. 565, 579 (1975) (mandating "some kind of notice and afforded some kind of hearing" for disciplined public school students, without adding much more detail).

[253] Rights often have implementation-related costs that spill over to others, who are not exercising the right. For instance, criminal procedure rights can make law enforcement more costly, and thus reduce the extent to which the state can generate public security for all. The argument here focuses on a different possibility: that the exercise of a right has costs that are spread across the population putatively benefiting from the right.

[254] Adam M. Samaha, *Undue Process*, 59 STAN. L. REV. 601, 630 (2006). Samaha was careful to acknowledge the novelty of this possibility, and its tension with extant doctrine. I invoke his idea here not so much to suggest that there might be a cognizable 'undue process' claim, but rather to press the perversity of insisting upon human involvement in certain cases.

[255] Friendly, *supra* note 76, at 1276.

Consider first the effect of a backstop human decision-maker for all the outputs of a machine decision-learning process in terms of net false positives and false negative rates.[256] Of course, "even well-designed" algorithmic tools will make mistakes.[257] But the addition of a human backstop to a well-designed machine decision will not necessarily increase the overall rate of accurate judgments. As noted, machine decisions are often less error-prone than close human substitutes.[258] It is not safe to assume that a human will be able to identify and correct all of, and only, the instances in which the machine erred. To the contrary, there is a real possibility that human input will lead to a *higher* error rate. Further, if human right is envisaged when one classification is reached, and not the other, there is risk that the resulting errors will be unevenly distributed across the population. Absent some reason to think the machine was itself biased, it is hard to see how a higher, asymmetrically distributed error rate is desirable. Where an algorithmic tool is flawed, moreover, it does not follow that ex post human review is therefore "due." Rather, there is every reason to believe that what is "due" is a better machine decision rather than a reliably unreliable human one. That is, even when a machine decision is flawed, the anticipated trade-off between expected accuracy and participation trade is likely to tilt against a human decision.

To motivate this analysis, consider a recent instance in which a non-machine algorithm was supplemented with an ex post right of human review. Starting in 2007, the Supreme Court expanded judges' discretion over sentencing in federal court by rejecting the binding force of a very elementary algorithm, the federal sentencing guidelines.[259] Interjudge sentencing disparities subsequently sharply increased, in some jurisdictions almost doubling.[260] So too did racial disparities.[261] Perhaps there is a happy story to be told here about the propensity of judges to match sentences on a range of offender characteristics beyond those contained in a presentencing report. I doubt it. To date, there is instead every reason to think judicial discretion has had dismaying, and socially destructive, effects.[262] Adding

---

[256] This was, recall, the position of the Wisconsin Supreme Court in respect to algorithmic sentencing tools. State v. Loomis, 881 N.W.2d 749, 760 (2016); *supra* text accompanying notes 79 to 82.

[257] Kroll, *supra* note 152, at 11.

[258] *See supra* text accompanying notes 136 to 138.

[259] United States v. Booker, 543 U.S. 220, 233, 247 (2005) (invalidating mandatory force of Federal Sentencing Guidelines); *see also* Rita v. United States, 551 U.S. 338, 350-55 (2007) (appellate presumption of reasonableness for within-Guidelines sentences); Gall v. United States, 552 U.S. 38, 48-49 (2007) (rejecting heightened appellate review for out-of-guidelines sentences); Kimbrough v. United States, 552 U.S. 85, 109-10 (2007) (rejecting statutory constraints on sentencing).

[260] Crystal S. Yang, *Have Interjudge Sentencing Disparities Increased in an Advisory Guidelines Regime? Evidence from* Booker, 89 N.Y.U. L. REV. 1268, 1333 (2014) (finding that "interjudge disparities have doubled from the period of mandatory Guidelines sentencing to post-*Booker* sentencing, with a defendant potentially receiving a six-month longer sentence if assigned by happenstance to a harsh judge"); *accord* Ryan W. Scott, *Inter-Judge Sentencing Disparity After* Booker: *A First Look*, 63 STAN. L. REV. 1, 3, 52-53 (2010) (similar result for a Massachusetts district court).

[261] Crystal S. Yang, *Free at Last? Judicial Discretion and Racial Disparities in Federal Sentencing*, 44 J. LEGAL STUD. 75, 77 (2015) (finding "significantly increased racial disparities after controlling for extensive offender and crime characteristics" post-*Booker*); *accord* Max Schanzenbach & Emerson H. Tiller, *Strategic Judging under the United States Sentencing Guidelines: Positive Political Theory and Evidence*, 23 J. L & ECON. & ORG. 23 (2007) (similar finding).

[262] I do not mean to suggest that the pre-*Booker* regime was without faults. Interjudge disparities derived from inconsistent punitive preferences, however, do not appear to have been one. The mere fact of disparities, moreover, does not alone demonstrate a flawed arrangement. Evaluating when a disparity is unwarranted "requires an idea of why we punish." Kevin Cole, *The Empty Idea of Sentencing Disparity*, 91 NW. U. L. REV. 1336, 1337 (1997). At least the data on racial disparities between similar defendants raises substantial questions as to whether adequate justifications can be identified.

human input to a (simple, non-machine) algorithm may well have done more harm than good. With this example in hand, it is possible to see that arguments for a right to a human decision focus on a specific person who has been wrongly denied a benefit are misleading:[263] We should be concerned not with one person's case, but with the overall mix of wrongful human or machine decisions produced by a system.

Now consider a variant of this argument: Rather than permitting ex post human review of all outcomes, only instances in which the regulated subject suffers a disadvantage would trigger human input. That is, the right would attach only to the Catherine Taylors of the world, and not to those who are granted benefits. This asymmetry could be sharpened. The algorithm could titrate human review as a means of limiting the latter's costs by isolating a subset of individuals who might plausibly offer salient, new facts that could result in a different outcome. Hence, an algorithm that generated a risk parameter as a continuous variable will have a numerical threshold as a classification rule.[264] Individuals classified as exceeding that threshold by a small margin might be allowed to appeal. Those who cleared the threshold by a large margin would have no right to a human decision. In short, the algorithm itself would be designed to select a subset of negative outputs for which the addition of ex post human review might be justified in terms of an accuracy gain. Why would a right to human review *hurt* in this narrow, asymmetrical form? How, that is, could it not be due?

I remain skeptical. To begin with, this argument (for asymmetrical, and narrow-gauged ex post human review) assumes that human review will correct false positives and only false positives. But there is no reason to assume that only meritorious subjects of an adverse decision will appeal. ¸It is hence possible that even narrowly calibrated human review will increase the net volume of errors by reversing more true positives than false positives. The class of Catherine Taylors, that is, won't all get relief, while a class of undeserving beneficiaries who have correctly had their claims denied—call them Elizabeth Taylors—will prevail. Lest this sound implausible, think about how selection into human appeals will operate. There is no reason a priori to think that only and all those with relevantly corrective private information will appeal ex post to a human. It is more plausible to think that wealth or epistemic resources or social class will predict the tendency to appeal.[265] Whether this results in more or less errors overall is anyone's guess. Indeed, different rates of falsification should be expected among the Elizabeth Taylors than the Catherine Taylors. The result will be a pooling equilibrium, rather than a separating equilibrium, in which the human decision-maker is presented with a mix of true and false private information. It is, once again, rather fantastical to think that a human decision-maker has a costless and frictionless mechanisms for sorting the earnest Catherines from the Machiavellian Elizabeths. What seems from O'Neill's threshold example to be a simple, relatively costless step turns out on inspection, therefore, to be highly problematic and very far from weightless.[266] The empirical and technological assumptions of these optimal cases for ex post human review are thus demanding to the point of implausibility.

---

[263] *See supra* note 17 and accompanying text.

[264] Camelia Simon, Sam Corbett-Davies & Sharad Goel, *The Problem of Infra-Marginality in Outcome Tests for Discrimination*, 11 ANN. APP. STAT. 1193, 1194 (2017) (setting out example).

[265] The original insight into the problem of signaling and the possibilities of both pooling and separating equilibrium is to be found in Michael Spence, *Job Market Signaling*, 87 Q. J. ECON. 355, 362-63 (1973); *see also* MARTIN J. OSBORNE & ARIEL RUBINSTEIN, A COURSE IN GAME THEORY 238 (1994) (describing Spence's signaling game and the resulting pooling and separating equilibria).

[266] Technical solutions to this problem are also costly. Imagine an optimal machine decision designed to be amenable to human interpretation. *See* Desai & Kroll, *supra* note 153, at 11 n.61. This machine produces "a tamper-evident

Second, the superficial appeal of this asymmetric, narrow-gauge human review rests (perhaps implicitly) on the assumption that there are no costs from the reversal of true positives. But this will rarely be so. In the bail context, for instance, the reversal of a true positive may cash out as the avoidable commission of a serious violent crime. In the welfare context, it means an undeserving person gets a benefit that otherwise could have gone to a needy beneficiary. It is wishful thinking to assume away the costs of reversed true positives under any system, although I suspect that the appeal of a right to a human review trades on some suppression of these costs.

Third, it will often be the case that even the asymmetrical, narrow-gauged right to human right will either be otiose from the start, or can be rendered irrelevance by reinforcement learning. An individual's opportunity to supply reasons to a human decision-maker is relevant only if those reasons have some likelihood of influencing a process's outcome. But for many of the decisions for which algorithms might be employed in official hands, such as benefits eligibility or parole revocation, the law delimits a closed set of relevant parameters. That is, the law often employs rules picking out ex ante a fixed set of relevant facts as opposed to standards. The latter allow for more open-ended consideration of enumerated and unanticipated factors.[267] Where an algorithm is applying a legal rule rather than a standard in this sense, it is not clear why the novel reasons or facts that an individual subject to classification hopes to point out ex post matter. The very fact of selecting a rule rather than a standard as the relevant law forecloses their assertion that unexpected facts are pertinent. Even when the algorithm is designed to apply a standard (e.g., dangerousness in the bail context), it may be that the routine application of that standard over hundreds or thousands of cases generates sub-rules based on closed and predictable sets of parameters.[268] These subrules will cover the field of possible facts asserted as salient to a machine decision. Human review ex post will then rarely add anything of value.[269] The interaction of closed legal rules and function selection may therefore render otiose any

record that provides nonrepudiable evidence of all nodes' actions." Haeberlen, Kuznetsov & Druschel, *supra* note 153, at 175. This record could then be examined manually to ascertain whether any error had occurred, or whether the data parameters employed in the classification rule failed to capture a particular parameter of relevance to the individual being ranked by the classifier.

[267] Louis Kaplow, *Rules Versus Standards: An Economic Analysis*, 42 DUKE L.J. 557, 559-63 (1992) (defining a rule as a legal norm given content before regulated subjects act, whereas a standard is a legal norm that is given content after regulated subjects act).

[268] Such rules-standards 'cycling' is observed in many legal domains. *See* Aziz Z. Huq & Jon D. Michaels, *The Cycles of Separation-of-Powers Jurisprudence*, 126 YALE L.J. 346 (2016) (mapping cycles in structural constitutional law); Carol M. Rose, *Crystals and Mud in Property Law*, 40 STAN. L. REV. 577, 598-99 (1988) (making this observation about rules in the property-law context); Adrian Vermeule, *The Cycles of Statutory Interpretation*, 68 U. CHI. L. REV. 149, 150 (2001) (identifying cycling in statutory interpretation).

[269] What of "the social commitment to try to understand each other" and "the potential for connection and community"—i.e., "empathy" and "ethical development"? EUBANKS, *supra* note 8, at 168. Considering the state's use of algorithms in welfare and public benefits contexts, Eubanks argues that absent empathy, bias against minorities and women is more possible. *Id.* I agree that animus is sometimes a failure of empathy. But expanding the institutional space for human empathy is an immediate, or even medium-range, solution for failures of compassion. Absent some dramatic improvement in the street-level quality of human judgment—and Eubanks gives no reason to think street-level officials are overnight going to become better and more fair decision-makers--such expansions will have precisely the opposite effect she advocates: It enables biased or motivated reasoning, and makes the ensuing distortions even harder to remedy. By contrast, reform of machine tools allows ambient bias to be mitigated centrally. Hence, the argument that one needs immediate human contact for better state decision-making is simply a fallacy. It is a fallacy is contradicted by some of the twentieth-century's great achievements of social democracy, from Britain's National Health Service to the American Great Society. Empathy, in short, is not enough: It must be acted upon in a strategic and thoughtful fashion, rather than used as a crutch for superficial and ultimately unavailing responses.

claim to a human decision (or, at a minimum, make a right to a dynamic rather than a static algorithm more plausible).

Fourth, recall that I have so far assumed that the parameters of the training data were selected precisely because they enable "faithful measurement" of an underlying property of interest relevant to the prediction of the target variable.[270] The participation-based argument for a human decision in effect may assume this is not so. It assumes that the training data's parameters are insufficient to maximize accurate results such that ex post reason-giving has a corrective value. But if that is so, the proper response is to improve the training data or to tweak the algorithm. Retail responses, such as a right to an ex post human decision, perversely maintain a deficient status quo, and may even delay the implementation of better, systemic fixes achievable through better machine decision-making.

This last point can be generalized to provide a final reason for resisting a participation-based right to a human decision in all cases. To the extent that human review is understood as a means of verifying the integrity of an algorithmic tool in an individual case, it is hardly clear that retail interventions in respect to specific classification decisions is a well-fitting solution in the long term (even if they might be a good diagnostic tool for sniffing out systemic problems). To the contrary, a range of static oversight tools—which focus on the algorithm's underlying source code—and dynamic oversight tools—which look at the algorithm's behavior in the wild – are alternatively available,[271] and over time might conduce to systematic improvements that are likely to elude a right to human review. Given scarce resources, devoting time to individualized ex post review in lieu of more systemic testing of an algorithm's integrity will likely often generate the perverse effects of lower overall accuracy. Hence, given the marginal nature of due process analysis—which focuses on the discrete positive or negative contribution from any given increment of procedural change—ex post human review of discrete decisions will rarely be an optimizing strategy.

* * *

I suspect that some readers will retain a nagging doubt as to cases such as Catherine Taylor's welfare denial.[272] What, they might ask, can be said to the person who is erroneously ranked? To begin with, it is worth reiterating that the mere fact of an erroneous determination does not, standing alone, establish a legal or a moral wrong. Use of the reasonable doubt standard in criminal trials implies that we are willing to tolerate a certain number of erroneous convictions. The question to ask is whether Taylor was classified by a flawed system, and not whether in her case an error was made.

At the system level, moreover, it is far from clear that addition of a human decision-maker would reduce the net volume of errors. Perhaps human review of both benefit grants and also denials would generate a higher rate or errors, or a more racially skewed distribution of errors, than a machine decision alone. And finally, to anticipate a point to which I will return in the conclusion, the flawed quality of a machine decision does not imply that a human decision-maker would do better: Perhaps the optimal system for all the Taylors of the world entails a better machine decision, rather than human

---

[270] David Lehr & Paul Ohm, *Playing with the Data: What Legal Scholars Should Learn About Machine Learning*, 51 U.C. Davis L. Rev. 653, 679 (2017).

[271] Kroll et al., *supra* note 2, at 547-52 (describing static and dynamic testing protocols); *see also* Wachter, Mittelstadt & Russell, *supra* note 155, at 13-14 (describing the use of counterfactuals to conduct tests of algorithmic integrity).

[272] *See supra* text accompanying notes 17 to 18.

review. Given the greater capacity for accuracy with machine decisions, indeed, there is every reason to think that this condition holds—as I explore briefly in the conclusion. So nothing about my analysis implies that some ex post verification or audits of such computational tools is unwarranted. It is, more narrowly, to reject *retail human* as opposed to *systemic* review as a warranted intervention. But since the right to a human decision in its GDPR Article 22 form, as well as in its less articulate variants, is best described as a retail right that attaches to distinct individuals subject to algorithmic classification, the continuing need for systemic review cannot supply it with a normative foundation.

C. Classifier-Facing Grounds

I turn now to arguments for a right to a human decision that hinge on the character of state action. The idea that an action can be impermissible as a legal or a moral matter because of the way in which the state has behaved, rather than because of its intrusion into some protected zone of individual interest, is familiar in American law. The Supreme Court, for instance, has recently affirmed that individuals are entitled to bring claims alleging that official action rests on a violation of structural constitutional principles such as federalism or the separation of powers,[273] even if the same action with the same impact on an individual's interests could have been achieved though properly constituted governmental action. Another version of this phenomenon arises when the state acts on the basis of impermissible considerations, such as racial or religious identity.[274] At least notionally, state action based on impermissible grounds is unlawful even if the same action on different motivational grounds would not be disallowed.[275] In both these lines of cases, state action is deemed flawed not because of its effects but because of the manner in which the state acted.

A classified-focused justification for the right to a human decision might home in upon one of two arguments. First, it might be argued that a characteristic of lawful state action against individuals is that it is reasoned. Algorithmic decisions, it might be argued, fail that minimal threshold criterion of rationality. Second, those decisions classify individuals on the basis of group-based generalization. As such, they fail to treat them as individuals. This latter point might be understood as a concern about "profiling," or it might be understood as a concern about dignity. Although both of these arguments draw upon deep normative wellsprings, tapping anchoring intuitions in American constitutional law, I ultimately suggest that neither provides a plausible grounding for a right to a human decision.

1. *Reasoned State Action*

The idea that state action, and in particular coercive state action, ought to be firmly grounded on reasons is deeply entrenched in Anglo-American law. Giving reasons, on this account, is "a way of

---

[273] Bond v. United States, 564 U.S. 211, 223 (2011) (stating that "individuals . . . are protected by the operations of separation of powers"). *But see* Aziz Z. Huq, *Standing for the Structural Constitution*, 99 VA. L. REV. 1435, 1490–514 (2013) (doubting this claim).

[274] For examples of decisions disallowing state action that would be permitted in the absence of impermissible considerations, see, e.g., Parents Involved in Cmty. Sch. v. Seattle Sch. Dist. No. 1, 551 U.S. 701, 720 (2007); *see also* Gratz v. Bollinger*,* 539 U.S. 244, 270 (2003) (describing the use of such classifications as "pernicious" (citation omitted)).

[275] For a recent exception, see Trump v. Hawaii, 138 S. Ct. 2392, 2421 (2018) (upholding an immigration-related executive order publicly justified on discriminatory grounds "because there is persuasive evidence that the entry suspension has a legitimate grounding in national security concerns, quite apart from any religious hostility").

showing respect for the subject, and a way of opening a conversation rather than forestalling one."[276] It "attach[es] value to the individual's being told why the agent is treating him unfavorably and to his having taken a part in the decision."[277] A decision, on this view, is consistent with the rule of law only if it is "comprehensible for those subject to the decision."[278] From a midcentury liberal perspective, state power is legitimate when based on "grounds of adequate neutrality and generality."[279] An ardent libertarian might add that reason-giving works as a salutary friction on state action, generating transaction costs that may be preclusive when no public-regarding ground for an action can be articulated. Something of that intuition seems at work in the application of Fourth Amendment law to preclude street stops on the basis of a "mere hunch."[280]

Let us set aside the possibility that algorithms can be designed to issue explained decisions.[281] The demand for reasoned state action still does not provide an adequate basis for a right to a human decision for the simple reason that this is itself only occasionally and incompletely met. Many state decisions issue absent any supporting reasoning. As Lon Fuller noted in his famous 1978 essay on adjudication, the "integrity of adjudication" does "not necessarily" require that "reasons be given for the decision rendered."[282] From street stops to certiorari denials, there are many discrete state interventions within and beyond the adjudicative context that typically lack an explicit justification. Beyond that, statutes can fashion extensive changes to social realities without offering anything by way of adequate normative justification.[283] Indeed, it is now conventional wisdom that legislatures often enact statutory text without reaching a consensus view of the meaning of an element in the statutory text. Ambiguous statutory text—the daily fare of appellate courts—can arise if "Congress had no particular intent on the matter."[284] Nor is it plausible to think that all important adjudicative actions are reasoned in a fulsome sense of that term. When a trial judge denies an evidentiary objection, when an appellate court exercises discretion to permit a nonmandatory appeal, or the Supreme Court denies certiorari, it is hardly clear there are articulable, let alone well-grounded, reasons for the action.[285] The sheer extent of insufficiently reasoned state action is suggestive: Taking that demand at face value would in effect choke the modern state before it could perform any of its basic obligations.

---

[276] Frederick Schauer, *Giving Reasons*, 47 STAN. L. REV. 633, 658 (1995).

[277] Frank I. Michelman, *Formal and Associational Aims in Procedural Due Process*, 18 NOMOS 126, 127 (1977)

[278] Mireille Hildebrandt, *Algorithmic regulation and the rule of law*, PHIL. TRANSACTIONS R. SOC. 1, 3 (2018)

[279] Herbert Wechsler, *Toward Neutral Principles of Constitutional Law*, 73 HARV. L. REV. 1, 15 (1959); *accord* LON L. FULLER, THE MORALITY OF LAW 34 (1964). On the putatively liberal origins of the demand for reasoned generality, see Mark V. Tushnet, *Following the Rules Laid Down: A Critique of Interpretivism and Neutral Principles*, 96 HARV. L. REV. 781, 783-85 (1983).

[280] United States v. Arvizu, 534 U.S. 266, 274 (2002) (citation and quotation marks omitted).

[281] *See supra* text accompanying notes 174 to 194.

[282] Lon L. Fuller, *The Forms and Limits of Adjudication*, 92 HARV. L. REV. 353, 387 (1978).

[283] Schauer, *supra* note 276, at 636.

[284] Antonin Scalia, *Judicial Deference to Agency Interpretations of Law*, 1989 DUKE L.J. 511, 516; *see also* ROBERT A. KATZMANN, JUDGING STATUTES 15-22 (2014) (describing deficits of awareness, agreement, foresight, precision drafting, and care as typical sources of legislative ambiguity).

[285] David Enoch has persuasively argued that the state's action ought to be evaluated solely on the basis of their foreseeable (positive or negative) consequences, rather than on their intended, or reasoned, ends. David Enoch, *Intending, Foreseeing, and the State*, 13 LEG. THEORY 69, 91-92 (2007) (grounding this conclusion on a comparison of individual and

But consider a narrower version of the argument from reasoned decision-making. Perhaps when the state engages in certain coercive actions—including the criminal justice and social welfare decision-making that now employs algorithmic tools—it cannot act on a "mere hunch."[286] In the narrower category of cases, officials must have sound cause for their actions. Even in these cases, the impression that machine decisions are not, or cannot be, reasoned is a misleading and incomplete one. For it is not the case that machine decisions are bereft of justifying grounds. It is rather that those reasons are supplied at a point in time removed from state action impinging upon the individual. Recall that the design, testing, and implementation of machine learning tools are all thoroughly imbricated with purposeful human choice and intentionality.[287] Human intentions guide the choice between supervised and unsupervised models; the process of feature selections; the choice of training data; and the ongoing process of refinement and calibration toward an optimal classifier.[288] Much of this intentional human action is necessarily oriented by an understanding of the ends that the machine will serve. The dearth of reasoned judgments in machine decisions, therefore, is something of an optical illusion. It is not so much that such judgments are wanting. Rather, they have already been embedded into a classifier by the time that an algorithm is working in the world. These encoded judgments, moreover, serve the same ends as a demand for ex post reason-giving: They function as a precommitment to generality and as a safeguard against personalistic or arbitrary state action.[289] Given their formalized–literally calcified as code—nature, the reasons embedded in algorithms may be more resilient to ex post manipulation than the reasons upon which judgments by courts are anchored.

In sum, even if we stand on solid ground when we demand reasoned state action—especially when it comes to the deprivation of important human interests—our demand need not end in a right to a human decision. To the contrary, the concerns underlying the demand might well point in the other direction: A robustly designed algorithm thoughtfully supplied with unbiased and illuminating training data.

2. *The Right to an Individualized Decision*

A right to a human decision, alternatively, might be justified by pointing to the general character of the grounds upon which an algorithm relies when it reaches a classification decision. Roughly stated, the intuition here is that the state should take action against a person solely on the basis of their own behavior or merits. It should treat them, that is, "as an individual."[290] Action based on traits shared by a larger social group ipso facto fails to take that person seriously as an individual.

---

state responsibility). It may be a fair implication of his account is that the reasoned quality vel non of state action is irrelevant.

[286] *Arvizu,* 534 U.S. at 274. *But see* Craig S. Lerner, *Judges Policing Hunches*, 4 J.L. ECON. & POL'Y 25, 25 (2007) (defending hunches as "indispensable heuristic devices that allow people to process diffuse, complex information about their environment and make sense of the world").

[287] *See supra* text accompanying notes 174 to 194.

[288] On this iterative process of algorithmic improvement, see Bengio, *supra* note 33, at 50-51; Amershi et al., *supra* note 182, at 106

[289] *Cf.* Schauer, *supra* note 276, at 651-52 (describing reason-giving as a precommitment mechanism that yields generality).

[290] Kasper Lippert-Rasmussen, '*We are all Different': Statistical Discrimination and the Right to be Treated as an Individual*, 15 J. ETHICS 47, 49 (2011).

Something like this concern with the generality of justificatory grounds is implicit in GDPR Article 22. The latter picks out "profiling" as a form of automated processing.[291] The regulation elsewhere defines profiling broadly as "the use of personal data … to predict aspects concerning [a] natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behavior, location, or movement."[292] The breadth of this definition, and its connection in Article 22 to automated processing, imply a concern with algorithmic systems that ingest large volumes of data in order to generate predictive classifications. The implicit distinction drawn between automated and non-automated profiles, moreover, suggests that the GDPR's intervention is predicated on concern about the impersonal generality, and correlative detachment from individual particulars, of certain machine decisions. A parallel thought can be detected in courts' skepticism about the use of certain kinds of statistical evidence to demonstrate the likelihood of defendant responsibility in tort cases.[293] Here there issues a call for "individualized evidence."[294] Both it and demand for a human decision rest on a call for a particularized (rather than a population-wide) evidentiary basis for state action.

The intuition of a right to a human decision based on a demand to be treated as an individual can be justified by appeal to a number of philosophical traditions. It might be warranted, first, by the Kantian notion that an individual cannot be treated as a 'mere' means to an end.[295] The German Constitutional Court, for example, has invoked the impermissibility of state action that "verdinglicht und zugleich entretlicht" ("reifies and then disenfranchises") to invalidate a provision in the 2004 Air Transport Security Act that allowed a hijacked plane to be shot down under certain circumstances under the German Basic Law's commitment to human dignity.[296] Alternatively, this demand might be linked to the luck egalitarian demand that one aim to "eliminate so far as is possible the impact on people's lives of bad luck that falls on them through no fault or choice of their own."[297] Because machine decisions often rely on traits over which a person has no control, it would fall afoul of this demand. The latter ground, however, confronts substantial difficulties given the mismatch between

[291] GDPR art. 22. This Article does not prohibit profiling: It prohibits certain "decisions" based on profiling.

[292] GDPR art. 4(4).

[293] Courts have sometimes said that merely "mathematical chances," Smith v. Rapid Transit Inc., 317 Mass. 469, 470 (1945), or "quantitative probability," Day v. Boston & Me. R.R., 96 Me. 207, 217 (1902), are never sufficient for the imposition of tort liability. For a survey of different jurisdictions' responses, see Rebecca Haw, Note, *Prediction Markets and Law: A Skeptical Account*, 122 HARV. L. REV. 1217, 1229 (2009).

[294] Judith Jarvis Thompson, *Liability and Individualized Evidence*, 49 L. & CONT. PROB. 199, 205 (1996)

[295] I. KANT, FOUNDATIONS OF THE METAPHYSICS OF MORALS 47 (L. Beck trans. 1959) (1st ed. Riga 1785) ("Act so that you treat humanity, whether in your own person or that of another as, always as an end and never as a means only.") This Kantian notion, it should be noted, is malleable enough that it has also been put to work to justify the right to bare participation. Edward Pincoffs, *Due Process, Fraternity, and a Kantian Injunction*, 18 NOMOS 172, 179 (1977) ("[P]articipation is morally valuable to the degree that it makes determinate the moral principle that we should never treat a man as a mere means."). On the wide range of interpretations of this version of the categorical imperative, see Thomas E. Hill, Jr., *Humanity as an End in Itself*, 91 ETHICS 84, 84 (1980); see also Alexander Somek, *German Legal Philosophy and Theory in the Nineteenth and Twentieth Centuries*, *in* A COMPANION TO PHILOSOPHY OF LAW AND LEGAL THEORY 343, 343-43 (Dennis Patterson ed., 1996) (situating this idea in the history of German legal theory).

[296] Bundesverfassungsgericht (BverfG – Federal Constitutional Court), 59 Neue Juristiche Wochenschrift 751m at C II b) aa) and bb) (2006).

[297] Richard J. Arneson, *Luck egalitarianism and prioritarianism* 110 ETHICS 339, 339 (2000).

many criteria of social treatment and individual choice, as well as the difficulty of disentangling unchosen traits from those over which choice has been exercised.[298]

At first blush, it is not at all clear why algorithmic decisions should be singled out as failing to individuate. Algorithms can be designed to take account of "all relevant information, statistical or non-statistical" that is "reasonably available."[299] Nor is differentiated treatment of individuals based on their different traits and behaviors always a moral wrong. To the contrary, a uniform rule that imposes on each an equal "share in the cost of maintaining and preserving" the common good—think of a general draft for the military—will often be morally compelling.[300] Indeed, "even good judgment" is often predicated on nonspurious generalizations.[301] Hence, the bare claim that a decision is morally flawed because it is predicated on population-wide data rather than individualized evidence is simply not tenable.

I think a more subtle approach is necessary to make sense of this argument. Although I am not convinced that it yields a general objection to machine decisions, I think that with certain assumptions and under certain conditions, it can be deployed to resist specific substitutions of machine for human decisions. What issues is much narrower, that is, than the GDPR regime. To motivate this more fine-grained argument, it is necessary to assume that the human decision-maker will have access to individualized evidence, whereas the machine decision-maker would have access only to statistical, or population-wide information. Notice that there is nothing that compels this division of epistemic labor; a machine might be supplied with individualized evidence, while a human decision-maker might rely on statistical evidence. The assumption, however, seems to be baked into a right to a human decision as illuminated by the anti-"profiling" direction in GDPR Article 22.

With this assumption in hand, we can distinguish between different kinds of machine decisions. Where the decision is a prediction, there is no obvious objection to reliance on non-individualized evidence. If population-wide evidence is sufficient, say, to impose seatbelt mandates for automobiles[302] or vaccine regimes for school-age children,[303] why should it be taken as inferior as a basis for more granular state actions that are predictive in nature, such as bail and parole decisions? The absolute epistemic quality of different kinds of evidence cannot be a basis for distinction. Individualized evidence and population-wide evidence both vary in quality. There is no a priori reason to think decisions based on one will be less accurate than decisions based on the other.[304]

---

[298] Both problems are delineated in Samuel Scheffler, *What is egalitarianism?*, 31 PHIL & PUB. AFFAIRS 5, 17-21 (2003); *see also* Elizabeth S. Anderson, *What is the Point of Equality?*, 109 ETHICS 287, 289 (1999) (developing three further critiques of luck egalitarian).

[299] Lippert-Rasmussen, *supra* note 290, at 54.

[300] Annabelle Lever, *Why racial profiling is hard to justify: A response to Risse and Zeckhauser*, 33 PHIL & PUB. AFF. 94, 110 (2005).

[301] FREDERICK SCHAUER, PROFILES, PROBABILITIES AND STEREOTYPES 215 (2003).

[302] *See, e.g.,* 1984 N.J. Laws c. 179 § 2, last amended by 2009 N.J. Laws c. 318 § 1.

[303] For instance, both the District of Colombia and Virginia mandate by statute the Human Papillomavirus vaccine for school-age girls. *See* D.C. Code § 7-1651.04 (b)(1)(B)(iii) (2008); Va. Code Ann. § 32.1-46 (D)(3) (2008).

[304] *See* Thompson, *supra* note 294, at 200-02 (setting forth an example).

But when population-based evidence is used to substantiate a matter of historical fact for the purpose of assigning responsibility, subtly different considerations emerge. A problem arises particularly when the purpose of the decision is to generate a deterrence effect in the future. In a phrase developed by the philosopher Martin Smith, there are many instances in which individualized evidence "normically supports" the conclusion for which it is proffered,[305] and hence is capable of generating a deterrence effect. To see Smith's point, imagine I have a laptop with a screensaver that shows a blue screen nine-tenths of the time. While I am out at work, my friend walks past my computer and sees a blue screen. My friend's belief that the screen is blue is "normically supported" by her perception; my analog belief that the screen was blue is not.[306] This can be stated in another way. My friend's belief that the screen is blue is *counterfactually sensitive* to the truth whereas my evidence is not.[307] If we learn later that the screen was not blue at that moment, I might simply shrug about my unlucky guess. For my friend, such indifference would seem "out of place."[308] She should, perhaps, have her vision checked for colorblindness.

This distinction can be transposed to the legal context in the following way. If employed as a basis for adverse action, my friend's evidence is sensitive to historical facts in a way my evidence is not. This suggests that whereas both kinds of evidence can be rationally employed to form beliefs and predictions, only counterfactually sensitive evidence can be used to generate a deterrence effect.[309] Where liability is imposed on the basis of insensitive grounds (i.e., statistical evidence), it will not deter. Hence, sensitivity matters for deterrence, even if it does not matter for prediction or perhaps knowledge.[310] When a machine decision relies on population-wide evidence, therefore, there is a loss of deterrence effect.

So it may be that the objection to the use of non-individualized evidence comes down to a demand for optimal deterrence, and also perhaps to our social practices of blaming.[311] But if evidence that is not individualized improves accuracy, while failing to create desirable incentives, why should that be the basis of an *individual's* objection? It is the state, not the regulated individual, who has the interest in deterrence. Moreover, recall that this line of argument is also premised on the (probably flawed) assumption that machines only rely on population-wide evidence, whereas human decisionmakers always have access to individualized evidence. This line of argument, finally, is not enough to explain a general right to a human decision given the manner in which algorithmic decisions are presently employed. Most of the present uses of machine learning by the state involve predictions, rather than findings of historical fact upon which deterrence is based.[312] Not all concern blame, and if

---

305 Martin Smith, *What Else Justification Could Be*, 44 NOÛS 10, 13-14 (2010).

306 *Id.* (offering a more complex version of this hypothetical).

307 David Enoch, Levi Spectre & Talia Fisher, *Statistical Evidence, Sensitivity, and the Legal Value of Knowledge*, 40 PHIL & PUB. AFF. 197, 209-10 (2012).

308 *Id.* at 209.

309 *Id.* at 218-19.

310 There is a literature on whether probabilistic evidence can be a basis for knowledge or rational belief. *See, e.g.,* HENRY KYBERG, PROBABILITY AND THE LOGIC OF RATIONAL BELIEF (1961). I do not think that form of strong skepticism has salience to the questions of legal and institutional design here.

311 Enoch, Spectre, & Fisher, *supra* note 307, at 215 (noting that blame may also require counterfactually sensitive evidence).

312 *See supra* text accompanying notes 133 to 134.

the allocation of blame is viewed as the central function of a decision tool, then machines may be just as inapt as for causal questions. For all these reasons, I am skeptical that a concern with counterfactual sensitivity can redeem the right to a human decision.

*   *   *

The classifier-facing grounds for a right to a human decision, in short, fare no better than arguments that begin with individuals' rights. Neither a worry about reasoned state action nor a concern with the individuated character of evidence upon which state action rests proves satisfying.

D. Systemic Concerns

A final potential ground upon which the right to a human decision might be defended dilates the analytic lens beyond the immediate transaction between an individual and a machine to consider the dynamic consequences of exclusive reliance on machine decisions on wider patterns of state action. Although a greater number of human decision may have desirable systemic consequences—discussed below—these could only with difficulty be used to sustain a free-standing individual *right*. Rather, all these interests might more precisely be targeted and advanced through alternative interventions that do not rely on the happenstance of individuals exercising discretion over whether, or if, to press their legal interests. While the system concerns identified here might thus provide collateral support for the right at issue, they cannot plausibly work as its principal buttresses.

First, the enforcement of legal rights against state actors is commonly associated with "decreased activity levels" close to judicially enforced threshold of liability.[313] A right to a human decision would be no different. Exclusive reliance on machine decisions lowers the marginal cost of exercising a given state power. The right can hence be thought of as a kind of enervating friction on state action. But of course, whether this is desirable will obviously depend on the nature of the activity. For example, consider the possibility of fully automating unmanned drone plans capable of exercising deadly force on a distant battlefield.[314] Even to sympathetic eyes, a thorough excision of the human role in warfare raises complex ethical issues.[315] One of these might focus on the effect of maintaining a human role on activity levels. If it were the case that maintaining a human role led to a lower activity level, without preclusive costs to a war effort, one might plausibly speak of an obligation to maintain a human in the loop as a way to forestall the rapid inflation of lethal drone use.

Second, several commentators have worried about "automation bias," or the "use of automation as a heuristic replacement for vigilant information seeking and processing."[316] In effect, humans adopt a heuristic of reliance upon automated decisions "as a replacement for more vigilant

---

[313] John C. Jeffries, Jr., *The Right-Remedy Gap in Constitutional Law*, 109 YALE L.J. 87, 105 (1999). The effect of the liability rule is disputed. Conventional wisdom holds that both negligence and strict liability regimes are associated with a risk of excessive activity. STEVEN SHAVELL, ECONOMIC ANALYSIS OF ACCIDENT LAW 66-71 (1987).

[314] For an acute description, see HUGH GUSTAFSON, DRONES: REMOTE-CONTROL WARFARE 2-25 (2017).

[315] *See* Robert Sparrow, *Robots and Respect: Assessing the Case Against Autonomous Weapon Systems*, 30 ETHICS & INT'L AFF. 93, 94-95 (2016) (summarizing key ethical questions).

[316] Linda J. Skitka et al., *Automation Bias and Errors: Are Crews Better Than Individuals?*, 10 INT'L J. AVIATION PSYCHOLOGY 85, 86 (2000).

system monitoring or decision making."[317] The right to a human decision works as a prophylactic against the possibility that humans will place excessive faith in machine decisions because of the veneer of expertise and objectivity that infuses that technology.[318] However forceful this concern may be—empirical evidence is thin on the ground—there are likely a number of ways to ensure against complacent reliance on automated decisionmakers—not least some kind of frequent auditing. Reliance on the individuals subject to classification may be one of a range of solutions, but it is hardly an inevitable design choice.

A third argument for a right to a human decision focuses on the effects of machine decisions on the distribution of social power. Machine learning allows for gains to social welfare as a result of new or more accurate predictions. But these gains might be unevenly distributed in ways that trigger deep normative concern. Because machine-learning tools requires large pools of data and robust computational resources, it is likely that they will be adopted and used by organizational entities, not least the state, that already have asymmetrical relationships with the public at large. Adoption of machine learning might exacerbate these imbalances in undesirable ways. This raises the possibility that unease concerning machine decisions rests not on their distinctive quality, but on their effects upon the relationship between the state and its subjects, or large corporations and individual market participants. Asymmetrical distributions of such power might undermine the possibility conditions of participatory democracy, if machine decisions are used to shape political preferences. Alternatively, they might enable new, highly intrusive forms of regulation consistent with some normatively well-grounded account of individual liberty.

I am sympathetic to these concerns. But I am skeptical that an individual right provides a meaningful response given the technological realities and normative implications mapped in Parts II and earlier in this Part. A right to a human decision is a response to technologies that generate troublesome asymmetries between persons and concentrated organizational power. The problem is that it requires heroic assumptions to conclude that dispersed individuals—vulnerable to state or corporate pressure along multiple margins—will be capable of using a right to engage in collective action that effectually redresses asymmetrical social arrangements. That is, the mere provision of rights does not resolve the asymmetry of power. This much is evident from a half-century of experience with procedural entitlements in the criminal justice context, which suggests that rights' efficacy is tightly constrained by the ability of the state (or similarly regulated actor) to find substitute vectors of influence.[319] Recent experience with individual entitlements to privacy in the social media and internet platform contexts also furnishes cause for pessimism.[320] A central problem with consent-based privacy regimes is that consumers seem to place different values on that good depending on whether they were asked to consider how much money they would accept to disclose otherwise private information

---

[317] Linda J. Skitka, *Does Automation Bias Decision-making?*, 51 INT. J. HUMAN-COMPUTER STUD. 991, 992 (1999).

[318] A recent example of automation bias arguably having catastrophic results was an oil-pipeline breach in Marshall, Michigan, in 2010. David Wesley & Luis Alfonso Dau, *Complacency and Automation Bias in the Enbridge Pipeline Disaster*, 25 ERGONOMICS IN DESIGN 17, 19-20 (2017).

[319] The classic statement of this concern is William J Stuntz, *The Uneasy Relationship Between Criminal Procedure and Criminal Justice*, 107 YALE L. J. 1, 64 (1997).

[320] *See, e.g.,* Alessandro Acquisti, Leslie K. John & George Loewenstein, *What is privacy worth?*, 42 J. LEG. STUD. 249, 250-51 (2013) (identifying sensitivity of privacy-related preferences to subtle contextual cues).

or how much they would pay to protect otherwise public information.[321] They also appear to have time-inconsistent preferences, in the sense that they are willing to accept low rewards now in exchange for a "possibility permanent negative annuity in the future."[322] Rights, in short, often (although not always) fall short of catalyzing reallocations of power.

If rights are not necessarily the best instruments to challenge concentrated social power, at least when conceptualized as a stand-alone instrument, is there an alternative? A more direct approach entails a frontal attack on asymmetries of power or influence by fragmenting the extant concentrations of social power.[323] One might also think of ways to redistribute the surplus that results from aggregated epistemic authority. But absent evidence that the right to a human decision can facilitate this sort of mobilization—and I don't think such evidence exists—we should not precipitously conclude that an individual right provides an effectual solution to a structural and systemic dysfunction.

Finally, and related to the concern about power, a human decision may be preferred to a human decision in order to pursue an institutional design goal such as the diffusion of state authority or the continued evolution of legal rules. In effect, the claim would be that human decisional authority has certain positive spillovers beyond the individual case. An argument of this kind might justify localized substitution of human for machine decision-making. But it would not provide a global reason for such substitution. Indeed, even its local application would hinge on other elements of institutional design.

A human decision might be preferred, for example, because it maintains the open-endedness of legal criterion, or because it injected an element of uncertainty into law's implementation. A supervised machine learning tool's goals must be fully specified in order to be implemented. A resort to human decision-making allows for an under-specification of those ends. Seemingly problematic from the perspective of legality, such under-specification of law's ends might be desirable under certain circumstances. It might, for example, allow law's dynamic updating over time. It might also constitute a limitation on the authority of law-makers to fully define law by preserving a redoubt of free-wheeling discretionary judgment by a back-end human decision-maker. An institutional perch for revision and second-guessing of authority has a particular attraction as an element of a liberal constitutional democracy. Of course, there is no reason why the institutionalization of corrigibility needs to take the form of a human decision. It is also possible to imagine a similar goal being pursued through a public system of audits focusing on the systemic operation of the algorithm. Something of the kind can also be installed elsewhere in a political system, rendering a right to a human decision nugatory. Alternatively, a specific machine decision tool might be rejected on the ground that it will impede the dynamic development of new legal rules. A machine learning tool, that is, might refine its classification rule through reinforcement learning over time, but this will not necessarily yield detailed new legal guidance for primary conduct. Again, this argument for human decision-making is

---

321 Alessandro Acquisti, Leslie K. John, and George Loewenstein, *What is privacy worth?*, 42 J. LEG. STUD. 249, 249-51 (2013).

322 Alessandro Acquisti et al., *The Economics of Privacy*, 54 J. ECON. LITERATURE 442, 442-43 (2016). For similar results, see Kirsten Martin, *Privacy Notices as Tabula Rasa: An Empirical Investigation into How Complying with a Privacy Notice Is Related to Meeting Privacy Expectations Online*, 34 J. Pub. Pol'y & Marketing 210 (2015).

323 For an argument along these lines with respect to corporate power, see TIM WU, THE CURSE OF BIGNESS: ANTITRUST IN THE NEW GILDED AGE (2018).

contingent on the absence of other platforms for refinement and publication of new, more detailed rules for primary conduct.

None of these systemic concerns, in short, is sufficient to motivate a freestanding individual right. At best, a human decision in lieu of a machine decision is a useful, although not essential instrument of institutional design for the production of positive spillovers. Whether that substitution is desirable, in short, will depend on other elements of institutional design. It is poorly described as a 'right.'

E. The 'Right' to a Human Decision *Dubitante*

Fully automated decision-making provokes unease in many minds.[324] And a taxonomy of the potential justifications for a right to a human decision reveals a range of variegated normative considerations. Some can be loosely ranked under the labels 'autonomy' and 'dignity.' Some concern the opportunity to proffer reasons, while others turn on the reasoned nature of state action. Yet others concern the balance of power between individuals and either the state or corporate actors. Careful analysis of all of these reasons, however, suggests that none can justify a right to a human decision. In some instances, such as the right to bare participation, the absence of any recognition of an analog right in cognate fields of law seems preclusive. In other instances, including the interest in proffering reasons and in receiving a reasoned decision, the argument for the right is predicated on doctrinal and technological circumstances that will only occasionally, and only by happenstance, be satisfied. As a result, the right to individualized consideration is more ephemeral than appears at first blush; to the extent that it stands on firm normative ground, it probably does not apply to most practical state applications of machine-learning technology.

There are many reasons to be cautious about the avulsive advance of new algorithmic technologies, ranging from their effects on the labor market to their propagation of historical bias and reinforcement of social stratification. Normative and legal concern, however, should not be directed at the articulation or enforcement of a right to a human decision akin to that found in GDPR art. 22. At least for now, it may well be a game that is simply not worth the candle.

Conclusion: A Right to a Well-Calibrated Machine Decision?

The aim of this paper has been to explore potential justifications for a right to a human decision. Based on that survey of the normative and technological terrain, I have rejected that right as an attractive general amendment to the law. Although I cannot rule out the desirability of such a right as applied to the state (the main object of my analysis) in all circumstances, my analysis undermines the plausibility of the right to a human decision wrought as a general, across-the-board regime.

In concluding, let me offer what I hope is a provocative, albeit tentative, thought: Rather than thinking about a right to a human decision, might we be better off limning a right to a well-calibrated machine decision? Glimpses of such a right's foundations can be caught, scattered across the previous analysis: I can only gather them here briefly as a way of stimulating reflection on the possibility.

---

[324] Binns et al., *supra* note 7, at 385.

An account of the right to a machine decision would begin with the observation that under most likely conditions, a well-designed machine-learning tools is both more accurate and less likely to discriminate on invidious grounds than human decision-makers. But many of the algorithmic instruments now implemented by government are highly flawed.[325] Yet this does not impel a reversion to equally flawed human decision-making. Rather, legal rules should incentivize the correction of such errors. Its dynamic goal should be a better machine decision. To the extent that Due Process interests are justified by the interest in an accurate decisional process, they arguably point in favor of more and better, not less or worse, automation. Human supplements, indeed, might be a form of baleful "undue process,"[326] which induces a sense of psychological comfort at the price of an increase in false positives and negatives. Provided that an algorithmic tool is well-calibrated in the sense of not relying on flawed training data, and not employing a predictive tool that reinforces pernicious forms of stratification, that tool may well perform better than a human being. Hence, a better right to formulate into law would entitle individuals to a robustly designed algorithm aimed at a legitimate goal, which is supplied with unbiased and illuminating training data, and then effectively monitored to prevent unwarranted results from emerging.[327]

It is also possible that well-calibrated machine decision-makers will have underappreciated advantages that sound in dignitary and autonomy terms. Consider the possibility of dignitary gains from such a right. Algorithmic therapists such as 'Woebot,' for example, now interact with between one to two million people online; apart from being free, the algorithmic tool is "easier to talk to" because users "don't feel judged."[328] The same point might be made by pointing to dating algorithms, which occupy an increasing share of the matchmaking market and which squeeze out family and friends as intermediaries; depersonalization might facilitate new possibilities and avoid certain forms of humiliation.[329] It is quite possible that an algorithmic interface for welfare recipients –who are a notoriously stigmatized group[330]—might also reduce the psychological costs attendant on those benefits. These examples are intended to be suggestive rather than conclusive. But they point to ways in which new technologies, strategically deployed, might mitigate inequality and enable humanity—rather than the reverse.

In sum, algorithmic technologies used by machine decisions are still in their infancy. And they can be flawed in many ways. It seems too early, however, to assume that human decisions will be globally superior to machine decisions such that a right to the former is warranted. Sometimes the opposite might be true. We should, therefore, at least consider the possibility that under certain circumstances a right to a well-calibrated machine decision might be the better option.

---

[325] AI Now, *supra* note 39, at 18-22.

[326] Samaha, *supra* note 254, at 630.

[327] Obviously, this compressed definition packs in a lot of detail, which I hope to unpack in later work.

[328] Clive Thompson, *May A.I. Help You?*, N.Y. TIMES, Nov. 14, 2018.

[329] *How the Internet has Changed Dating*, THE ECONOMIST, August 18, 2018, https://www.economist.com/briefing/2018/08/18/how-the-internet-has-changed-dating.

[330] For a classic treatment, see Joel F. Handler & Ellen Jane Hollingsworth, *Stigma, privacy, and other attitudes of welfare recipients*, 22 STAN. L. REV. 1, 4-5 (1969) (documenting experienced stigma).