A RIGHT TO A HUMAN DECISION

Aziz Z. Huq[*]

Recent advances in computational technologies have spurred anxiety about a shift of power from human to machine decision makers. From welfare and employment to bail and other risk assessments, state actors increasingly lean on machine-learning tools to directly allocate goods and coercion among individuals. Machine-learning tools are perceived to be eclipsing, even extinguishing, human agency in ways that compromise important individual interests. An emerging legal response to such worries is to assert a novel right to a human decision. European law embraced the idea in the General Data Protection Regulation. American law, especially in the criminal justice domain, is moving in the same direction. But no jurisdiction has defined with precision what that right entails, furnished a clear justification for its creation, or defined its appropriate domain.

This Article investigates the legal possibilities and normative appeal of a right to a human decision. I begin by sketching its conditions of technological plausibility. This requires the specification of both a feasible domain of machine decisions and the margins along which machine decisions are distinct from human ones. With this technological accounting in hand, I analyze the normative stakes of a right to a human decision. I consider four potential normative justifications: (a) a concern with population-wide accuracy; (b) a grounding in individual subjects' interests in participation and reason giving; (c) arguments about the insufficiently reasoned or individuated quality of state action; and (d) objections grounded in negative externalities. None of these yields a general justification for a right to a human decision. Instead of being derived from normative first

[*] 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, Lior Strahilevitz, Rebecca Wexler, and Annette Zimmermann for thoughtful conversation. Workshop participants at the University of Chicago, Stanford Law School, the University of Houston, William and Mary Law School, and Syracuse University School of Law also provided thoughtful feedback. I am grateful to Christiana Zgourides, Erin Brown, and the other law review editors for their careful work on this Article. All errors are mine, not the machine's.

principles, limits to machine decision making are appropriately found in the technical constraints on predictive instruments. Within that domain, concerns about due process, privacy, and discrimination in machine decisions are typically best addressed through a justiciable "right to a well-calibrated machine decision."

INTRODUCTION

Every tectonic technological change—from the first grain domesticated to the first smartphone set abuzz[1]—begets a new society. Among the ensuing birth pangs 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 human 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

[1] For recent treatments of these technological causes of social transformations, see generally James C. Scott, Against the Grain: A Deep History of the Earliest States (2017), 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 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 5 (3d ed. 2009) (defining an algorithm as "any well-defined computational procedure that takes some value, or set of values, as input and produces some value, or set of values, as output" (emphasis 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 1–4 (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 2–3 (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, 355 Science 483, 483 (2017) (noting the use of machine learning to solve prediction problems). I discuss the technological scope of the project, and define relevant terms, infra at text accompanying note 111. I will use the terms "algorithmic tools" and "machine learning" interchangeably, even though the class of algorithms is technically much larger.

[4] Kartik Hosanagar & Vivian Jair, We Need Transparency in Algorithms, But Too Much Can Backfire, Harv. Bus. Rev. (July 23, 2018), https://hbr.org/2018/07/we-need-transparency-in-algorithms-but-too-much-can-backfire [https://perma.cc/7KQ9-QMF3]; accord Cary Coglianese & David Lehr, Regulating by Robot: Administrative Decision Making in the Machine-Learning Era, 105 Geo. L.J. 1147, 1149 (2017).

[5] Shoshana Zuboff, Big Other: Surveillance Capitalism and the Prospects of an Information Civilization, 30 J. Info. Tech. 75, 75 (2015) (describing a "new form of information capitalism

that are beyond individuals' ken or control.[6] Many 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 an important decision on the basis of past data seem[s] unfair."[7] Machines, it is said, want fatally for "empathy."[8] For others, machine decisions seem dangerously inscrutable, non-transparent, and so 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.[10] Even the basic terms of political choice seem compromised.[11] At the same time that machine learning is poised to recalibrate the ordinary forms 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.[12] A fearful future looms, one

[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, 558 Nature 357, 357 (2018) (documenting concerns among the public that algorithmic risk scores for detecting child abuse fail to account for an "effort . . . to turn [a] life around").

[7] Reuben Binns et al., 'It's Reducing a Human Being to a Percentage'; Perceptions of Justice in Algorithmic Decisions, 2018 CHI Conf. on Hum. Factors Computing Systems 9 (emphasis omitted).

[8] Virginia Eubanks, Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor 168 (2017).

[9] Will Knight, The Dark Secret at the Heart of AI, MIT Tech. Rev. (Apr. 11, 2017), https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/ [https://perma.cc/L94L-LYTJ] ("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] For consideration of these issues, see Mariano-Florentino Cuéllar & Aziz Z. Huq, Economies of Surveillance, 133 Harv. L. Rev. 1280 (2020), and Mariano-Florentino Cuéllar & Aziz Z. Huq, Privacy's Political Economy and the State of Machine Learning: An Essay in Honor of Stephen J. Schulhofer, N.Y.U. Ann. Surv. Am. L. (forthcoming 2020).

[11] See, e.g., Daniel Kreiss & Shannon C. McGregor, Technology Firms Shape Political Communication: The Work of Microsoft, Facebook, Twitter, and Google with Campaigns During the 2016 U.S. Presidential Cycle, 35 Pol. Comm. 155, 156–57 (2018) (describing the role of technology firms in shaping campaigns).

[12] 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, 1:00 AM), https://www.theguardian.com/commentisfree/2018/feb/01/robots-take-our-jobs-amazon-go-seattle [https://perma.cc/2CFP-3JJV]. For a more nuanced account, see Martin Ford, Rise of the Robots: Technology and the Threat of a Jobless Future 282–83 (2015).

characterized by massive economic dislocation, wherein people have lost control of 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 an individual being assigned a benefit or a coercive intervention has a right to a human decision rather than a decision reached by a purely automated process (a "machine decision"). European law has embraced the idea. American law, especially in the criminal justice domain, is flirting with it.[13] 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. My focus is not on the form of such a right—statutory, constitutional, or treaty-based—or how it is implemented—say, in terms of liability or property rule protection—but more simply on what might ab initio justify its creation.

To motivate this inquiry, consider some of the anxieties unfurling already in public debate: A nursing union, for instance, launched a campaign urging patients to demand human medical judgments rather than technological assessment.[14] And a majority of patients surveyed in a 2018 Accenture survey preferred treatment by a doctor in person to virtual care.[15] When California proposed replacing money bail with a "risk-based pretrial assessment" tool, a state court judge warned that "[t]echnology cannot replace the depth of judicial knowledge, experience, and expertise in law enforcement that prosecutors and defendants' attorneys possess."[16] In 2018, the City of Flint, Michigan, discontinued the use of a highly effective machine-learning tool designed to identify defective water pipes, reverting under community pressure to human decision making

[13] See infra text accompanying notes 70–73.

[14] 'When It Matters Most, Insist on a Registered Nurse,' Nat'l Nurses United, https://www.-nationalnursesunited.org/insist-registered-nurse [https://perma.cc/MB66-XTXW] (last visited Jan. 19, 2020).

[15] Accenture Consulting, 2018 Consumer Survey on Digital Health: US Results 9 (2018), https://www.accenture.com/_acnmedia/PDF-71/Accenture-Health-2018-Consumer-Survey-Digital-Health.pdf#zoom=50 [https://perma.cc/TU5F-9J82].

[16] Quentin L. Kopp, Replacing Judges with Computers Is Risky, Harv. L. Rev. Blog (Feb. 20, 2018), https://blog.harvardlawreview.org/replacing-judges-with-computers-is-risky/ [https://perma.cc/WS5S-ARVF]. On the current state of affairs, see California Set to Greatly Expand Controversial Pretrial Risk Assessments, Filter (Aug. 7, 2019), https://filtermag.org/-california-slated-to-greatly-expand-controversial-pretrial-risk-assessments/ [https://perma.cc-/2FNX-U3C9].

with a far lower hit rate for detecting defective pipes.[17] Finally, and perhaps most powerfully, consider the worry congealed in an anecdote told by data scientist Cathy O'Neil: An Arkansas woman named Catherine Taylor is denied federal housing assistance because she fails an automated, "webcrawling[,] data-gathering" background check.[18] 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.[19] O'Neil's plainly troubling anecdote powerfully captures the fear that machines will be unfair, incomprehensive, or incompatible with the flexing of elementary human agency: it provides a sharp spur to the inquiry that follows.

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 a General Data Protection Regulation (GDPR).[20] Unlike the legal regime it superseded,[21] 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).[22] Hence, it can be directly enforced in court through hefty financial penalties.[23] Article 22 of the GDPR endows a natural individual 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."[24] That right covers private

[17] Alexis C. Madrigal, How a Feel-Good AI Story Went Wrong in Flint, Atlantic (Jan. 3, 2019), https://www.theatlantic.com/technology/archive/2019/01/how-machine-learning-found-flints-lead-pipes/578692/ [https://perma.cc/V8VA-F22W].

[18] Cathy O'Neil, Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy 152–53 (2016).

[19] Id. at 153.

[20] Regulation 2016/679, of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L 119) (EU) [hereinafter GDPR]; see also Christina Tikkinen-Piri, Anna Rohunen & Jouni Markkula, EU General Data Protection Regulation: Changes and Implications for Personal Data Collecting Companies, 34 Computer L. & Security Rev. 134, 134–35 (2018) (documenting the enactment process of the GDPR).

[21] 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].

[22] Bryce Goodman & Seth Flaxman, European Union Regulations on Algorithmic Decision Making and a "Right to Explanation," AI Mag., Fall 2017, at 51–52 (explaining the difference between a non-binding directive and a legally binding regulation under European law).

[23] Id. at 52.

[24] GDPR, supra note 20, arts. 4(1), 22(1) (inter alia, defining "data subject").

and some (but not all) state entities.[25] On its face, it fashions an opt-out of quite general scope from automated decision making.[26]

The GDPR also has extraterritorial effect.[27] It reaches platforms, such as Google and Facebook, that offer services within the EU.[28] 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.[29] 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 complaints sounding in procedural justice and fairness terms.[30] Further, the Sixth Amendment's right to a jury trial has to date principally been deployed to resist *judicial* factfinding.[31] But there is no conceptual reason why the Sixth Amendment could not be invoked to preclude at least some forms of algorithmically generated inputs to criminal sentencing. Indeed, it would seem to follow a fortiori that a right precluding a jury's substitution with a judge would also block its displacement by a mere machine.

[25] See id. art. 4(7)–(8) (defining "controller" and "processor" as key scope terms). The Regulation, however, does not apply to criminal and security investigations. Id. art. 2(2)(d).

[26] 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 53–58.

[27] GDPR, supra note 20, art. 3.

[28] There is sharp divergence in the scholarship over the GDPR's extraterritorial scope, which ranges from the measured, see Griffin Drake, Note, 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.").

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

[30] 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 2 (May 23, 2016), https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing [https://perma.cc/Q9ZU-VY6J] (criticizing machine-learning instruments in the criminal justice context).

[31] 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))).

In this Article, I start by situating a right to a human decision in its contemporary technological milieu. I can thereby specify the feasible domain of machine decisions. I suggest this comprises decisions taken at high volume in which sufficient historical data exists to generate effective predictions. Importantly, this excludes many matters presently resolved through civil or criminal trials but sweeps in welfare determinations, hiring decisions, and predictive judgments in the criminal justice contexts of bail and sentencing. Second, I examine 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 the form of artificial intelligence diffusing most rapidly today.[32] A right to a human decision cannot be defined or evaluated without some sense of 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 distinctions between human and machine decisions are less crisp than might first appear. Claims about a right to human decision, 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 evaluate the right to a human decision in relation to four normative ends it might plausibly be understood to further. A first possibility turns on overall accuracy worries. My second line of analysis takes up the interests of an individual exposed to a machine decision. The most pertinent of these interests hinge upon an individual's participation in decision making and her opportunity to offer reasons. A third analytic salient tracks ways that a machine instrument might be intrinsically objectionable because it uses a deficient decisional protocol. I focus here on worries about the absence of individualized consideration and a machine's failure to offer reasoned

[32] See infra text accompanying note 88 (defining machine learning). I am not alone in this focus. Legal scholars are paying increasing attention to new algorithmic technologies. For leading examples, see 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 . . .)"); Andrew Guthrie Ferguson, Big Data and Predictive Reasonable Suspicion, 163 U. Pa. L. Rev. 327, 383–84 (2015) (discussing the possible use of algorithmic prediction in determining "reasonable suspicion" in criminal law); Kroll et al., supra note 2, at 636–37; 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).

judgments. Finally, I consider dynamic, system-level effects (i.e., negative spillovers), in particular in relation to social power. None of these arguments ultimately provides sure ground for a legal right to a human decision.

Rather, I suggest that the limits of machine decision making be plotted based on its technical constraints. Machines should not be used when there is no tractable parameter amenable to prediction. For example, if there is no good parameter that tracks job performance, then machine evaluation of those employees should be abandoned. Nor should they be used when decision making entails ethical or otherwise morally charged judgments. Most important, I suggest that machine decisions should be subject to *a right to a well-calibrated machine decision* that folds in due process, privacy, and equality values.[33] This is a better response than a right to a human decision to the many instruments now implemented by the government that are highly flawed.[34]

My analysis here focuses on state action that imposes benefits or coercion on individuals—and not on either private action or a broader array of state 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, the range of private uses of algorithmic tools is vast and heterogenous. Algorithms are now deployed in private activities ranging from Google's PageRank instrument,[35] to "fintech" applied to generate new revenue streams,[36] to medical instruments used to calculate stroke risk,[37] to engineers' identification of new stable

[33] A forthcoming companion piece develops a more detailed account of how this right would be vindicated in practice through a mix of litigation and regulation. See Aziz Z. Huq, Constitutional Rights in the Machine Learning State, 105 Cornell L. Rev. (forthcoming 2020).

[34] For a catalog, see Meredith Whittaker et al., AI Now Inst., AI Now Report 2018, at 18–22 (2018), https://ainowinstitute.org/AI_Now_2018_Report.pdf [https://perma.cc/2BCG-M4-54].

[35] See, e.g., David Segal, The Dirty Little Secrets of Search: Why One Retailer Kept Popping Up as No. 1, N.Y. Times, Feb. 13, 2011, at BU1.

[36] 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/DG8N-8NVS] (describing how fintech firms use artificial intelligence to improve investment strategies and analyze consumer financial activity).

[37] 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 Applied Stat. 1350, 1350 (2015).

inorganic compounds.[38] Algorithmic tools are also embedded within new applications, such as voice recognition software, translation software, and visual recognition systems.[39] In contrast, the state is to date an unimaginative user of machine learning, with a relatively constrained domain of deployments.[40] This makes for a more straightforward analysis. Third, where the state does use algorithmic tools, it often results directly or indirectly in deprivations of liberty, freedom of movement, bodily integrity, or basic income. These normatively freighted machine decisions present arguably the most compelling circumstances for adopting a right to a human decision and so are a useful focus of normative inquiry.

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 fanciful. Part II defines the class of computational tools to be considered, explores the manner in which such instruments can be used, and teases out how they are (or are not) distinct from human decisions. Doing so helps illuminate the plausible forms of a right to a human decision. 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 Article's analytic lens to gesture at the possibility that 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. jurisprudence to European directives, and to range across both private and public law domains capturing the regulation of state and nonstate action. I take this wide-angle view in this Part 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

[38] See, e.g., Paul Raccuglia et al., Machine-Learning-Assisted Materials Discovery Using Failed Experiments, 533 Nature 73, 73 (2016) (identifying new vanadium compounds).

[39] Yann LeCun et al., Deep Learning, 521 Nature 436, 438–41 (2015).

[40] See infra text accompanying notes 117–21 (describing state uses of machine learning).

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 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 Article 22 of the GDPR also cast some light on the difficulties of fashioning such a right and discerning its justifications.

1. Antecedents

An early antecedent of a right to a human decision is France's 1978 law on information technologies, datafiles, and civil liberties.[41] Article 2 of this law, as originally enacted, prohibited official use of automated profiling or personality screening.[42] So far as I can tell, this measure was never applied to machine learning, which, in any case, was not in general usage at the time. Seventeen years later, the European Parliament and Council promulgated the Data Protection Directive.[43] The latter lacked independent legal effect on individuals but obliged European Union member states to enact conforming laws. Article 15 of the Directive required member states to create a right "not to be subject to a decision which produces legal effects concerning him [or her] or significantly affects him [or her] and which is based solely on automated processing of data intended to evaluate certain personal aspects."[44] Countries implemented this provision in various ways, at times differentiating

[41] Loi 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés [Law 78-17 of January 6, 1978 on Information Technologies, Datafiles and Civil Liberties], Journal Officiel De La République Française [J.O.] [Official Gazette of France], Jan. 7, 1978, p. 227.

[42] "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. art. 2 ("No judicial decision involving an appraisal of human conduct may be based on any automatic processing of data which describes the profile or personality of the person concerned."). A second clause extended the same rule to administrative decisions. Id.

[43] Directive 95/46, supra note 21.

[44] Id. art. 15(1).

between private and public actors.[45] As with the 1978 French law, however, these measures appear to have had little effect on the ground.[46] There is also little evidence of the concerns that motivated Article 15's inclusion in the Data Protection Directive.[47] Accordingly, its history yields scant guidance 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."[48] 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 contains penalty provisions envisaging fines running into the millions of euros.[49] Its ninety-nine articles cover a broad array of other topics. But my discussion trains largely on Article 22.

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

[45] See, e.g., D.Lgs. 30 giugno 2003, n.196, in G.U. July 29, 2003, n.174 (It.), http://www.normattiva.it/uri-res/N2Ls?urn:nir:stato:decreto.legislativo:2003-06-30;196!vig= [https://perma.cc/7X3A-D23Q] (distinguishing public actors, and imposing an absolute prohibition on exclusively using automated data processing in relation to judicial proceedings, but creating a right to object to automated decisions by private actors).

[46] See Lee A. Bygrave, Minding the Machine: Article 15 of the EC Data Protection Directive and Automated Profiling, 17 Computer L. & Security Rev. 17, 21 (2001) (describing the right as a "house of cards"). Only one German decision seems to have touched on this right. Isak Mendoza & Lee A. Bygrave, The Right Not to Be Subject to Automated Decisions Based on Profiling, *in* EU Internet Law: Regulation and Enforcement 77, 87–88 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).

[47] Bygrave asserts that the Article was motivated by "the potential for . . . automatisation to diminish the role played by persons in shaping important decision-making processes." Bygrave, supra note 46, at 18. But this is a tautology, not an explanation of why machine decisions are to be disfavored in relation to human ones.

[48] European Commission Fact Sheet MEMO/17/1441, Questions and Answers—Data Protection Reform Package (May 24, 2017).

[49] Jan Philipp Albrecht, How the GDPR Will Change the World, 2 Eur. Data Protection L. Rev. 287, 287 (2016) (describing the sanctions regime).

significantly affects him or her."[50] According to the European Commission Data Protection Working Party created by the EU, Article 22(1) applies only if "there is no human involvement in the decision process."[51] Further, the Working Party guidance document suggests that "meaningful" ex post review of an automated decision would remove it from the scope of Article 22(1).[52] The GDPR does not explain how the quality of such review is to be assessed.[53] The precise range of automated machine-learning tools captured by the prohibition thus 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."[54] So far, no European state has carved exceptions under Article 22(2). The potential scope of the consent exception might be limited.[55] But it is also 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.[56]

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

[50] GDPR, supra note 20, art. 22(1).

[51] 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/media/pdf/resource_center/W29-auto-decision_profiling_02-2018.pdf [https://perma.cc/ZT2L-BVXT] [hereinafter Working Party Guidelines].

[52] Id. at 20–21 (excluding from Article 22(1) instances in which a human "reviews and takes account of other factors in making the final decision").

[53] It is not clear what "meaningful" supervision entails. Michael Veale & Lilian Edwards, Clarity, Surprises, and Further Questions in the Article 29 Working Party Draft Guidance on Automated Decision-Making and Profiling, 34 Computer L. & Security 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 51, at 21–22 (noting that while targeted advertising is not typically covered, the "intrusiveness" of the targeting, an individual's expectations, the way the advertisement is delivered, and the operator's knowledge of the "vulnerabilities" of the person might render it covered by Article 22(1)).

[54] GDPR, supra note 20, art. 22(2).

[55] The GDPR also defines consent in very narrow and demanding terms. Id. art. 4(11).

[56] 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.

"obtain . . . restriction[s]" on unlawful or inaccurate data processing,[57] 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.[58] Given the broad definitions of "processing"[59] and "profiling,"[60] these other provisions might be construed to bar some machine decisions. Again, the absence of implementation by national governments or enforcement actions by private, national, or supranational authorities means that there is much 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.[61] The right to a human decision, whether anchored in Article 22(1) or elsewhere, has precipitated contrastingly little scholarly attention to date.[62]

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

[57] GDPR, supra note 20, art. 18.

[58] Id. art. 21(1).

[59] 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).

[60] 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).

[61] 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, 7 Int'l Data Privacy L. 76, 77–78 (2017) (arguing against the inference of a right to an explanation), with Andrew D. Selbst & Julia Powles, Meaningful Information and the Right to Explanation, 7 Int'l Data Privacy L. 233, 234 (2017) (offering a "positive conception of the right").

[62] See Veale & Edwards, supra note 53, 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. Sci. 216, 224 (2017) (flagging the existence of Article 22, and noting that it provides something like a right to a human decision). This is one of very few previous articles to address this topic.

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 modes of decision making that are inconsistent with at least certain applications of machine learning. Consider the Sixth Amendment right to a jury trial in criminal cases.[63] Ignore for present purposes the ubiquity of plea bargaining,[64] and construe the Constitution's entitlement to a jury determination as a right to a human decision. In the original Constitution's scheme, juries were interposed at critical instances when coercive state power was wielded against individuals.[65] Jurors' obligation was "an active one" presupposing an autonomous exercise of human judgment.[66] So if the Sixth Amendment is violated by substitution of a judge for jurors, it is hard to see how the pivotal decisional role in a criminal trial could be played by a machine. Similarly, the use of a machine decision to increase a defendant's sentence beyond a statutory maximum may violate a defendant's Sixth Amendment right to a jury trial.[67] Further, the Confrontation Clause might create an additional friction on the adoption of machine decisions.[68]

Second, the idea of due process might also be grounds for a mandatory human decision rather than a machine judgment. At its core, the idea of procedural due process is thought to entail "notice and some kind of

[63] U.S. Const. art. III, § 2, cl. 3 ("The Trial of all Crimes . . . shall be by Jury . . . ."); id. amend. VI (right to "an impartial jury"); id. amend. 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 factfinding).

[64] See George Fisher, Plea Bargaining's Triumph, 109 Yale L.J. 857, 859 (2000).

[65] 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).

[66] 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 of Democracy 22–24, 68–75 (1994) (discussing the history of the development of the jury as a political institution in colonial America).

[67] See 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 L. Skeem, Risk Assessment in Criminal Sentencing, 12 Ann. Rev. Clinical Psychol. 489, 500–01 (2016).

[68] See Andrea Roth, Machine Testimony, 126 Yale L.J. 1972, 2039–51 (2017).

hearing."[69] 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.[70] 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.[71]

To date, there has not been a frontal challenge to algorithmic tools on Sixth Amendment or due process grounds. This is perhaps because such instruments have so far been used to support human decision making rather than formally ousting it. But the Wisconsin Supreme Court, in a 2016 decision called *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).[72] The defendant's argument in part rested on the defendant's limited ability to review 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.[73] The Wisconsin court rejected the defendant's constitutional arguments. It reasoned that the algorithm employed only publicly available data and data that the defendant herself supplied. Hence, the defendant could deny or explain the epistemic ground of a prediction.[74] The court flagged the group-based,

[69] 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, see 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).

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

[71] Friendly, supra note 69, 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.").

[72] State v. Loomis, 881 N.W.2d 749, 757 (Wis. 2016); see also Aziz Z. Huq, Racial Equity in Algorithmic Criminal Justice, 68 Duke L.J. 1043, 1081 (2019) (discussing the sentencing instrument used in Wisconsin).

[73] 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 sentenc[ing].

Loomis, 881 N.W.2d at 756–57.

[74] Id. at 761–62.

rather than individualized, nature of the data used to train the sentencing algorithm in that case. And it refused to view the use of a gender criterion as unconstitutional, since it was but one of several inputs to the defendant's sentence.[75] Predictions, the court nevertheless warned, "may not be considered as the determinative factor in deciding whether the offender can be supervised safely and effectively in the community" while remaining consistent with due process.[76]

Finally, a right to a human decision might be created by statute. In November 2019, Representatives Anna Eshoo and Zoe Lofgren introduced the Online Privacy Act into the House. Section 105 of that bill mandated "a reasonable mechanism by which [an] individual may request human review" of an "automated processing" decision with "reasonably foreseeable significant privacy harms."[77] And in April 2019, two Democratic Senators introduced a bill requiring the Federal Trade Commission (FTC) to enact regulations mandating "automated decision system impact assessments."[78] For certain algorithms, that assessment is required "prior to implementation."[79] The FTC would have to promulgate regulations to enforce this requirement, and both the FTC and Attorneys General of the several states could enforce it.[80] Although this does not contain an individual right of action, this bill may allow suits against the algorithms' users that may de facto preclude their deployment.

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 and due process, or as a side effect of statutory data protection measures. To the extent that American regulators follow the lead of the GDPR, the intellectual foundations for such a right nevertheless exist.

[75] 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.").

[76] 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. See 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, so I leave it to one side here.

[77] Online Privacy Act, H.R. 4978, 116th Cong. § 105 (2019).

[78] Algorithmic Accountability Act, S. 1108, 116th Cong. § 3(b)(1)(A) (2019).

[79] Id. § 3(b)(1)(A)(ii).

[80] Id. § 3(a)(1), (d), (e).

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. But it would be equally premature to deny that such a right is finding footing in both criminal and civil domains. Article 22(1) of the GDPR will likely be the cynosure of such a right. Although some version of that right has existed in national or supranational European law since 1978, technical advances in the capacity of machine learning, and in particular deep-learning tools that have emerged since the early 2000s,[81] place new strain on this status quo. New fears might spark conflict in unexpected sites.[82] But the rate of ensuing legal change is hard to predict.[83] Prohibitory regulation of machine decisions, for example, may undermine the business strategy of business entities that use some form of machine learning.[84] As a result, powerful lobbies favoring expansive use of data may oppose a broad construction of Article 22. On the other hand, those lobbies may alternatively perceive beneficial compliance-related economies of scale in regulation such as the GDPR. Rather than risk a balkanized regulatory terrain, these interest groups might accept some kind of right to a human decision as a lesser evil.

Despite this uncertainty, 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 formally and functionally articulated in quite varied guises. In form, it can vary from the explicit

[81] See infra text accompanying notes 106–10.

[82] But not all such fears are unwarranted. See, e.g., Edward Geist & Andrew J. Lohn, How Might Artificial Intelligence Affect the Risk of Nuclear War?, Rand Corp., https://www.rand.org/pubs/perspectives/PE296.html [https://perma.cc/8N4T-6CD2] (last visited Apr. 2, 2020) ("AI has the potential to exacerbate emerging challenges to nuclear strategic stability by the year 2040 even with only modest rates of technical progress.").

[83] 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 9–12 (1983) (defining "technical change" as the "manufacture and modification of tools" and discussing possible 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.

[84] MIT Tech. Review Insights, Machine Learning: The New Proving Ground for Competitive Advantage, MIT Tech. Rev. 4 (2017), https://www.technologyreview.com/s/603872/machine-learning-the-new-proving-ground-for-competitive-advantage/ [https://perma.cc/Z92G-WHFD] (finding in a survey of businesses that sixty percent already employed machine learning in some way, while only five percent had no interest in doing so in the future).

commitment contained in GDPR Article 22(1) to the commitment to a (human) jury decision in the Sixth Amendment. Such a right can in operation be imagined as a complete opt-out, or a right to an appeal. Exposure to non-human decision making might also be minimized by regulating the sheer volume of data flows too, for instance via individual opt-outs of data sharing.[85]

Second, 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, and concentrates instead on data processing by non-state actors. Perhaps this divergence can be glossed as another instance of the European concern with "respect and personal dignity" working out differently from an American fervor for "values of liberty."[86] 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 different circumstances in the two continents, perhaps leading over time to quite divergent legal regimes.

Still, 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. So there is at present a theoretical gap between the emergent right (however it is framed) and an empirically and normatively persuasive justification. It is that gap that this Article explores.

II. THE DIFFERENCE BETWEEN MACHINE DECISIONS AND HUMAN DECISIONS

To decide whether a right to a human decision is a good idea, we need to clearly understand what kind of non-human (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

[85] See, e.g., Assemb. B. 375, 2017–18 Leg. (Cal. 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.

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

that end, this Part undertakes three tasks. First, it offers a brief and non-technical account of the relevant technologies. I focus in particular on a class of machine-learning tools (also sometimes rather inaccurately labeled artificial intelligence[87]) that holds the most immediate promise for displacing human decisions. Second, I sketch the plausible technical domain of machine decisions. Finally, I parse the technical differences between machine and human decisions. All this sets up a more extended normative inquiry in 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."[88] Less abstractly, we might start with the observation that most such algorithms come in two forms: supervised and unsupervised.[89] We can usefully consider each in turn.

Supervised machine-learning algorithms define a function *f(x)* which produces an output *y* for any given input *x*.[90] Its outputs hence take the form of a sorting of *x* into categories of *y*:[91] 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" and "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

[87] I avoid this term because it has a potentially wider and more ambiguous scope. Cf. Stuart J. Russell & Peter Norvig, Artificial Intelligence: A Modern Approach 2–14 (3d ed. 2010) (offering a series of alternative definitions of AI that include thinking and acting humanly as well as rationally).

[88] M.I. Jordan & T.M. Mitchell, Machine Learning: Trends, Perspectives, and Prospects, 349 Science 255, 255 (2015).

[89] See LeCun et al., supra note 39, at 436, 442 (discussing the relatively higher frequency of supervised as opposed to unsupervised instruments). I do not address reinforcement learning here, although it is arguably a distinct form.

[90] See id. at 436. 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, 55 Comm. ACM 78, 79 (2012) ("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*.").

[91] The values of *y* must be identified ex ante by the programmer. Nat'l Research Council, Frontiers in Massive Data Analysis 104 (2013) (noting that in supervised learning, the analyst must actively specify a variable of interest); see Flach, supra note 3, at 14 (noting that "multi-class classification" is "a machine learning task in its own right").

to discern an actual causal relationship between *x* and *y*.[92] 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."[93] At bottom, this task is analogous to the function performed by familiar tools such as ordinary least squares and logistic regression analysis.[94] 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.[95] 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.[96]

An unsupervised machine-learning algorithm employs 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.[97] 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. dogs v. rats; people v. objects, etc. These categories 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,

[92] Jordan & Mitchell, supra note 88, 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).

[93] Alpaydin, supra note 3, at 3.

[94] Ian Goodfellow, Yoshua Bengio & Aaron Courville, Deep Learning 34–35 (2016) (noting functional similarities). 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 in hand and then proceeds to estimate its parameters, with 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, 16 Stat. Sci. 199, 199 (2001).

[95] 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, Matías Gámez & Noelia García, 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).

[96] Ziad Obermeyer & Ezekiel J. Emanuel, Predicting the Future—Big Data, Machine Learning, and Clinical Medicine, 375 New Eng. J. Med. 1216, 1217 (2016); accord Pedro Domingos, The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World 6–7, 23 (2015).

[97] Flach, supra note 3, at 14–15.

with each layer having a different level of abstraction.[98] 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.[99] The aim of unsupervised machine learning, in colloquial terms, is thus "to see what generally happens and what does not."[100]

Supervised and unsupervised machine-learning tools generally rest on distinct computational architectures.[101] Among the prominent forms of unsupervised learning methods are associational learning, cluster analysis, principal component analysis, and multi-dimensional scaling.[102] Supervised learning relies upon a number of different computational strategies. Two common ones are worth mentioning to give just a quick flavor of the technology. First, some supervised learning employs what are called "neural networks."[103] 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 yet further transformations—until at last they reach an output layer.[104] Relations between the neurons are recalibrated constantly by a learning algorithm, which reinforces connections between neurons that

[98] Geoffrey E. Hinton, Learning Multiple Layers of Representation, 11 Trends Cognitive Sci. 428, 429 (2007).

[99] John D. Kelleher & Brendan Tierney, Data Science 100–02 (2018). The algorithm can be understood as maximizing these parameters.

[100] Alpaydin, supra note 3, at 11.

[101] Some tools are used for both. Id. at 116 (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/N-459-CRUY] (describing the history of the SVM by one of the scientists that modified the algorithm in the 1990s).

[102] For a more general discussion of this approach, see Trevor Hastie, Robert Tibshirani & Jerome Friedman, Unsupervised Learning, *in* The Elements of Statistical Learning 485 (2d ed. 2009). In clustering, for example, "we generate a tree structure with clusters at different levels of granularity and clusters higher in the tree that are subdivided into smaller clusters" to "find structure in the data" that was previously unknown. Ethem Alpaydin, Machine Learning: The New AI 117 (2016).

[103] The following description of the neural net's internal operation draws on the lucid account in Alpaydin, supra note 102, at 88–103.

[104] For a useful graphical representation, see Yoshua Bengio, Machines Who Learn, Sci. Am., June 2016, at 46, 49.

are activated at the same time.[105] A second approach is the "random forests" school of algorithms, which generate predictions by producing thousands of decision trees mapping the data.[106] 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 trees in the forest.[107] 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.[108]

A recent development that is pertinent here is 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."[109] 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."[110]

[105] This is called a Hebbian learning rule. Kelleher & Tierney, supra note 99, 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).

[106] Leo Breiman, Random Forests, 45 Machine Learning 5, 5 (2001).

[107] Kelleher & Tierney, supra note 99, at 141–42.

[108] Id. at 136.

[109] LeCun et al., supra note 39, 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."); see Kelleher & Tierney, supra note 99, at 131 ("Deep-learning networks are simply neural networks that have multiple layers of hidden units . . . ." (emphasis and footnote omitted)).

[110] Alpaydin, supra note 3, at 309.

B. The Plausible Domain of Machine Decisions by the State

The ends to which machine-learning tools can be put by the state are not without limit.[111] Technical constraints on those tools have so far guided the state's adoption decisions. They also provide markers for the domain of potential machine decision making that is the focus of my inquiry. In brief, the state uses machine learning to make a range of high-volume decisions turning on empirical predictions for which a large pool of historical data is available. Absent sufficient training data, and absent some empirical parameter to predict, machine decisions are not appropriate.

Most machine-learning tools are used for the (non-causal) prediction of outcomes, in the case of supervised instruments, or the identification of clusters or associations, in the case of recommendation systems (such as those employed by Netflix and Amazon). Kelleher and Tierney 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.[112] In contrast, machine-learning tools are presently of less use when a problem requires "estimating the causal effect of an intervention."[113] An example of the perils of causal estimation with the current crop of instruments is a recent study of the allocation of hip replacement surgery among otherwise eligible patients.[114] 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.[115] Indeed, at present in the use of machine learning, "causal inference is only

[111] Judea Pearl, Theoretical Impediments to Machine Learning with Seven Sparks from the Causal Revolution (Jan. 15, 2018) (manuscript at 1–2), https://arxiv.org/pdf/1801.04016.pdf [https://perma.cc/RV3J-UNMK] (arguing that inability of machine learning to analyze counterfactuals to infer causation is a major impediment).

[112] Kelleher & Tierney, supra note 99, 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 1, 4 (2017).

[113] Athey, supra note 3, at 483.

[114] Kleinberg et al., supra note 95, at 493–94.

[115] Id. at 493.

possible when the analyst makes assumptions beyond those required for prediction methods."[116]

More generally, there is no suggestion in the literature that machine learning can be employed to resolve non-empirical questions. For instance, they cannot be used to answer ethical or other normatively inflected questions. There is also no suggestion that machines can resolve moral questions of priority, distribution, and belonging commonly associated with the political domain.

At least to date, state adoptions of machine learning have reflected these constraints. Criminal justice appears to be a leading adapter. States use machine-learning tools to determine how to allocate policing resources and to decide when to grant or deny bail.[117] More controversially, machine tools are being used to amplify the (already large) footprint of immigration law.[118] There are also, however, uses of machine tools beyond the carceral state. In a handful of states, welfare bureaucracies have started to use algorithmic tools to sort between recipients.[119] The same is true at the federal level, although the extent of such adoption remains unclear. The Environmental Protection Agency, for instance, uses machine learning to evaluate the effects of some toxins.[120] The Internal Revenue Service uses a machine-learning instrument to predict fraud and abuse.[121]

[116] Athey, supra note 3, at 484; see also id. at 485 (flagging limitations due to "incentives and manipulability").

[117] See Huq, supra note 72, at 1068–76 (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).

[118] 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/ [https://perma.cc/9LDE-RKA-Q] (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").

[119] See Eubanks, supra note 8, at 14–38 (describing the way in which algorithms affect low-income defendants).

[120] See U.S. Envtl. Prot. Agency, Using ToxCast to Predict Chemicals Potential for Developmental, Reproductive and Vascular Development Toxicity (Nov. 10, 2011), https://www.epa.gov/sites/production/files/2013-12/documents/toxcast-models-fact-sheet.-pdf [https://perma.cc/SB8W-LE5P].

[121] David DeBarr & Maury Harwood, Relational Mining for Compliance Risk (2004), http://www.irs.gov/pub/irs-soi/04debarr.pdf [https://perma.cc/VM4U-92ZM]. For a more recent, critical view, see Kimberly A. Houser & Debra Sanders, The Use of Big Data Analytics

Still, many decisions related to the law fall outside the domain of the technologically plausible. There is no suggestion that a machine can make ethical decisions of the kind infused into many public and private law domains. Hence, machine learning provides no substitute for paradigmatic complex civil and criminal disputes.[122] "Machine judges" are much debated,[123] but it is difficult to see how they get off the ground at least for hard cases. Machines cannot (yet?) resolve the difficult and inescapably normative questions of aggregation, distribution, and belonging that characterize politics.

C. 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 two perspectives on those differences. First, probably the most obvious discontinuities are related to the scale, capacity, and underlying mechanisms of machine learning. In what follows, I first accept the efficiency gap between machines and humans, but query its significance for an inquiry into rights. Second, it is also sometimes said that machine decisions are more opaque than human ones. I doubt, though, whether a transparency gap exists. Rather, even as 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 compared to humans are well-founded, at least when framed as a generalization

by the IRS: Efficient Solutions or the End of Privacy as We Know It, 19 Vand. J. Ent. & Tech. L. 817 (2017).

[122] There is a technical literature on predicting judicial decisions. See, e.g., 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 PeerJ Computer Sci. 1 (2016). But prediction is quite distinct from the task of ethically infused judging. Chris Johnston, Artificial Intelligence 'Judge' Developed by UCL Computer Scientists, Guardian (Oct. 23, 2016), https://www.theguardian.com/technology/-2016/oct/24/artificial-intelligence-judge-university-college-london-computer-scientists [https://perma.cc/T78A-2SYM] (quoting Nikolas Aletras, creator of the algorithm) ("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.").

[123] There is some fanciful speculation on this point. See, e.g., 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"); 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 [https://perma.cc/3VLP-GSTN]. 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).

rather than case-to-case comparisons.[124] In the last Section, I take up 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 their architecture of reasoning; in their relative 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 support 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."[127] It is in practice more akin to regression analysis than to the operation of human neurons. Among the first neural networks was, for instance, the "perceptron," comprising a single "neuron" or node, which executed a single non-linear function.[128] This 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

[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?, 538 Nature 20, 21 (2016) (asserting that neural networks are modeled on the brain).

[126] Bengio, supra note 104, at 49.

[127] Jayesh Bapu Ahire, Artificial Neural Network: Some Misconceptions, Medium (Jan. 27, 2018), https://medium.com/swlh/artificial-neural-network-some-misconceptions-cb93e80b3-4bb [https://perma.cc/7EXM-N6TD]; see Adam H. Marblestone, Greg Wayne & Konrad P. Kording, Toward an Integration of Deep Learning and Neuroscience, 10 Frontiers Computational Neuroscience 1, 1–2 (2016), https://www.frontiersin.org/articles/10.3389/-fncom.2016.00094/full [https://perma.cc/JMZ4-EPBV] (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 Psychol. Rev. 386, 387 (1958).

information storage," the machine-learning field "has largely focused on instantiations of a single principle: function optimization."[129] It 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, importantly, is also 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 infect the underlying texts.[135] Incautiously deployed, predictive 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] For 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

[129] Marblestone et al., supra note 127, at 1.

[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 Jeffrey J. Rachlinski, A Positive Psychological Theory of Judging in Hindsight, 65 U. Chi. L. Rev. 571, 572 (1998) (discussing hindsight bias in judicial decisions).

[132] See 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] See Daniel L. Chen & Jess Eagel, Can Machine Learning Help Predict the Outcome of Asylum Adjudications?, Proc. 16th Edition Int'l Conf. on Artificial Intelligence & L. 237 (2017) (finding that case-relevant factors explain only about one-third of the outcomes in asylum decisions).

[134] See Huq, supra note 72, at 1111–33 (discussing flawed training data and moral choices created by background racial inequalities).

[135] Aylin Caliskan, Joanna J. Bryson & 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, Proc. 17th SIAM Int'l Conf. on Data Mining 777–85 (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).

a decision, moreover, studies in a variety of fields suggest that large gains in human well-being 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-learning tools, so too it would be an error to ignore the range of social goods that can be generated by their employ. Rather, a key question is whether the flaws of machine learning are easier to identify and remedy in practice than the flaws of its human analog.

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. Of course, this difference in computational capacity 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. And no human can see as far as a modern telescope or as deep into materials as a modern electron 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. Indeed, it is striking that the most sophisticated

[138] Jon Kleinberg et al., Human Decisions and Machine Predictions, 133 Q.J. Econ. 237, 237–39 (2018) (discussing bail); Kleinberg et al., supra note 95, at 494 (discussing medical decisions). 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 & David G. Groves, The Enemy of Good: Estimating the Cost of Waiting for Nearly Perfect Automated Vehicles 29 (2017), https://www.rand.org/-pubs/research_reports/RR2150.html [https://perma.cc/39WV-3SAQ] (recommending rapid introduction of driverless vehicles on safety grounds).

[140] See infra Section III.C.

[141] Obermeyer & Emanuel, supra note 96, at 1217.

forms of deep learning are used today to execute functions such as speech or handwriting recognition that most children manage with ease.[142]

At least in theory, therefore, machine learning is capable of better (in the sense of more accurate) predictions than humans. This capacity, to be sure, is not always realized. But legal design, including the design of rights, is best thought of as 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 threshold hint that a right to a human decision might risk stymying beneficial institutional changes. And at least absent some reason to think that machine errors are irremediable in a way that human errors are not, there is no reason to prefer the latter.

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."[143] It rests instead on "the high-dimensionality of data, complex code, and changeable decision-making logic."[144] This concern with transparency seems to motivate in part the demand for a right to a human decision.[145] The empirical predicates of this claim hence warrant separate treatment.

We should first rule out here a common complaint about algorithmic transparency. This hinges on the unwillingness of the corporate entities

[142] See, e.g., Cheng-Lin Liu et al., Handwritten Digit Recognition: Benchmarking of State-of-the-Art Techniques, 36 Pattern Recognition 2271, 2271 (2003) (reporting the accuracy of handwriting recognition by state-of-the-art machine learning). Consider whether causal inference is a task that, per Hume, is better viewed as a matter of imaginative conjuring rather than induction—and hence beyond the remit of machines. Cf. Sendhil Mullainathan & Jann Spiess, Machine Learning: An Applied Econometric Approach, 31 J. Econ. Persp. 87, 87–88 (2017) (explaining how "machine learning algorithms are not built" for certain applications).

[143] Tal Z. Zarsky, Transparent Predictions, 2013 U. Ill. L. Rev. 1503, 1519, 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 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); Alyssa M. Carlson, Note, The Need for Transparency in the Age of Predictive Sentencing Algorithms, 103 Iowa L. Rev. 303, 324–26 (2017) (advocating that freedom of information laws extend to non-state providers of algorithmic tools).

[144] Brent Daniel Mittelstadt et al., The Ethics of Algorithms: Mapping the Debate, 3 Big Data & Soc'y 1, 6 (2016).

[145] See supra text accompanying notes 9–13.

that own the machine-learning instrument to disclose their details for fear of economic harm.[146] 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.[147] There is nothing distinctive, that is, in the fact that it is a machine rather than a human 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."[148] The classification rule identified by a machine-learning tool can be a dynamic function of a neural network rather the result of one sequence of calculations.[149] 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.[150] 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 non-

[146] For examples of this concern, see Frank Pasquale, The Black Box Society: The Secret Algorithms that Control Money and Information 12–15 (2015); 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").

[147] 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").

[148] Jenna Burrell, How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms, 3 Big Data & Soc'y 1, 6 (2016).

[149] Id. at 8.

[150] Kroll et al., supra note 2, at 647; Joshua A. Kroll, The Fallacy of Inscrutability, 376 Phil. Transactions Royal Soc'y A 1, 9 (2018) [hereinafter Kroll, The Fallacy of Inscrutability] (describing such disclosure as "neither necessary nor sufficient" for improving understanding).

repudiable evidence of all nodes' actions."[151] From this record, "faulty" nodes in a machine-learning system can be detected.[152] 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.[153] This has a human analog of sorts in experimental tests of the trolley problem.[154] Finally, some kinds of algorithmic design may be more amenable to interpretation than others.[155] Models that are simple to learn tend to be simpler to represent.[156] A designer thus can, within limits, select a computational architecture precisely for its amenability to post hoc explanation. Alternatively, it is possible to "approximate the model" in simpler form even after it has been created and applied in the wild.[157] 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

[151] Andreas Haeberlen, Petr Kouznetsov & Peter Druschel, PeerReview: Practical Accountability for Distributed Systems, Proc. 21st ACM Symposium on Operating Systems Principles 175, 175 (2007); Kroll et al., supra note 2, at 662–72 (describing various methods of ensuring accountability in machine decisions even without full transparency, such as cryptographic commitments, zero-knowledge proofs, and fair random choices); see 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).

[152] Haeberlen et al., supra note 151, at 175.

[153] Sandra Wachter, Brent Mittelstadt & Chris Russell, Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR, 31 Harv. J.L. & Tech. 841, 851–52 (2018) ("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 Sandy Huang, Nicolas Papernot, Ian Goodfellow, Yan Duan & Pieter Abbeel, Adversarial Attacks on Neural Network Policies (Feb. 8, 2017) (manuscript at 1), https://arxiv.org/pdf/1702.02284.pdf [https://perma.cc/-BVL7-399X]. These are both "subject-centric" 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, 16 Duke L. & Tech. Rev. 18, 56 (2017).

[154] See 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) (finding that people change their moral reasoning in response to different variations of the trolley problem).

[155] Desai & Kroll, supra note 151, at 11 n.61 (suggesting that decision tree, naïve Bayes, and rule learners are easier to interpret than neural networks or support vector machines).

[156] Michael Gleicher, A Framework for Considering Comprehensibility in Modeling, 4 Big Data 75, 82 (2016).

[157] Andrew D. Selbst & Solon Barocas, The Intuitive Appeal of Explainable Machines, 87 Fordham L. Rev. 1085, 1110–12 (2018).

world is "merely a design choice, not an inevitability of the complexity of large systems."[158]

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 machines are any more opaque than humans.[159] 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 for 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 effective.

If there are not systemic reasons to think that machine decisions are always more opaque than human decisions, consider instead the possibility of a rough equality between human and machine transparency, at least across the mine-run of cases. Notice that other minds are just as much black boxes as are machine-learning instruments. Thinking remains obdurately difficult to theorize. There is "no generally accepted theory of how cognitive phenomena arise from computations in cortex"[160] or of whether consciousness serves any "significant function" for an organism.[161] 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

[158] Kroll, The Fallacy of Inscrutability, supra note 150, 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 (Feb. 5, 2018) (manuscript), https://arxiv.org/pdf/1802.00682.pdf [https://perma.cc/Q5CF-KWXY].

[159] Anthony J. Casey & Anthony Niblett, A Framework for the New Personalization of Law, 86 U. Chi. L. Rev. 333, 355–56 (2019).

[160] Leslie G. Valiant, What Must a Global Theory of Cortex Explain?, 25 Current Opinion Neurobiology 15, 15 (2014).

[161] David M. Rosenthal, Consciousness and Its Function, 46 Neuropsychologia 829, 831, 839 (2008) (arguing that it does facilitate "rational thinking, intentional action, executive function, or complex reasoning").

conclusion."[162] The bigger the decision, moreover, the less amenable it can be to reasoned resolution.[163]

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.[164] 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. This is a capability that precedes sophisticated cognition. Most people (at least those over the age of four who are not law professors) navigate the social world on the working assumption that other people have minds.[165] Some reasonable proportion of the time, our beliefs about other minds are close enough to the mark to permit effectual social interactions.[166] Further, the

[162] Vijay Pande, Opinion, Artificial Intelligence's 'Black Box' Problem Is Nothing to Fear, N.Y. Times (Jan. 25, 2018), https://www.nytimes.com/2018/01/25/opinion/artificial-intelligence-black-box.html [https://perma.cc/D6AG-78S3]; see also Frank Fagan & Saul Levmore, The Impact of Artificial Intelligence on Rules, Standards, and Judicial Discretion, 93 S. Cal. L. Rev. 1, 9 (2019) ("The best judges, like the best athletes and teachers, are often unable to identify the reasons for their successes.").

[163] Edna Ullmann-Margalit, Big Decisions: Opting, Converting, Drifting, 58 Royal Inst. Phil. Supplements 157, 158–59 (2006) (describing a category of "big" decisions that cannot be resolved by standard cost-benefit analysis).

[164] Finale Doshi-Velez et al., Accountability of AI Under the Law: The Role of Explanation (Nov. 21, 2017) (manuscript at 9), https://arxiv.org/pdf/1711.01134.pdf [https://perma.cc/-TN8K-4GHC].

[165] 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.

[166] 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

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 when it comes to other people's mental processes, in short, is a problem for philosophers.[167] It is not necessarily a worry for the rest of us.[168]

So it is far from clear that transparency is systematically more inaccessible for machine rather than human interlocutors.[169] 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.[170] A difference of a predictable and stable sort between the two domains, as a general matter, is hard to discern.[171] 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

other people and thereby to alter their behavior" as an important component of "social intelligence").

[167] 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).

[168] Recall here Samuel Johnson's famous refutation of Bishop Berkeley. See James Boswell, Life of Samuel Johnson, LL.D. 131 (London, Henry Washbourne 1847).

[169] 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 164, 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.").

[170] 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 153, at 67; accord Mike Ananny & Kate Crawford, Seeing Without Knowing: Limitations of the Transparency Ideal and Its Application to Algorithmic Accountability, 20 New Media & Soc'y 973, 979–80 (2018) (offering a critique of transparency as a "neoliberal" solution).

[171] 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. See Doshi-Velez et al., supra note 164, at 11. It is hard to think of a parallel with humans.

arguments explored in Part III, therefore, must be carefully framed to avoid any assumption about necessary differentials in opacity.

D. Entanglements of Human and Machine Action

There is yet one other perplexing preliminary of an empirical rather than a legal cast: to what extent it is even 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."[172] But is there really such a beast?

Surely not—for three reasons. *First*, all machine-learning tools are at their origin the fruit of specific human design and engineering choices. There is simply no such thing as a wholly endogenous algorithm.[173] And the design of a machine-learning tool is not a mechanical task. It is freighted with normative choices. A designer must select (inter alia) 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 is 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.[174]

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. The designer must then decide how to connect the network's different elements. She might choose to create a recurrent

[172] GDPR, supra note 20, art. 22(1) (emphasis added).

[173] 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, This A.I. Can Build A.I. Itself: Big Tech Bypasses Humans to Accelerate Advances in Machine Learning, N.Y. Times, Nov. 6, 2017, at B1. Similar research is ongoing at Carnegie Mellon University. See Renato Negrinho & Geoff Gordon, DeepArchitect: Automatically Designing and Training Deep Architectures (Apr. 28, 2017) (manuscript at 1), https://arxiv.org/-abs/1704.08792 [https://perma.cc/G3K7-S8PB]. A human engineer is still needed to guide some of the search processes over different potential architectures. Id.

[174] Kelleher & Tierney, supra note 99, at 99–100; see also Casey & Niblett, supra note 159, at 354 ("[H]umans are involved in all stages of setting up, training, coding, and assessing the merits of the algorithm.").

neural network, in which the network's topology is looped.[175] Each neuron processes inputs in the context of the previous inputs processed, creating a sort of "memory."[176] 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).[177] 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.[178] For example, in the policing context, there is a concern that historical arrest data, if used to motivate future force deployment decisions, will reflect and reproduce any troubling assumptions about racial proclivities toward crime that have been prevalent among police officers in the past.[179] Having collected data, a designer needs to pre-process 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.[180] The process of learning then "require[s] close involvement by a human," who must craft labels for training data and then generate hypotheses to guide the process of optimization.[181] Saul Levmore and Frank Fagan, arguing for the

[175] Bengio, supra note 104, at 50.

[176] Kelleher & Tierney, supra note 99, at 132–33.

[177] Id. at 133; see also LeCun et al., supra note 39, at 438–39 (discussing convolutional networks in general terms).

[178] 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 [https://perma.cc/R8LY-D652] (noting problems that can arise from missing data).

[179] Huq, supra note 72, at 1053–54, 1115–23.

[180] Saleema Amershi, Maya Cakmak, William Bradley Knox & Todd Kulesza, Power to the People: The Role of Humans in Interactive Machine Learning, AI Mag., Winter 2014, at 105, 105.

[181] Bengio, supra note 104, 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

inevitability of human-machine partnerships, additionally suggest that missing data problems mean that human "thinking about data limitations and goal-setting" will inevitably be needed even after a system is up and running.[182] On their account, the implementation as well as the design process for machine learning will usually be punctuated by "lengthy and asynchronous iterations" of human-machine interaction that ensure that machine learning is always critically molded by a human hand.[183]

Third, once up and running, machine-learning tools still need "human caretakers," tasked with everything from moderating the results of deep-learning algorithms used for simulating vision to serving as maintenance workers who clean and repair the data centers used to house large pools of information necessary for algorithms' operation.[184]

The idea of a machine that will run of its own accord, in short, appears far from plausible. To be sure, this could change. Some deep-learning instruments already require "very little engineering by hand" at the operational stage (although their design still presents considerable challenges).[185] 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,[186] lies beyond the technological event horizon.

All this brings us back to the question of how precisely the notion of a decision "based solely on automated processing"[187] should be construed. 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 decision should be jettisoned. Another possible interpretation of the intuition behind a right to a human decision would focus not on the

the neural network by adjusting parameters slightly"); see also Amershi et al., supra note 180, at 106 (describing "iterative exploration of the model space" by practitioners once a learning algorithm is up and running).

[182] Fagan & Levmore, supra note 162, at 9–10.

[183] Amershi et al., supra note 180, at 105.

[184] Alex Campolo, Madelyn Sanfilippo, Meredith Whittaker & Kate Crawford, AI Now, AI Now 2017 Report 12 (2017), https://assets.ctfassets.net/8wprhhvnpfc0/1A9c3ZTCZa2KEY-M64Wsc2a/8636557c5fb14f2b74b2be64c3ce0c78/_AI_Now_Institute_2017_Report_.pdf [https://perma.cc/FYA4-W9AV] (internal quotation marks omitted).

[185] LeCun et al., supra note 39, at 436.

[186] 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. Vincent C. Müller & Nick Bostrom, Future Progress in Artificial Intelligence: A Survey of Expert Opinion, *in* Fundamental Issues of Artificial Intelligence 553, 566 (Vincent C. Müller ed., 2016).

[187] GDPR, supra note 20, art. 22(1).

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. In other words, it demands that in the immediate transactional frame 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 that a human engineer once upon a time calibrated the machine with which one interacts 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.[188] Just as the rules-standards distinction hinges on the timing (and to some extent the degree of specification[189]) 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—or by eliminating the machine entirely.[190] This suggests that to the extent that the Working Party Guidelines to Article 22(1) of the GDPR are ambiguous, they should be interpreted as mandating "meaningful" and ex post review.[191] This suggests that Article 22(1)

[188] Louis Kaplow, Rules Versus Standards: An Economic Analysis, 42 Duke L.J. 557, 559–63 (1992).

[189] 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.").

[190] The explanatory memorandum for the draft Universal Guidelines for 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, Universal Guidelines for Artificial Intelligence, Explanatory Memorandum and References (Oct. 2018), https://thepublicvoice.org/ai-universal-guidelines/memo/ [https://perma.cc/63G8-2G2S]. To view the Guidelines, see The Public Voice, Draft Universal Guidelines for Artificial Intelligence (Oct. 23, 2018), https://epic.org/international/AIGuidleinesDRAFT20180910.pdf [https://perma.cc/G3WN-ZSKB].

[191] Working Party Guidelines, supra note 51, at 20–21 (excluding from Article 22(1) instances in which a human "reviews and takes account of other factors in making the final decision"); see also infra text accompanying note 192.

should be read to be violated only when there is an absence of ex post, as opposed to ex ante, human involvement. This does not solve all of Article 22's interpretive difficulties. The GDPR, as I have noted, does not explain how to define the content of "meaningful" review.[192] Should it be defined in terms of substantive reconsideration of the decision on its own merits? Or is it better to think of it as review for procedural and technical regularity?

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, at least 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.

* * *

Machine learning denotes a large field of heterogenous and evolving computational forms. What is mapped here barely scratches its surface. I have stressed some of the central taxonomical lines sufficiently to allow intelligent discussion of the right to a human decision. This allows us to sketch a (limited) domain in which machine decisions are plausibly used.

That sketch, I hope, clarifies the following points. 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 rights of some sort. 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

[192] See Veale & Edwards, supra note 53, at 401 ("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 51, 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)).

incentivize the creation of better machines rather than their substitution with humans. In contrast, transparency may be a focal point for much legal scholarship, but it 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. Can a Right to a Human Decision Be Justified?

With this technological context in hand, this Part explores the available normative justifications for a right to a human decision. I focus on the direct state applications of machine-learning tools to individuals for the purpose of allocating benefits or burdens. Criminal justice and welfare administration have to date been characterized by the most rapid uptake of such tools.[193] The case of the person who is subject to coercion, or denied a benefit, because of a mistaken machine decision is a compelling circumstance in which to ask about a human decision right. I thus assume that something material is at stake when the machine (or human) decides.

I identify and discuss four potential normative arguments for a right to a human decision. The first relates to *accuracy* concerns at the population level. The second set of *subject-focused* arguments trains on the actions, or potential actions, of the individual exposed to algorithmic classification. Do machines, for example, foreclose certain opportunities for the exercise of meaningful human agency? The third cluster of reasons is *classifier-focused*. These build on the intuition that the state in particular owes to individuals a certain species of decision making, even if the choice between processes would not be outcome-determinative. The final group of reasons focuses on *systemic effects*. This class of reasons dilates the analytic 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. Instead, I argue that the domain of plausible machine decisions is best delineated by reference to the

[193] See supra text accompanying notes 30, 117, 119.

technical constraints on such tools. These practical grounds do suggest outer bounds to machine decision making—but are not derived from normative theorizing.

I try here to consider the widest possible range of normative theories. This means drawing on, but not leaning conclusively upon, precedential arguments. So it is possible (even likely) that the Sixth Amendment right to a jury trial would, as presently interpreted by the Supreme Court, foreclose a machine decision on facts related to guilt or innocence.[194] This might be relevant evidence of a widely shared moral intuition. But I do not assume that it disposes of the normative question. I also try to avoid tautological reliance on ambiguous and contested concepts such as "autonomy" and "dignity."[195] Such concepts, to be sure, play important roles in normative theorizing about the state's obligations in respect to, say, "due process."[196] But they require specification. A 2011 survey of Supreme Court deployments of the word "dignity" found five different semantic usages by the Justices alone: "institutional status as dignity, equality as dignity, liberty as dignity, personal integrity as dignity, and collective virtue as dignity."[197] In the philosophical tradition, Schopenhauer pointedly described dignity as "the shibboleth of all the

[194] See supra text accompanying notes 63–65.

[195] 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 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, 129 (2016) (invoking "autonomy-related concerns that also involve harms to individual dignity"); see also Margaret Hu, Algorithmic Jim Crow, 86 Fordham L. Rev. 633, 695–96 (2017) (offering a similar argument on grounds including disparate impact and rights of informational privacy).

[196] 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, 899–906 (1981) [hereinafter Mashaw, Administrative Due Process] (discussing dignitary definition of due process).

[197] Leslie Meltzer Henry, The Jurisprudence of Dignity, 160 U. Pa. L. Rev. 169, 189–90 (2011) (emphasis omitted); see also Vicki C. Jackson, Constitutional Dialogue and Human Dignity: States and Transnational Constitutional Discourse, 65 Mont. L. Rev. 15 (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, 877–78 (1994) (distinguishing a "descriptive" and an "ascriptive" sense of autonomy in reference to First Amendment debates).

perplexed and empty-headed moralists who concealed behind that imposing expression their lack of any real basis of morals."[198] Abstract normative terms such as "autonomy" and "dignity" are useful only after they have been colored and bounded through the invocation of a more full-throated normative theory.[199]

A. Accuracy in Decision Making

The Supreme Court has at times suggested that the Due Process Clause of the Fourteenth Amendment creates an entitlement to an accurate decision.[200] 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 factfinding, is to minimize the risk of erroneous decisions."[201] 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 post-conviction review of capital punishment, the Supreme Court has shied away from a personal right to a *true* determination.[202] A right to

[198] Michael Rosen, Dignity: Its History and Meaning 1 (2012) (citation omitted).

[199] For example, Martha Nussbaum has employed the capabilities approach to human wellbeing to give the idea of dignity meaningful content. Martha Nussbaum, Human Dignity and Political Entitlements, *in* Human Dignity and Bioethics: Essays Commissioned by the President's Council on Bioethics 351, 351 (Adam Schulman & Thomas W. Merrill 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 13–78 (Meir Dan-Cohen ed., 2012). Neither of 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).

[200] 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.").

[201] Greenholtz v. Inmates of the Neb. Penal & Corr. 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.").

[202] See Dist. Attorney's Office v. Osborne, 557 U.S. 52, 71 (2009) (explaining that whether "actual innocence" exists as a federal right remains an "open question" (internal quotation

accuracy is instead understood in probabilistic terms, such as "beyond a reasonable doubt."[203] 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"[204] is sufficient for constitutional purposes.

Can a due process right to an accurate decision, understood in these systemic, population-wide, and probabilistic terms, provide a normative foundation for a right to a human decision? Let us 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,[205] and instead focus on the suggestion that humans are better than machines overall. 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 those amenable to human decisions. Because the current crop of algorithmic tools generally identify correlational rather than causal relationships, there is a cluster of empirical questions that they are not well designed to answer.[206] Although an individual might have a legitimate complaint if subject to an algorithmic decision on a matter of causal inference or normative reasoning, 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 false positives and negatives in the aggregate than most human decision making.[207] It is said that this is true in contexts such

marks omitted)); Herrera v. Collins, 506 U.S. 390, 404 (1993) (explaining that "actual innocence" has never been held to be an independent constitutional claim (internal quotation marks omitted)).

[203] 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.").

[204] Patrick Woolley, Rethinking the Adequacy of Adequate Representation, 75 Tex. L. Rev. 571, 590 (1997).

[205] See infra Subsection III.B.2.

[206] See Mullainathan & Spiess, supra note 142, 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. Note that work on causal inference through machine learning is ongoing. Perhaps in less than a decade, this caveat will therefore be otiose.

[207] See supra text accompanying notes 136–38138.

as pretrial bail[208] and domestic violence-related arraignments,[209] although the technical plausibility of predicting violence at an individual level has been challenged.[210] At least if it proves feasible to predict individual violence, it is likely implausible to hold that a right to an accurate decision maker (in this sense) entails a right to a human decision maker in these cases.[211] Indeed, as the conclusion explores, perhaps the opposite is true, such that legal rules might be designed with an eye to improving, rather than ousting, extant machine decisions.

To be sure, 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 quite difficult. Accuracy does not provide a useful lens. Studies of algorithmic bail tools demonstrate that the most common population-wide measures of false positive rates cannot simultaneously be equalized.[212] 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.[213] 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

[208] Kleinberg et al., supra note 138, at 237–38.

[209] Richard A. Berk, Susan B. Sorenson & Geoffrey Barnes, Forecasting Domestic Violence: A Machine Learning Approach to Help Inform Arraignment Decisions, 13 J. Empirical Legal Stud. 94, 110 (2016) (finding that the release rate of 20% repeat offenders in a pool of domestic violent defendants could be dropped to a 10% rate through a move from judicial to machine-led determinations).

[210] Technical Flaws of Pretrial Risk Assessments Raise Grave Concerns 2 (2019), https://dam-prod.media.mit.edu/x/2019/07/16/TechnicalFlawsOfPretrial_ML%20site.pdf [https://perma.cc/D5E9-FA8M] [hereinafter Technical Flaws].

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

[212] 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, Proc. 23rd ACM SIGKDD Int'l Conf. on Knowledge Discovery & Data Mining 797, 798–99, 804 (2017); Jon Kleinberg, Sendhil Mullainathan & Manish Raghavan, Inherent Trade-Offs in the Fair Determination of Risk Scores (Nov. 17, 2016) (manuscript at 4, 9), https://arxiv.org/pdf/1609.05807 [https://perma.cc/2W8C-X65Y].

[213] See, e.g., Huq, supra note 72, at 1128–33 (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).

positives, with human decisions. As a result, this problem of equality and non-discrimination should be scrutinized 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 individual exposed to a machine decision. She may feel disempowered and enervated by her exclusion from any effectual role in the process. That is, she experiences 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 seek an exception from the rule normally applicable to their case.

I explore here whether a right to a human decision might be predicated on the non-instrumental interest in bare participation, or alternatively the interest in giving (broadly understood) reasons.[214] In the end, I conclude that there are more grounds for skepticism than hope here for finding solid foundations on which to rest a right to a human decision.

1. Participation Interests

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 non-instrumental claim might sound in a "deep-rooted historic tradition that everyone should have his own day in court,"[215] even if that participatory entitlement will not necessarily alter the outcome of a proceeding. It might also be founded on the felt human "need to explain and justify our actions," such that "the loss of the

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

[215] Martin v. Wilks, 490 U.S. 755, 762 (1989) (internal quotation marks omitted) (quoting 18 Charles Alan Wright, Arthur R. Miller & Edward H. Cooper, Federal Practice and Procedure § 4449, at 417 (1981)).

opportunity to do so denies our self-worth."[216] Participation, on this view, creates immediate hedonic gain.[217] 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."[218] In this non-instrumental vein, participation has been glossed as a morally important manifestation of autonomy,[219] or as a manifestation of dignity.[220] The overlap in the scope of those two general terms is suggestive of their joint and overlapping ambiguity.[221]

I do not think that a non-instrumental interest in participation can be used to justify a right to a human decision. Consider first when and how such an interest is recognized in present law. Observed practice of course is not a precise measure of moral value. But it provides a rough guide to the weight that an interest is generally accorded and so is a starting point for reflection. 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 gives us a pro tanto reason to be skeptical of arguments on its basis toward a right to a human decision.

[216] Mashaw, Administrative Due Process, supra note 196, at 903.

[217] Cf. Amershi et al., supra note 180, at 111 (noting that transparency about an algorithm's operation is related to greater user satisfaction).

[218] T.R.S. Allan, Procedural Fairness and the Duty of Respect, 18 Oxford J. Legal 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.

[219] 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.").

[220] 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); Mashaw, Due Process Calculus, supra note 201, at 49–52 (identifying dignity as an important consideration for due process).

[221] 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. See 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").

A place where a bare right to participation is embodied especially crisply in jurisprudence is the Sixth Amendment right to counsel.[222] The latter extends to create rights to self-representation as well as to counsel-of-choice in the criminal adjudication context. This is so even if the former yields no better outcomes for defendants.[223] The strength of the autonomy-related justification for these rights remains contested among academics.[224] In a recent choice-of-counsel case, though, a plurality of the Supreme Court described the defendant's right to elect counsel as "fundamental,"[225] perhaps suggesting something more than an instrumental concern.

Yet I am skeptical that the participation interest reflected in these decisions is one of *personal* involvement. True, 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 (the modal case), 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. Given the centrality of lawyers in the criminal process, it is reasonable to think that the relevant autonomy protected by the Sixth Amendment is the interest in *electing one's counsel*—and not participation for its own sake. Consistent with this intuition is the fact that, when a criminal defendant exercises a right to self-representation,

[222] See U.S. Const. amend. VI ("In all criminal prosecutions, the accused shall . . . have the Assistance of Counsel for his defence.").

[223] 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."). See also Flanagan v. United States, 465 U.S. 259, 268 (1984) (reasoning that this right "reflects constitutional protection of the defendant's free choice independent of concern for the objective fairness of the proceeding"). The relationship between representation and adjudicative outcomes is a complex and difficult one. For an insightful empirical analysis in the civil context, see D. James Greiner et al., The Limits of Unbundled Legal Assistance: A Randomized Study in a Massachusetts District Court and Prospects for the Future, 126 Harv. L. Rev. 901, 903–04 (2013).

[224] Compare John Rappaport, The Structural Function of the Sixth Amendment Right to Counsel of Choice, 2016 Sup. Ct. Rev. 117, 118 (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).

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

that right is circumscribed in highly formalized ways that establish "a certain distance" between judges and parties.[226] 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). So the interest in self-representation, on this view, is a byproduct of this more general right. It is not an instantiation of a specifically protected participation interest.

Might instead a participation-based right to a human decision be justified by the dignity interest in hearing and understanding directly the reasons for a decision? A threshold problem with this argument is that machine-learning instruments may be roughly as opaque as human decisions.[227] Indeed, it is important to notice that as a practical matter there may well not be much phenomenological distance between the bafflement an unschooled criminal defendant reasonably feels 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. And 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.

Worse, the argument from dignity helps itself to an empirical premise to which it is not plainly entitled. This is the idea that personhood is respected more by a human decision maker than a machine. In practice, quite the opposite might well be true. Especially in the context of mass adjudicative systems (such as welfare determinations and criminal justice), the experience of going before a human decision maker who rapidly, perhaps summarily, ranks you may be fraught with indignity. Not least, there is the prospect of having one's flaws aired and evaluated by a powerful stranger. Then, there is the risk that the decision maker may take against you, perhaps for bad (animus-related) reasons, or perhaps because they simply dislike you. Finally, there is an unavoidable publicity attendant on that human decision that might weigh heavily. In contrast, the impersonal and non-judgmental character of a machine decision might

[226] 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.").

[227] See supra text accompanying notes 143–71.

well be more conducive to human dignity than any human-driven process. Dignity, in short, should not simply be assumed to run with human decision making. Instead, it may well be that in many cases our sense of integrity and standing are best preserved by insulation from human scrutiny. Under plausible empirical assumptions, it may well cut in favor of a machine decision.

I am not convinced there is a hedonic argument for a participation-based right to a human decision. An ex post human decision will often be an ineffectual 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. It may be instead the intervention of another machine. At least one algorithmic tool has been developed, for instance, as a means of "explain[ing] the predictions of *any* classifier or regressor in a faithful way."[228] 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"[229] the grievances of affected individuals falls away. It 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 in turn be a contingent feature of social experience, and that which strikes us today as dehumanizing or insensitive will appear to our children as merely sensible and mundane.[230] 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.

Finally, perhaps an individual subject to a machine decision seeks 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

[228] Marco Tulio Ribeiro, Sameer Singh & Carlos Guestrin, "Why Should I Trust You?" Explaining the Predictions of Any Classifier, Proc. 22nd ACM SIGKDD Int'l Conf. on Knowledge Discovery & Data Mining 1135, 1135 (2016).

[229] Matthew J.B. Lawrence, Mandatory Process, 90 Ind. L.J. 1429, 1432 (2015).

[230] 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 Neurorobotics 1 (2018), https://www.frontiersin.org/articles/10.3389/fnbot.2018.00034/full [https://perma.cc/HL2V-5DUM].

"official exercise of discretion to mitigate a legal consequence that is otherwise a person's lawful fate."[231] 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 the most plausible berth.[232] Its presently penurious titration may be unwise. But it is still 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 non-instrumental terms. Instead, it may be better to focus on the tangible ways in which participation can alter outcomes in ways that lead us to a right to a human decision. 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 a particular individual's ability to supply information, and more generally to offer reasons, for a decision to be made in her favor to a human decision maker.[233] Recall that this is the force of Cathy O'Neil's anecdote about the welfare applicant Catherine Taylor, which I recounted earlier[234]: Only Taylor, we are to presume, had the information needful for a right resolution of her own case. And only human review could have elicited that information.

[231] Aziz Z. Huq, The Difficulties of Democratic Mercy, 103 Calif. 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 classifier has been used to impose a decision.

[232] 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).

[233] 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 assertions. To distinguish them here serves no useful purpose.

[234] See supra text accompanying notes 18–19.

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.[235] 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 respond directly to that decision by drawing it to a human's attention. Such a right might 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.[236] Or such a right to give reasons might also 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.[237] 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.[238] Systemic legitimacy, on the procedural justice view, flows from an embedding of the opportunity to embed reasons in the texture of citizen interactions with the legal system.[239]

Arguments of either form might also seek doctrinal footing in the Due Process Clause.[240] The requirement of a hearing, to be sure, does not

[235] See supra text accompanying notes 173–86.

[236] Domingos, supra note 90, at 79; see also LeCun et al., supra note 39, at 436 (discussing feature selection in algorithmic design).

[237] For an instrumental argument to this effect sounding in behavioral economic terms, see Lawrence, supra note 229, at 1432 ("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.").

[238] Tom R. Tyler & Yuen J. Huo, Trust in the Law: Encouraging Public Cooperation with the Police and Courts 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).

[239] 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." (emphasis omitted)).

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

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. Yet individuals have no entitlement to individualized participation in respect to them.[241] The boundary line between permissible legislative fiat (where due process does not apply) and forms of government action to which due process does attach is not wholly clear.[242] 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,"[243] 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[244]: its basis is not quite that the machine itself is inaccurate overall, but rather that the addition of human review can eliminate a particular class of false negatives (positives) by leveraging private information held by regulated subjects. For example, false negatives (positives) might be generated by flawed records or erroneous information in an otherwise accurate database—as in the case of a machine welfare denial documented by Cathy O'Neil.[245] The fact that state agencies may have financial incentives to maintain highly flawed records (or at least not to correct them) and may face interest-group pressure to precipitously adopt deeply flawed machine-learning systems only adds to this argument's appeal.

Without denying for a moment the normative concerns raised by cases such as Catherine Taylor's, I want to push back on the idea that a right to a human decision can be grounded on this argument for reason giving. In developing this point, 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 may well, in

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

[242] Friendly, supra note 69, at 1276–77 ("[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.").

[243] Adrian Vermeule, Conventions of Agency Independence, 113 Colum. L. Rev. 1163, 1213 (2013).

[244] See supra Section III.A.

[245] See O'Neil, supra note 18, at 152–53.

the long term, shape both desirable and undesirable outcomes. I also want to keep in view the possibility that there is a better alternative vehicle for addressing the normative concerns raised by Catherine Taylor's case—a possibility to which my conclusion returns.

A first, concededly tenuous, reason for hesitation is doctrinal. It is not clear that the Due Process Clause requires a supplemental human action to verify the sufficiency of machine decisions. The adjudicative forms that due process can take are often desultory.[246] 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, might warrant relatively little weight. Due process jurisprudence, after all, might simply be wrong and in need of updating in light of technological change.

A second reason for rejecting the argument from reason giving 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.[247] Depending on the empirics of the situation, the addition of a human decision may become a form of "undue process" that traduces "constitutionally mandated ceilings on government process."[248] This second point 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."[249] Where the addition of a procedural step has the expected systemic effect

[246] 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).

[247] 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.

[248] 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.

[249] Friendly, supra note 69, at 1276.

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.

To see why a right to a human decision might be "undue process," consider the effect of a backstop human decision maker for all the outputs of a machine decision-learning process in terms of net false positive and false negative rates.[250] Of course, "even well-designed" algorithmic tools will make mistakes.[251] 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.[252] And it is not safe to assume that a human will be able to identify and correct all of the instances in which the machine erred, while also not generating new errors. To the contrary, there is a real possibility that human input will lead to a *higher* net error rate. Further, if a human-decision right is installed 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.

There is a recent and familiar instance in which a non-machine algorithm was supplemented with an ex post right of human review—to dismaying systemic results. Starting in 2005, 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.[253] Inter-judge sentencing disparities subsequently sharply increased, in some jurisdictions almost doubling.[254] So too did racial

[250] Recall that this was the position of the Wisconsin Supreme Court in respect to algorithmic sentencing tools. State v. Loomis, 881 N.W.2d 749, 760 (Wis. 2016); supra text accompanying notes 72–75.

[251] Kroll, The Fallacy of Inscrutability, supra note 150, at 11.

[252] See supra text accompanying notes 136–38.

[253] United States v. Booker, 543 U.S. 220, 233, 248–49 (2005) (invalidating mandatory force of Federal Sentencing Guidelines); see also Gall v. United States, 552 U.S. 38, 49 (2007) (rejecting heightened appellate review for out-of-guidelines sentences); Kimbrough v. United States, 552 U.S. 85, 109–11 (2007) (rejecting statutory constraints on sentencing); Rita v. United States, 551 U.S. 338, 350–55 (2007) (allowing appellate presumption of reasonableness for within-Guidelines sentences).

[254] Ryan W. Scott, Inter-Judge Sentencing Disparity After *Booker:* A First Look, 63 Stan. L. Rev. 1, 4–5, 52 (2010) (similar result for a Massachusetts district court); 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*

disparities, at least according to some studies.[255] 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.[256] Adding 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 focusing on a specific person who has been wrongly denied a benefit have the potential to mislead[257]: 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.

Sentencing is no isolated case. To the contrary, there is a bushel of cases in which "professional overrides *decrease* accuracy in predicting reoffending, compared to unadjusted actuarial estimates."[258] A recent study of probation officers found that human overrides of an actuarial assessment tool decreased the overall accuracy of the system.[259] Another study of overrides by probation officers of the Post-Conviction Risk Assessment Instrument reached a similar finding.[260] And a 2005 meta-

sentencing, with a defendant potentially receiving a six-month longer sentence if assigned by happenstance to a harsh judge").

[255] 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*).

[256] I do not mean to suggest that the pre-*Booker* regime was without faults. Inter-judge 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.

[257] See supra note 18 and accompanying text.

[258] Sharad Goel et al., The Accuracy, Equity, and Jurisprudence of Criminal Risk Assessment (Dec. 26, 2018) (manuscript at 3), https://papers.ssrn.com/sol3/papers.cfm?-abstractid=3306723 [https://perma.cc/L7SJ-M6D9].

[259] Jean-Pierre Guay & Geneviève Parent, Broken Legs, Clinical Overrides, and Recidivism Risk: An Analysis of Decisions to Adjust Risk Levels with the LS/CMI, 45 Crim. Just. & Behav. 82, 94–96 (2018). For a now classic treatment, see Paul E. Meehl, Clinical Versus Statistical Prediction: A Theoretical Analysis and a Review of the Evidence 119–20, 136–38 (1954) (predicting that mechanical predictive methods would outperform clinical ones).

[260] Thomas H. Cohen, Bailey Pendergast & Scott W. VanBenschoten, Examining Overrides of Risk Classifications for Offenders on Federal Supervision, 80 Fed. Prob. 12, 18–19, 21 (2016).

study of sex offender recidivism algorithms identified actuarial predictions as the most accurate measures available.[261] Across many different contexts, therefore, there is no *empirical* reason to expect that addition of a human override will increase overall accuracy. To the contrary, human review seems generally to increase net error rates.

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.[262] This asymmetry could be sharpened. The algorithm could be used to limit the cost of human review by isolating a subset of individuals who might plausibly offer salient, new facts that could result in a different outcome. For instance, an algorithm that generated a risk parameter as a continuous variable could have a numerical threshold as a classification rule.[263] Individuals classified as exceeding that threshold by a small margin might be allowed to appeal. Those who cleared the threshold by a large margin, in contrast, would have no entitlement to a human decision through an appeal. 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 am skeptical of even this refinement of the argument from reason giving for three reasons. *First*, the argument for asymmetrical and narrow-gauged ex post human review still assumes that human review will correct false positives and only false positives—which, we have already seen, is no sure thing. Moreover, it assumes that only meritorious subjects of an adverse decision will appeal. But this is not likely. The class of Catherine Taylors, that is, will not all get relief, while a class of undeserving beneficiaries who have correctly had their claims denied—call them Elizabeth Taylors—will prevail. The latter as much as the

[261] R. Karl Hanson & Kelly E. Morton-Bourgon, The Characteristics of Persistent Sexual Offenders: A Meta-Analysis of Recidivism Studies, 73 J. Consulting & Clinical Psychol. 1154, 1155, 1158 (2005).

[262] See supra note 18 and accompanying text.

[263] For a useful exposition how prediction using a cut-off with a continuous variable might look, see Camelia Simoiu, Sam Corbett-Davies & Sharad Goel, The Problem of Infra-Marginality in Outcome Tests for Discrimination, 11 Annals Applied Stat. 1193 (2017).

former will select into the override procedure in ways that generate new errors.

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.[264] The prospect that this results in less errors overall is slim. Indeed, different rates of falsification should be expected among the Elizabeth Taylors as opposed to 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, moreover, fantastical to think that a human decision maker has costless and frictionless mechanisms for sorting the earnest Catherines from the Machiavellian Elizabeths. What seems from O'Neil's threshold example to be a simple, relatively costless step turns out on inspection, therefore, to be highly problematic.[265]

Instead, 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. I rather

[264] 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 237–38 (1994) (describing Spence's signaling game and the resulting pooling and separating equilibria). Spence's signaling model here is useful insofar as it introduces the idea of both pooling and separating equilibria arising in a situation in which one actor is trying to select among a population.

[265] 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 151, at 11 n.61 (citing Jatinder Singh et al., Responsibility & Machine Learning: Part of a Process (Oct. 27, 2016) (manuscript at 4), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2860048 [https://perma.cc/YF6Q-Z9JU]). This machine produces "a tamper-evident record that provides non-repudiable evidence of all nodes' actions." Haeberlen, Kouznetsov & Druschel, supra note 151, 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.

suspect that some of the appeal of a right to a human decision trades on an elision of these costs.

Second, it will often be the case that even an asymmetrical, narrow-gauged right to human decision will either be otiose from the start or else can be rendered irrelevant by certain forms of machine 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 open-ended standards. The most familiar example is numerical speed limits (e.g., 50 miles per hour) versus commands to "drive reasonably." The latter allow for consideration of enumerated and unanticipated factors.[266] 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 even matter. The very fact of selecting a rule rather than a standard as the relevant law forecloses an 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 parameter values.[267] These sub-rules 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.[268]

[266] Kaplow, supra note 188, at 559–63 (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).

[267] 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, 349–50 (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). On the possibility of new rules emerging through algorithmic interaction with new data, a method called reinforcement learning, see Mnih et al., supra note 130.

[268] 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. at 167–68. 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

The interaction of closed legal rules and function selection may therefore render otiose any claim to a human decision (or, at a minimum, make a right to a dynamic rather than a static algorithm more plausible).

Third, recall that I have so far assumed that the parameters of the training data were selected precisely because they enable a "faithful measurement" of an underlying property of interest relevant to the prediction of the target variable.[269] The participation-based argument for a human decision in effect may assume this is not so. That is, it assumes that the training data's parameters are insufficiently numerous 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 feature set. Retail responses, such as a right to an ex post human decision, perversely maintain a deficient status quo. In this fashion, they may even delay the implementation of holistic, systemic fixes achievable through better machine decision making. Correcting individual wrongs today, therefore, need not alleviate justice in the long run.

This last point can be generalized. 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 with respect to specific classification decisions is a well-fitting solution in the long term (even if they might be a possible diagnostic tool for sniffing out systemic problems along with audits). To the contrary, a range of static oversight tools—which focus on the algorithm's underlying source code—and dynamic oversight instruments—which look at the algorithm's behavior in the wild—are alternatively available,[270] and over

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 fairer 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. This 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.

[269] 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).

[270] Kroll et al., supra note 2, at 647–52 (describing static and dynamic testing protocols); see also Wachter, Mittelstadt & Russell, supra note 153, at 843–44 (describing the use of counterfactuals to conduct tests of algorithmic integrity).

time these tools might conduce to systematic improvements that are likely to render a right to human review superfluous. 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.

* * *

So what might be said to the Catherine Taylors who have been erroneously classified by a machine?[271] To begin, we should recall 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 most pertinent question to ask is whether Taylor was classified by a flawed system, not whether in her case an error was made. And then it is far from clear that addition of a human decision maker would reduce the net volume of errors. Evidence suggests it will likely do the opposite. The flawed quality of a machine decision does not imply that a human decision maker would do better. Nor will a human decision maker better serve a putative non-instrumental interest in participation.

Where an algorithmic tool is flawed, therefore, it does not follow that ex post human review is "due." Rather, there is every reason to believe that what is "due" is a better machine decision rather than a reliably unreliable human one—as I explore briefly in the conclusion. So nothing about my analysis here implies that some ex post verification or audits of such computational tools to which Taylor was subject are unwarranted. My aim is, more narrowly, to reject *retail* as opposed to *systemic* human review as a warranted intervention. Since the right to a human decision in its GDPR Article 22 form, as well as in its less articulate variants, is a retail right that attaches to distinct individuals subject to algorithmic classification, a case for systemic review simply cannot supply it with a normative foundation.

[271] See supra text accompanying notes 18–19.

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 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,[272] 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.[273] At least notionally, state action based on impermissible grounds is unlawful even if the same action on different motivational grounds would not be disallowed.[274] 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 classifier-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 a minimal 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

[272] 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).

[273] For examples of decisions disallowing state action that would be permitted in the absence of impermissible considerations, see, e.g., Parents Involved in Cmty. Schs. v. Seattle Sch. Dist. No. 1, 551 U.S. 701, 704 (2007); see also Gratz v. Bollinger, 539 U.S. 244, 270 (2003) (describing the use of such classifications as "pernicious" (citation omitted)).

[274] 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").

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 upon public reasons is deeply entrenched in Anglo-American law. Giving reasons, on this account, is "a way of showing respect for the subject, and a way of opening a conversation rather than forestalling one."[275] 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."[276] A decision, on this view, is consistent with the rule of law only if it is "comprehensible for those subject to the decision."[277] From a mid-century liberal perspective, state power is legitimate when based on "grounds of adequate neutrality and generality."[278] 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."[279]

Let us set aside the possibility that algorithms can be designed to issue explained decisions.[280] 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 it 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

[275] Frederick Schauer, Giving Reasons, 47 Stan. L. Rev. 633, 658 (1995).

[276] Frank I. Michelman, Formal and Associational Aims in Procedural Due Process, 18 Nomos 126, 127 (1977) (emphasis omitted).

[277] Mireille Hildebrandt, Algorithmic Regulation and the Rule of Law, Phil. Transactions Royal Soc'y A 1, 3 (2018) (emphasis omitted).

[278] Herbert Wechsler, Toward Neutral Principles of Constitutional Law, 73 Harv. L. Rev. 1, 15 (1959); accord Lon L. Fuller, The Morality of Law 34–38 (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, 782–85 (1983).

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

[280] For a survey of issues raised by the question of explainability, see Amina Adadi & Mohammed Berrada, Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI), 6 IEEE Access 52138, 52148–49 (2018).

rendered."[281] 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.[282] Indeed, it is now conventional wisdom that legislatures often enact statutory text without reaching a consensus view of the meaning of certain clauses or sentences. Ambiguous statutory text—the daily fare of appellate courts—arises because "Congress had no particular intent on the subject."[283] 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 non-mandatory appeal, or the Supreme Court denies certiorari, it is hardly clear there are articulable, let alone well-grounded, reasons for the action.[284] 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.

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."[285] In this delimited 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 far

[281] Lon L. Fuller, The Forms and Limits of Adjudication, 92 Harv. L. Rev. 353, 387 (1978).

[282] Schauer, supra note 275, at 636.

[283] Antonin Scalia, Judicial Deference to Administrative 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).

[284] 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 Legal Theory 69, 91–92 (2007) (grounding this conclusion on a comparison of individual and state responsibility). It may be a fair implication of his account that the reasoned quality vel non of state action is irrelevant.

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

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.[286] Human intentions necessarily guide the choice between supervised and unsupervised models; the process of feature selection; the selection of training data; and the ongoing process of refinement and calibration toward an optimal classifier.[287] 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 pre-commitment to generality and as a safeguard against personalistic or arbitrary state action.[288] 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 (re-)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,

[286] See supra text accompanying notes 172–92.

[287] On this iterative process of algorithmic improvement, see Bengio, supra note 104, at 50–51; see also Amershi et al., supra note 180, at 106.

[288] Cf. Schauer, supra note 275, at 651–52 (describing reason giving as a pre-commitment mechanism that yields generality).

"as an individual."[289] Action based on traits shared by a larger social group ipso facto fails to take that person seriously as an individual.

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.[290] 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, behaviour, location or movements."[291] 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 so as to generate predictive classifications of individuals. 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 specific kinds of statistical evidence to demonstrate the likelihood of defendant responsibility in tort cases.[292] Instead, courts call for "individualized evidence."[293] Both it and the 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.[294] The German Constitutional Court, for example, has invoked the

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

[290] GDPR, supra note 20, art. 22. Article 22 does not prohibit profiling: it prohibits certain "decision[s]" based on profiling.

[291] Id. art. 4(4).

[292] Courts have sometimes said that merely "mathematical chances," Smith v. Rapid Transit Inc., 317 Mass. 469, 470 (1945), or "[q]uantitative probability," Day v. Bos. & Me. R.R., 96 Me. 207, 217 (1902), are never sufficient for the imposition of tort liability. For a more detailed account on the informational problem of adjudication, see Rebecca Haw Allensworth, Note, Prediction Markets and Law: A Skeptical Account, 122 Harv. L. Rev. 1217, 1228–29 (2009).

[293] Judith Jarvis Thomson, Liability and Individualized Evidence, 49 Law & Contemp. Probs. 199, 203–05 (1986).

[294] Immanuel Kant, Foundations of the Metaphysics of Morals and What Is Enlightenment? 47 (L. Beck trans., Macmillan Publishing Co. 1959) (1785) ("Act so that you treat humanity, whether in your own person or in that of another, 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

German Basic Law's commitment to human dignity to hold impermissible state action that "verdinglicht und zugleich entrechtlicht" ("treated as objects and at the same time deprived of their rights"), and more specifically to invalidate a provision in the 2004 Air Transport Security Act that allowed a hijacked plane to be shot down under certain circumstances.[295] 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."[296] 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 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.[297]

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."[298] 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

to work to justify the right to bare participation. Edmund L. 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, 343–44 (Dennis Patterson ed., 2d ed. 1999) (situating this idea in the history of German legal theory).

[295] Bundesverfassungsgericht [BverfG] [Federal Constitutional Court], Mar. 13, 2006, 59 Neue Juristische Wochenschrift [NJW] 751, 753, 758, 2006, *translated in* BVergG, 1 BvR 357/05, Feb. 15, 2006, https://www.bundesverfassungsgericht.de/SharedDocs/Entscheidungen/EN/2006/02/rs20060215_1bvr035705en.html [https://perma.cc/W857-5CU8]. My thanks to Annette Zimmermann for discussion and help with understanding this phrase in the context of German law.

[296] Richard J. Arneson, Luck Egalitarianism and Prioritarianism, 110 Ethics 339, 339 (2000).

[297] Both problems are delineated in Samuel Scheffler, What is Egalitarianism?, 31 Phil. & Pub. Aff. 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).

[298] Lippert-Rasmussen, supra note 289, at 54.

draft for the military—will often be morally compelling.[299] Indeed, "even good judgment" is often predicated on non-spurious generalizations.[300] Hence, the bare claim that a decision is morally flawed because it is predicated on population-wide data rather than individualized evidence is untenable.

I think a subtler 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[301] or vaccine regimes for school-age children,[302] why should it be inadequate 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.[303]

[299] Annabelle Lever, Why Racial Profiling Is Hard to Justify: A Response to Risse and Zeckhauser, 33 Phil. & Pub. Aff. 94, 110 (2005).

[300] Frederick Schauer, Profiles, Probabilities, and Stereotypes 215 (2003).

[301] See, e.g., Act of Nov. 8, 1984, ch. 179, 1984 N.J. Laws 948 (last amended by Act of Jan. 18, 2010, ch. 318, 2009 N.J. Laws 2339).

[302] For instance, both the District of Columbia and Virginia mandate by statute the Human Papillomavirus vaccine for school-age girls. See D.C. Code § 7-1651.04 (2007); Va. Code Ann. § 32.1-46 (2016).

[303] See Thomson, supra note 293, 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. To adopt a phrase developed by the philosopher Martin Smith, there are certain limited instances in which individualized evidence "normic[ally] support[s]" a conclusion for which it is proffered,[304] and hence is relevant to a legal decision. 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, my friend walks past my computer and sees a blue screen. My friend's belief that the screen is blue is "normic[ally] support[ed]" by her perception; my analog belief that the screen was blue is not.[305] 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.[306] 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."[307] She should, perhaps, have her vision checked for color blindness.

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.[308] Where liability is imposed on the basis of counterfactually 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 some sorts of historical knowledge.[309] When a machine decision relies on population-wide evidence, therefore, there is a loss of deterrence effect.

[304] Martin Smith, What Else Justification Could Be, 44 Noûs 10, 13–14 (2010).

[305] Id. (offering a more complex version of this hypothetical).

[306] David Enoch, Levi Spectre & Talia Fisher, Statistical Evidence, Sensitivity, and the Legal Value of Knowledge, 40 Phil. & Pub. Aff. 197, 209–10 (2012).

[307] Id. at 209.

[308] Id. at 218–19.

[309] There is a literature on whether probabilistic evidence can be a basis for knowledge or rational belief. See, e.g., Henry E. Kyberg, Jr., 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.

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.[310] 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 decision makers 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.[311] Not all concern blame, and if 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 and Negative Externalities

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 decisions 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

[310] Enoch, Spectre & Fisher, supra note 306, at 215 (noting that blame may also require counterfactually sensitive evidence).

[311] See supra text accompanying notes 117–21.

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 systemic 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 a judicially enforced threshold of liability.[312] 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 planes capable of exercising deadly force on a distant battlefield.[313] A thorough excision of the human role in this context raises difficult ethical issues.[314] If it were the case that maintaining a human role led to a lower activity level, without serious cost 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."[315] In effect, humans adopt a heuristic of reliance upon automated decisions "as a replacement for more vigilant system monitoring or decision making."[316] 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

[312] 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).

[313] For an acute description, see Hugh Gusterson, Drone: Remote Control Warfare 2–25 (2016).

[314] 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).

[315] Linda J. Skitka et al., Automation Bias and Errors: Are Crews Better Than Individuals?, 10 Int'l J. Aviation Psychol. 85, 86 (2000).

[316] Linda J. Skitka et al., Does Automation Bias Decision-Making?, 51 Int'l J. Human-Computer Stud. 991, 992 (1999).

expertise and objectivity that infuses that technology.[317] However forceful this concern may be—there is no strong empirical evidence in the machine-learning context to substantiate the concern available as of yet—there are likely a number of ways to ensure against complacent reliance on automated decision makers. Not least, one could resort to 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. It may well be that individuals provide too erratic and uncertain a safeguard, such that an alternative institutional check is wise.

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 require 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 for conditions of participatory democracy, if machine decisions are used to shape political preferences. Alternatively, they might enable new, highly intrusive forms of regulation inconsistent 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 Part II and earlier in this Part. A right to a human decision is best thought of as a response to specific technologies, and these instruments do generate troublesome asymmetries between persons and concentrated organizational power. But there is no reason from this motivation alone to simply assume that the right helps mitigate those harms. The problem is that it requires heroic

[317] 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 Design 17, 18–20 (2017).

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 alleviate the underlying 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.[318] Recent experience with individual entitlements to privacy in the social media and internet platform contexts also furnishes cause for pessimism.[319] 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 or how much they would pay to protect otherwise public information.[320] They also appear to have time-inconsistent preferences, in the sense that they are willing to accept low rewards now in exchange for a possible "permanent negative annuity in the future."[321]

If a right to a human decision is not necessarily the best instrument 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.[322] One might also think of different 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 do not think such evidence exists—we should not precipitously conclude that an

[318] 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).

[319] See, e.g., Alessandro Acquisti, Leslie K. John & George Loewenstein, What Is Privacy Worth?, 42 J. Legal Stud. 249, 250–51 (2013) (identifying sensitivity of privacy-related preferences to subtle contextual cues).

[320] Id. at 249–51.

[321] Alessandro Acquisti & Jens Grossklags, Privacy and Rationality in Individual Decision Making, 3 IEEE Security & Privacy 26, 31 (2005); accord 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, 220 (2015).

[322] 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).

individual right provides an effectual solution to a structural and systemic dysfunction.[323]

Finally, and related to the concern about power, a human decision may be preferred to a machine decision in order to pursue a more general 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 details of institutional design.

A human decision might be preferred, for example, because it maintains the open-endedness of legal criteria, or because it injects 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 still 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 lawmakers 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. 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 in a common-law fashion. A machine-learning tool, that is, might refine its classification rule over time, but this will not necessarily yield detailed new legal guidance for primary conduct. Again, this argument for human decision making is contingent on the absence of other platforms for refinement and publication of new, more detailed rules for primary conduct.

[323] I develop further possible strategies in Huq, supra note 33 (manuscript at 42–47).

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, however, will depend on other elements of institutional design. It is poorly described as an individual "right."

E. Practical Constraints on Machine Decisions

In my view, efforts to derive a right to a human decision from normative first principles do not succeed, despite the unease that fully automated decision making provokes in many minds.[324] Yet this does not mean that machines can or should always displace human action. As Section II.A explained, machine learning is not an all-purpose tool. It excels at empirical predictions when sufficient training data is available. There is no reason to think that it can effectively make causal decisions now or, more importantly, decisions with an ethical or other normative element (now, or, indeed, any time in the foreseeable future). From these technical bounds on machine decisions emerge a corresponding set of guide rails for the use of such instruments.

To begin, machine decisions are inappropriate when there is insufficient historical data or no tractable parameter that can be predicted. The state may wholly lack the necessary training data to measure the variable of interest, the available data may be misleading, or there may simply be no sound conceptualization of the ultimate result of interest. An objection in this vein has been leveled at criminal risk assessment tools, on the ground that there is no good measure of propensity to violence.[325] A similar argument has been leveled against machine-based teacher evaluation tools, on the ground that there is no well-defined parameter that elicits a useful measure of teachers' contributions and can be used as an outcome variable when training a machine tool.[326]

Finally, machine decisions are not presently appropriate for decisions with ethical or normative components. Many legal and even factual questions resolved through civil and criminal adjudication have some such component. To be sure, some have entertained the prospect of

[324] Binns et al., supra note 7, at 9.
[325] See Technical Flaws, supra note 210, at 2–3.
[326] O'Neil, supra note 18, at 135–40 (critiquing existing models of teacher evaluation).

automated adjudication.[327] But the most sophisticated of these, an argument for machine-tooled "micro-directives" forcefully pressed by Anthony Casey and Anthony Niblett, assumes that the law pursues a single goal (such as efficiency) and lacks normative multi-criteriality.[328] At least in public law, there are few issues with this complexion. Rather, the presence of normative shades in many matters presently resolved through adjudication suggests that—at least until it is possible to settle morals by machine—many functions that adjudication presently plays cannot be performed by machine.

CONCLUSION: A RIGHT TO A WELL-CALIBRATED MACHINE DECISION?

There are reasons aplenty to be cautious about the avulsive advance of new algorithmic technologies. These range 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 Article 22. At least where machine decisions are plausibly employed, there is no reason to establish a countervailing right to a human decision. Of course, there is always a risk that machines will be deployed where their particular strengths do not fit, above and beyond the risks of negligent or discriminatory design choices. But avoiding those possibilities simply requires a modicum of sensitivity to the capabilities and limits of machine learning. It does not compel technological abstinence.

In concluding, let me offer what I hope is a provocative, albeit tentative, thought: where machines are appropriate decision makers, there is no reason to relax one's guard against the manifold ways in which machine learning can go awry. Rather than thinking about a right to a human decision, however, 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, and I leave for another day a more fulsome treatment.[329]

[327] See sources cited supra note 123.

[328] Casey & Niblett, supra note 189, at 1419 (hinting at efficiency as a relevant single criterion).

[329] For amplifications of the points raised in this conclusion, see Huq, supra note 33.

An account of the right to a machine decision would begin with the observation that while machine-learning tools have the capacity to improve on humans' accuracy and neutrality, many of those now implemented by government are highly flawed.[330] Even if this does not impel a reversion to (equally flawed) human decision making, the legal system should incentivize the correction of such errors. Its dynamic goal should be a machine decision well-calibrated in light of constitutional concerns. Most basically, an algorithmic tool is well-calibrated if it does not rely on flawed training data and otherwise meets common standards of industry performance. More work, however, needs to be done to describe the circumstances in which algorithms are in compliance with due process, privacy, and equality norms.

In the end, this well-calibrated machine decision maker may have underappreciated advantages that sound in dignity and autonomy terms. Consider the possibility of dignity 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."[331] 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.[332] It is quite possible that an algorithmic interface for welfare recipients—who are a notoriously stigmatized group[333]—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.

Algorithmic technologies used by machine decisions are still in their infancy. Now, they can be flawed in many ways. It seems too early,

[330] Whittaker et al., supra note 34, at 18–22.

[331] Clive Thompson, May A.I. Help You?, N.Y. Times (Nov. 18, 2018), https://www.nytimes.com/interactive/2018/11/14/magazine/tech-design-ai-chatbot.html [https://perma.cc/G7GP-MNTK].

[332] How the Internet Has Changed Dating, Economist (Aug. 18, 2018), https://www.economist.com/briefing/2018/08/18/how-the-internet-has-changed-dating [https://perma.cc/SA9W-X5VN].

[333] 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).

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.