NOTES

BEYOND INTENT: ESTABLISHING DISCRIMINATORY PURPOSE IN ALGORITHMIC RISK ASSESSMENT

Equal protection doctrine is bound up in conceptions of intent. Unintended harms from well-meaning persons are, quite simply, nonactionable.[1] This basic predicate has erected a high bar for plaintiffs advancing equal protection challenges in the criminal justice system. Notably, race-based challenges on equal protection grounds, which subject the state to strict scrutiny review, are nonetheless stymied by the complexities of establishing discriminatory purpose. What's more, the inability of courts and scholars to coalesce on a specific notion of the term has resulted in gravely inconsistent applications of the doctrine.[2]

States' increasing use of algorithmic systems raises longstanding concerns regarding prediction in our criminal justice system — how to categorize the dangerousness of an individual, the extent to which past behavior is relevant to future conduct, and the effects of racial disparities.[3] Integration of algorithmic systems has been defined by ambivalence: while they are touted for their removal of human discretion,[4] they also readily promote and amplify inequalities — for example, through their consideration of protected characteristics and their interaction with existing systems tainted by legacies of inequality.[5] Furthermore, algorithms, especially those incorporating artificial intelligence (AI), may operate in ways that are opaque, unpredictable, or not well understood.

---

[1] *See* Washington v. Davis, 426 U.S. 229, 239–40 (1976).

[2] *See* Aziz Z. Huq, *What Is Discriminatory Intent?*, 103 CORNELL L. REV. 1211, 1240–63 (2018) (presenting five definitions from the literature).

[3] *See* Sonia K. Katyal, *Private Accountability in the Age of Artificial Intelligence*, 66 UCLA L. REV. 54, 58 (2019) ("The idea that algorithmic decisionmaking, like laws, are [sic] objective and neutral obscures . . . the causes and effects of systematic and structural inequality, and thus risks missing how AI can have disparate impacts on particular groups."); Sandra G. Mayson, *Bias In, Bias Out*, 128 YALE L.J. 2218, 2224 (2019) (arguing that the problem of disparate impact stems from "the nature of prediction itself"). *See generally* Barbara D. Underwood, *Law and the Crystal Ball: Predicting Behavior with Statistical Inference and Individualized Judgment*, 88 YALE L.J. 1408, 1409 (1979) (discussing promising and concerning aspects of prediction).

[4] *See* Mayson, *supra* note 3, at 2280.

[5] *See generally* VIRGINIA EUBANKS, AUTOMATING INEQUALITY: HOW HIGH-TECH TOOLS PROFILE, POLICE, AND PUNISH THE POOR (2018); CATHY O'NEIL, WEAPONS OF MATH DESTRUCTION: HOW BIG DATA INCREASES INEQUALITY AND THREATENS DEMOCRACY (2016); Ifeoma Ajunwa, *The Paradox of Automation as Anti-bias Intervention*, 41 CARDOZO L. REV. 1671 (2020); Solon Barocas & Andrew D. Selbst, Essay, *Big Data's Disparate Impact*, 104 CALIF. L. REV. 671 (2016).

As commentators have demonstrated, intent-based notions inherent in equal protection jurisprudence are ill-suited to artificial intelligence.[6]

In response, scholars have presented a host of proposals to address the use of race by algorithmic systems, including colorblindness, affirmative action, outright prohibition, and the extension of effects-based statutory regimes.[7] Many of these suggestions entail a sharp departure from the current discriminatory purpose requirement in favor of an effects-based framework attuned to the need to provide redress for unjust treatment.[8] A doctrinal shift of this nature would be normatively beneficial, but is unlikely in the near future given the composition of the Supreme Court.[9]

This Note argues that transitioning to a test rooted primarily in evaluation of effect for equal protection challenges to the use of algorithmic risk assessment systems (RASs) would not present as significant a shift as described. Courts already rely extensively on effects when determining whether a discriminatory purpose exists in jury selection and voting cases.[10] This Note proposes that the Supreme Court extend its current jurisprudence in these contexts to the use of algorithmic RASs in sentencing. Part I describes algorithmic RASs and how current intent-based notions are ill-suited to them. Part II illustrates how equal protection doctrine may incorporate primary evaluation of effects for algorithms. Part III demonstrates how an effects-plus framework may resolve equal protection challenges to algorithmic RASs. A brief conclusion follows.

---

6 *See* Aziz Z. Huq, *Racial Equity in Algorithmic Criminal Justice*, 68 DUKE L.J. 1043, 1083 (2019); Yavar Bathaee, *The Artificial Intelligence Black Box and the Failure of Intent and Causation*, 31 HARV. J.L. & TECH. 889, 893 (2018).

7 *See* Crystal S. Yang & Will Dobbie, *Equal Protection Under Algorithms: A New Statistical and Legal Framework*, 119 MICH. L. REV. 291, 346–50 (2020) (proposing two statistical approaches to purge race effects of algorithms); Jacob D. Humerick, Note, *Reprogramming Fairness: Affirmative Action in Algorithmic Criminal Sentencing*, 4 COLUM. HUM. RTS. L. REV. ONLINE 213, 239–40 (2020); Charles R. Lawrence III, *The Id, the Ego, and Equal Protection: Reckoning with Unconscious Racism*, 39 STAN. L. REV. 317, 324 (1987) (proposing a test that would evaluate whether an action "conveys a symbolic message to which the culture attaches racial significance"); Huq, *supra* note 6, at 1128–33 (arguing for evaluation of algorithms to account for impact on racial stratification); Bathaee, *supra* note 6, at 894 (suggesting the "modif[ication of] intent and causation tests with a sliding scale based on the level of AI transparency and human supervision"); Mayson, *supra* note 3, at 2262–81 (describing strategies).

8 *See* Michal Saliternik, *Big Data and the Right to Political Participation*, 21 U. PA. J. CONST. L. 713, 754–55, 755 n.223 (2019).

9 *See* Daniel B. Rice & Jack Boeglin, *Confining Cases to Their Facts*, 105 VA. L. REV. 865, 867 (2019) ("The Roberts Court . . . has tended to move the law in an incremental fashion, rather than to effect seismic doctrinal shifts all at once."); Hayden Johnson, Note, *Vote Denial and Defense: A Strategic Enforcement Proposal for Section 2 of the Voting Rights Act*, 108 GEO. L.J. 449, 462 & n.53 (2019) (noting that the Roberts Court "has heavily scrutinized and . . . disfavored disparate impact statutes," *id.* at 462).

10 *See* Daniel R. Ortiz, *The Myth of Intent in Equal Protection*, 41 STAN. L. REV. 1105, 1137 (1989).

I. WRESTLING A SQUARE PEG INTO A ROUND DOCTRINAL HOLE

A. Increasingly Artificial Governance

Algorithmic systems are already a consequential tool of the administrative, judicial, and carceral state. Their most controversial application is arguably in the criminal justice system, in which courts have upheld their use in risk assessment for policing, bail, charging, probation, sentencing, and parole determinations.[11] Throughout the country, RASs weigh factors correlated with risk to inform decisions relating to diverse categories of crimes.[12] For example, pretrial risk assessment relies on actuarial data to determine a defendant's risk of failing to appear in court and committing new criminal activity before trial.[13] Post-adjudication risk assessment evaluates factors related to recidivism to inform the imposition of sentences, supervision, and treatment.[14]

Despite their increasing use and significance, algorithmic systems are not widely understood. Algorithmic RASs commonly used in sentencing are simple from a technological standpoint;[15] most resemble automated checklists in that they estimate risk by assigning weight to a limited number of risk factors.[16] Nevertheless, the operation of today's RASs is practically inscrutable to defendants given access hurdles: such systems are often developed by private-sector companies that operate with limited

---

[11] *See, e.g.*, Malenchik v. State, 928 N.E.2d 564, 574–75 (Ind. 2010) (concluding that RASs could supplement judicial consideration at sentencing); People v. Younglove, No. 341901, 2019 WL 846117, at *3 (Mich. Ct. App. Feb. 21, 2019) (rejecting defendant's due process claim against the use of an RAS); State v. Guise, 921 N.W.2d 26, 29 (Iowa 2018) (same); Dorothy E. Roberts, *Digitizing the Carceral State*, 132 HARV. L. REV. 1695, 1716 (2019) (book review).

[12] *See* Doaa Abu Elyounes, *Bail or Jail? Judicial Versus Algorithmic Decision-Making in the Pretrial System*, 21 COLUM. SCI. & TECH. L. REV. 376, 389 (2020) ("[T]oday there are about sixty risk assessment tools used across the country [in the criminal justice system]."); Jessica M. Eaglin, *Constructing Recidivism Risk*, 67 EMORY L.J. 59, 114 (2017) (noting that a growing number of jurisdictions are requiring use of risk assessments during sentencing).

[13] PRETRIAL JUST. INST., PRETRIAL RISK ASSESSMENT: SCIENCE PROVIDES GUIDANCE ON ASSESSING DEFENDANTS 2–4 (2015), https://www.ncsc.org/__data/assets/pdf_file/0016/1654/pretrial-risk-assessment-science-provides-guidance-on-assessing-defendants.ashx.pdf [https://perma.cc/LW45-9ZM3].

[14] JENNIFER K. ELEK ET AL., NAT'L CTR. FOR STATE CTS., USING RISK AND NEEDS ASSESSMENT INFORMATION AT SENTENCING: OBSERVATIONS FROM TEN JURISDICTIONS 3–4 (2015), https://www.ncsc.org/__data/assets/pdf_file/0016/26251/final-pew-report-updated-10-5-15.pdf [https://perma.cc/HJ2Q-H2DQ].

[15] *See* Huq, *supra* note 6, at 1050, 1067.

[16] *See id.* at 1050; J. Stephen Wormith & James Bonta, *The Level of Service (LS) Instruments*, *in* HANDBOOK OF RECIDIVISM RISK/NEEDS ASSESSMENT TOOLS 117, 117 (Jay P. Singh et al. eds., 2018) [hereinafter RISK ASSESSMENT HANDBOOK]; Tim Brennan & William Dieterich, *Correctional Offender Management Profiles for Alternative Sanctions (COMPAS)*, *in* RISK ASSESSMENT HANDBOOK, *supra*, at 49, 49.

oversight and raise the shield of trade secret protection.[17] For example, the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS), an RAS in use in many jurisdictions, has a proprietary algorithm.[18]

Algorithms vary in their complexity and transparency: while structured algorithms may make carefully constrained decisions based on preprogrammed rules in a transparent manner,[19] machine-learning programs — which fall under the umbrella of artificial intelligence — learn from past data and experience, potentially through opaque processes with few or no rules constraining their ability to predict.[20] Some portend the wide use of technologically sophisticated AI processes,[21] which may consider myriad variables and evaluate not only risk but also associated social cost.[22] As a current example, although it primarily operates through standard regression models, COMPAS incorporates machine learning.[23]

---

[17] *See* Rebecca Wexler, *Life, Liberty, and Trade Secrets: Intellectual Property in the Criminal Justice System*, 70 STAN. L. REV. 1343, 1349–50 (2018); *see also, e.g.*, Hous. Fed'n of Tchrs., Loc. 2415 v. Hous. Indep. Sch. Dist., 251 F. Supp. 3d 1168, 1179 (S.D. Tex. 2017); State v. Loomis, 881 N.W.2d 749, 761, 769–71 (Wis. 2016). However, a New York court did recently compel disclosure. *See* Brennan Ctr. for Just. at N.Y. Univ. Sch. of L. v. N.Y.C. Police Dep't, No. 160541/2016, 2017 N.Y. Misc. LEXIS 5138, at *17 (N.Y. Sup. Ct. Dec. 22, 2017) (ordering disclosure of some output from predictive policing system).

[18] *Loomis*, 881 N.W.2d at 761. Some other systems currently employed by states are relatively open and transparent. *See, e.g.*, *Release Assessment*, N.Y.C. CRIM. JUST. AGENCY, https://www.nycja.org/release-assessment [https://perma.cc/VN6T-87UP]; *Prisoner Assessment Tool Targeting Estimated Risk and Needs (PATTERN) Interactive Tool*, URB. INST. (Sept. 4, 2019), https://apps.urban.org/features/risk-assessment [https://perma.cc/RQ5R-DM4R].

[19] THOMAS H. CORMEN ET AL., INTRODUCTION TO ALGORITHMS 5 (3d ed. 2009).

[20] Harry Surden, Essay, *Machine Learning and Law*, 89 WASH. L. REV. 87, 89 (2014); Melissa Hamilton, *The Biased Algorithm: Evidence of Disparate Impact on Hispanics*, 56 AM. CRIM. L. REV. 1553, 1558 (2019). Systems operating through deep neural networks may learn in complex ways that are multilayered and opaque. *See* L. Karl Branting, *Artificial Intelligence and the Law from a Research Perspective*, SCITECH LAW., Spring 2018, at 34–35. On the other hand, decision trees, another form of machine learning, may allow for visual representations of data in an intelligible way. *See* STUART J. RUSSELL & PETER NORVIG, ARTIFICIAL INTELLIGENCE 657 (4th ed. 2021). In general, the machine-learning approach involves (1) procuring an input (training) sample, (2) specifying an outcome for prediction, (3) selecting measurable variables (features), and (4) correlating features to predict the outcome. *See id.* Correlation does not imply a causal relationship. 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) (manuscript at 2), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3385594 [https://perma.cc/R6VA-QNY3].

[21] *See, e.g.*, Sandra G. Mayson, *Dangerous Defendants*, 127 YALE L.J. 490, 511 (2018); Cuéllar & Huq, *supra* note 20, at 2. But research suggesting advanced systems do not perform better than simple regressions counsels against their wide adoption. *See* Julia Dressel & Hany Farid, *The Accuracy, Fairness, and Limits of Predicting Recidivism*, SCI. ADVANCES, January 2018, at 3, https://advances.sciencemag.org/content/4/1/eaao5580.full [https://perma.cc/9EJS-MJ97] (finding that COMPAS's accuracy can be matched by a "simple linear classifier").

[22] *See* Huq, *supra* note 6, at 1067.

[23] Brennan & Dieterich, *supra* note 16, at 70–72; *cf.* Elyounes, *supra* note 12, at 423–24 (discussing an AI RAS developed by Professor Jon Kleinberg at Cornell University that is not yet in use).

The task of RASs, whether AI-enabled or not, in some ways is nothing different from what judges do every day.[24] However, their perceived scientific legitimacy may entrench their use, despite uncertainties and contrary evidence regarding their precision and prediction gains.[25] Moreover, algorithmic systems may be slower to change than human decisionmakers.[26] This resulting inertia is especially problematic when data the systems work from are distorted.

All of this is not to overlook the benefits of algorithmic systems in terms of consistency and uniformity.[27] Algorithmic RASs claim to be objective in reducing the opacity and subjectivity inherent in human decisionmaking, especially when issues of unconscious human bias are taken into account.[28] In part for these reasons, scholars note that subjective risk assessment, as performed by judges making decisions based on their past experience and judgment, has the potential to exacerbate inequalities associated with prediction.[29] A simple algorithm may at least explain itself through statistical measures of the correlations or weights associated with variables under consideration.[30]

B. The Standard Approach

Algorithmic systems do not easily fit the contours of modern equal protection doctrine, which provides two relevant paths for a plaintiff to establish a prima facie equal protection claim. First, in *Washington v. Davis*,[31] the Supreme Court established that constitutional equal protection challenges to government action yielding a disproportionate

---

[24] *See* Kevin R. Reitz, *"Risk Discretion" at Sentencing*, 30 FED. SENT'G REP. 68, 70 (2017) ("[P]rison sentence lengths in most U.S. jurisdictions are already based on predictions or guesses about offenders' future behavior, and this has been true — in multiple settings — for at least a century.").

[25] *See* Hamilton, *supra* note 20, at 1570 (demonstrating that "COMPAS does not predict as strongly for Hispanics"); Dressel & Farid, *supra* note 21, at 3 (showing that COMPAS "is no more accurate or fair than predictions made by people with little or no criminal justice expertise"); Elyounes, *supra* note 12, at 390–91 (discussing contrary findings of the validity and accuracy of RASs).

[26] *See* Huq, *supra* note 6, at 1066–67.

[27] *See* Mayson, *supra* note 3, at 2279–80.

[28] *See* Jon Kleinberg et al., *Discrimination in the Age of Algorithms*, 10 J. LEGAL ANALYSIS 113, 116 (2018) (noting that "people themselves may not know why and how they are choosing"); *cf.* Sonja B. Starr, *Evidence-Based Sentencing and the Scientific Rationalization of Discrimination*, 66 STAN. L. REV. 803, 804–05 (2014) (noting that many jurisdictions encourage judges to explicitly consider factors like socioeconomic status, gender, and neighborhood characteristics).

[29] *See* Mayson, *supra* note 3, at 2281 (arguing that algorithmic systems are more likely to be accountable than human decisionmakers); Mirko Bagaric et al., *Erasing the Bias Against Using Artificial Intelligence to Predict Future Criminality: Algorithms Are Color Blind and Never Tire*, 88 U. CIN. L. REV. 1037, 1064–66 (2020).

[30] *See* CHRISTOPHER SLOBOGIN, PRIMER ON RISK ASSESSMENT INSTRUMENTS FOR LEGAL DECISION-MAKERS 4, https://law.vanderbilt.edu/academics/academic-programs/criminal-justice-program/Primer_on_Risk_Assessment.pdf [https://perma.cc/KVQ4-YUHH] (explaining that adjusted actuarial RASs list factors used, how they should be measured, and a total score).

[31] 426 U.S. 229 (1976).

racial impact must first encounter the filter of discriminatory purpose.[32] A claim may pass upon only circumstantial evidence of effect plus intent to discriminate; the burden then shifts to the state to provide a nondiscriminatory justification.[33] Since *Davis*, the Court has rejected the incorporation of a pure effects test. In *Personnel Administrator of Massachusetts v. Feeney*,[34] it required a plaintiff to provide evidence of subjective intent to harm and rejected evidence that an action was taken "merely 'in spite of . . . ' its adverse effects upon an identifiable group."[35] Second, the *Feeney* Court laid out an alternate path that does not mandate a showing of discriminatory purpose: actions that are expressly conditioned on a racial classification or an "obvious pretext," "regardless of purported motivation," are "presumptively invalid," as they "in themselves supply a reason to infer antipathy."[36] The state may rebut with a sufficient nondiscriminatory rationale, although in practice, this presumption is hard to combat.[37]

The discriminatory purpose requirement has erected a practically insurmountable burden of persuasion for plaintiffs and produced a considerable chilling effect on equal protection claims.[38] It has accordingly engendered a host of criticisms: opponents have condemned the ease of hiding discriminatory motives and the myriad incentives to do so.[39] Scholars have emphasized that the element of intent does not align with modern conceptions of unconscious bias and racism.[40] They have also

[32] *Id.* at 233–36, 239–41, 245. The Court later explained that historical background, specific events leading up to a certain enactment, departures from normal procedures, and legislative or administrative history may evidence discriminatory purpose. Village of Arlington Heights v. Metro. Hous. Dev. Corp., 429 U.S. 252, 267–68 (1977).

[33] Hunter v. Underwood, 471 U.S. 222, 227–28 (1985) (applying heightened scrutiny once plaintiff had demonstrated racially discriminatory purpose and effect).

[34] 442 U.S. 256 (1979).

[35] *Id.* at 279; *see also Hunter*, 471 U.S. at 228 (explaining that a plaintiff must show that racial discrimination is a "substantial" or "motivating" factor).

[36] *Feeney*, 442 U.S. at 272.

[37] *See* Fisher v. Univ. of Tex. at Austin, 570 U.S. 297, 313 (2013) ("[T]he mere recitation of a 'benign' or legitimate purpose for a racial classification is entitled to little or no weight." (quoting City of Richmond v. J.A. Croson Co., 488 U.S. 469, 500 (1989))).

[38] K.G. Jan Pillai, *Shrinking Domain of Invidious Intent*, 9 WM. & MARY BILL RTS. J. 525, 538 (2001).

[39] Lawrence, *supra* note 7, at 319; *see* Charles R. Lawrence III, *Implicit Bias in the Age of Trump*, 133 HARV. L. REV. 2304, 2308–09 (2020) (reviewing JENNIFER L. EBERHARDT, BIASED: UNCOVERING THE HIDDEN PREJUDICE THAT SHAPES WHAT WE SEE, THINK, AND DO (2019)).

[40] *See* Reva Siegel, *Why Equal Protection No Longer Protects: The Evolving Forms of Status-Enforcing State Action*, 49 STAN. L. REV. 1111, 1131, 1134–36, 1141–45 (1997) (arguing that equal protection litigation employing a disparate impact standard would more successfully disestablish historic patterns of race stratification); Lawrence, *supra* note 7, at 321–22 ("Traditional notions of intent do not reflect the fact that decisions about racial matters are influenced in large part by factors that can be characterized as neither intentional . . . nor unintentional . . . ." *Id.* at 322.).

questioned the relevance of motive where disparate harm has been established.[41] Such concerns justify a shift to an impact standard.

C. The Doctrinal Challenge

Beyond the justifications for an impact standard presented in the prior section, the argument for abandoning evaluation of intent is pronounced in the case of algorithmic systems for several reasons.

1. The Lack of Intent. — The focus of the discriminatory purpose requirement on intent is inapposite to algorithmic systems. Algorithms do not possess intent.[42] Disparate impact rather stems from inevitable data biases or the intentional choices of their creators or the judges applying them.[43] However, one may not discern relevant intent by evaluating the motives of either human creator or judge because neither is sufficiently responsible for the decisions of algorithmic RASs. Relevant decisionmaking for equal protection challenges to RASs includes the features selected and factors weighed by an algorithm, which are outside the control of a human judge. The intent of human creators manifests in decisions regarding the selection of training data, the definition of the optimization problem, and, in some cases, the application of features in the optimization problem.[44] However, with complex AI systems, designers may not be responsible for the exact ways in which their tools operate or capable of explaining a system's choices.[45]

2. The Obscuring Effect of Proxies. — Determining whether a factor is "an obvious pretext" for a protected characteristic is a hard problem. Algorithmic systems largely avoid the sort of overt consideration of protected characteristics that would render them presumptively invalid.

---

[41] Siegel, *supra* note 40, at 1145–46.

[42] Huq, *supra* note 6, at 1066 (explaining that machine-learning systems "sever the connection between the human operator and the function"). While AI systems may mimic "cognitive" functions in the human brain, intent remains a largely human attribute. *See* Surden, *supra* note 20, at 89, 94.

[43] *See* Mayson, *supra* note 3, at 2224; Ben Grunwald & Jeffrey Fagan, *The End of Intuition-Based High-Crime Areas*, 107 CALIF. L. REV. 345, 398 (2019) (noting that racially discriminatory policing decisions yield definitions of high-crime areas in communities of color, creating a "high-crime feedback loop").

[44] *See* Kleinberg et al., *supra* note 28, at 23, 27–34 (explaining where discrimination is likely and unlikely to originate within algorithms); Roberts, *supra* note 11, at 1697; Noel L. Hillman, *The Use of Artificial Intelligence in Gauging the Risk of Recidivism*, 58 JUDGES' J. 36, 37 (2019) ("[R]ecidivism risk modeling still involves human choices about what characteristics and factors should be assessed, what hierarchy governs their application, and what relative weight should be ascribed to each."). Indeed, in some cases, courts have pierced the algorithmic veil to evaluate creators as the relevant persons of interest. *See, e.g.*, People v. Wakefield, 175 A.D.3d 158, 169–70 (N.Y. App. Div. 2019) (finding the human creator of an AI program the declarant for the purposes of Sixth Amendment confrontation rights).

[45] Jenna Burrell, *How the Machine "Thinks": Understanding Opacity in Machine Learning Algorithms*, BIG DATA & SOC'Y, Jan.–June 2016, at 1. *See generally* FRANK PASQUALE, THE BLACK BOX SOCIETY (2015).

RASs do not explicitly incorporate race as a factor.[46] They instead base predictions on proxies: facially neutral factors — for example, diagnoses, marital status, neighborhood features, childhood experiences, and family criminal background[47] — that tend to be strongly correlated with race.[48] Discriminatory purpose doctrine's probe for explicit classifications thus inadvertently insulates consideration of protected characteristics, the constitutionality of which is a matter of debate.[49]

3. The Countervailing Goal of Accuracy. — In a generalized way, AI systems aim to maximize accuracy — a model whose predictions do not reflect reality is worthless.[50] However, the promotion of accuracy as the be-all and end-all is likely to inflict harm on protected classes: statistically "fair" associations driving prediction may reflect racially distorted rates of inputs among protected groups.[51] The primary challenge is that the most significant variables from the standpoint of predictive accuracy tend to be correlated with protected attributes such as race due to unequally applied criminal justice practices — consider arrest rates or detention practices skewed against African Americans.[52] Moreover, fairness metrics abound, leading to contrary evaluations of the impact

---

[46] *See* Yang & Dobbie, *supra* note 7, at 297.

[47] *See* SLOBOGIN, *supra* note 30, at 6. Basic tools evaluate factors such as criminal history and attitudes toward crime while more advanced tools take into account fluid factors such as familial circumstances and employment history. Bagaric et al., *supra* note 29, at 1059.

[48] Yang & Dobbie, *supra* note 7, at 298–99, 311–19, 364–71 ("[I]f one believes that all racial proxies should be excluded from predictive algorithms, there remains no feasible way of designing an algorithm because every possible input is likely correlated with race." *Id.* at 318–19 (footnote omitted).); Anya E.R. Prince & Daniel Schwarcz, *Proxy Discrimination in the Age of Artificial Intelligence and Big Data*, 105 IOWA L. REV. 1257, 1265–66 (2020); Ajunwa, *supra* note 5, at 1679; Bernard E. Harcourt, *Risk as a Proxy for Race: The Dangers of Risk Assessment*, 27 FED. SENT'G REP. 237, 238 (2015). In some studies, the exclusion of protected variables has been shown to adversely affect accuracy and, in doing so, exacerbate negative outcomes for members of protected groups. *See* Yang & Dobbie, *supra* note 7, at 319–20; Deborah Hellman, *Measuring Algorithmic Fairness*, 106 VA. L. REV. 811, 853–55 (2020); Barocas & Selbst, *supra* note 5, at 721–22; Sam Corbett-Davies & Sharad Goel, The Measure and Mismeasure of Fairness: A Critical Review of Fair Machine Learning 2 (Sept. 11, 2018) (unpublished manuscript), https://arxiv.org/pdf/1808.00023.pdf [https://perma.cc/RE9M-543F]; Moritz Hardt et al., Equality of Opportunity in Supervised Learning 18–19 (Oct. 11, 2016) (unpublished manuscript), https://arxiv.org/pdf/1610.02413.pdf [https://perma.cc/CDX3-W7RQ].

[49] Some scholars contend that RASs' incorporation of race or factors correlated with it is per se unconstitutional under anti-classification norms. *See* Yang & Dobbie, *supra* note 7, at 302–19; Mayson, *supra* note 3, at 2224; Stephanie Bornstein, *Antidiscriminatory Algorithms*, 70 ALA. L. REV. 519, 568–69 (2018); Starr, *supra* note 28, at 812, 819–24; Renata M. O'Donnell, Note, *Challenging Racist Predictive Policing Algorithms Under the Equal Protection Clause*, 94 N.Y.U. L. REV. 544, 566–67 (2019). Others argue against the exclusion. *See* Yang & Dobbie, *supra* note 7, at 351.

[50] *See* Mayson, *supra* note 3, at 2225.

[51] Yang & Dobbie, *supra* note 7, at 294; Barocas & Selbst, *supra* note 5, at 721; Mayson, *supra* note 3, at 2224–25.

[52] *See* Yang & Dobbie, *supra* note 7, at 302; Barocas & Selbst, *supra* note 5, at 721.

of a specific algorithm.[53] For example, in 2016, ProPublica, an investigative news organization, reported that COMPAS produced racially disparate results.[54] Such results could be blamed on the tool's reliance on variables highly correlated with race, including criminal history and neighborhood crime rates.[55] But other scholars show that the choice of fairness metric impacts evaluations of COMPAS's results — under some metrics, there aren't clear racial imbalances.[56] All of this amounts to considerable ambiguity with respect to whether an individual is being discriminated against through algorithms.

4. Court Application. — The failure of the discriminatory purpose requirement is demonstrated by *State v. Loomis*,[57] which centered on a due process challenge to a trial court's use of COMPAS.[58] In upholding the use of COMPAS,[59] the Wisconsin Supreme Court underwent a similar process in trying to discern the intent of the system as would be required under current interpretations of the discriminatory purpose requirement. For the court, the task amounted to an evaluation of whether the system was utilizing a protected classification and how that classification factored into a human judge's analysis.[60] The court rejected the defendant's challenge to COMPAS's consideration of gender in sentencing, finding the goal of "promot[ing] accuracy" sufficed as a neutral explanation.[61] Moreover, the court held there was insufficient evidence that the sentencing court applying COMPAS had considered gender or rendered a decision solely on the basis of group characteristics.[62] The court thus found the retention of a human decisionmaker

---

[53] Huq, *supra* note 6, at 1134 (expounding upon different notions of algorithmic fairness); Corbett-Davies & Goel, *supra* note 48, at 2 (providing three formal definitions of fairness for AI).

[54] Julia Angwin et al., *Machine Bias*, PROPUBLICA (May 23, 2016), https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing [https://perma.cc/68JX-K54S] (finding that the software "was particularly likely to falsely flag black defendants as future criminals").

[55] Starr, *supra* note 28, at 838.

[56] *See* Sam Corbett-Davies et al., *Algorithmic Decision Making and the Cost of Fairness* 797 (KDD 2017 Research Paper), https://dl.acm.org/doi/pdf/10.1145/3097983.3098095 [https://perma.cc/B395-HBSC]; *cf.* Sarah L. Desmarais et al., *Performance of Recidivism Risk Assessment Instruments in US. Correctional Settings*, *in* RISK ASSESSMENT HANDBOOK, *supra* note 16, at 3, 5 (noting that studies have come to conflicting conclusions regarding performance of RASs along the dimensions of race).

[57] 881 N.W.2d 749 (Wis. 2016).

[58] *Id.* at 753.

[59] *Id.* at 757.

[60] *See id.* at 764.

[61] *Id.* at 767. This focus may have been dictated by the specific concerns of procedural due process, of which accuracy is forefront. William G. Young & Jordan M. Singer, *Bench Presence: Toward a More Complete Model of Federal District Court Productivity*, 118 PENN ST. L. REV. 55, 71 (2013). Loomis was not able to access the code itself due to the shield of trade secret protection. *Loomis*, 881 N.W.2d at 761. The court noted that Loomis had an opportunity to verify the accuracy of the information in the report because it was made from publicly available data. *Id.*

[62] *Loomis*, 881 N.W.2d at 767.

alleviated concerns about the lack of an individualized decision.[63] The Wisconsin Supreme Court's analysis did not take into account the nuances of the workings of algorithms described above. As discussed in the next Part, its approach is neither mandated under equal protection analysis nor normatively beneficial for the criminal justice system.

II. EMPHASIZING EFFECT IN THE EVALUATION OF ALGORITHMS

State and federal courts alike have varied in their application of the discriminatory purpose requirement to various categories of cases.[64] Notably, jury selection and political apportionment present two limited contexts in which the Supreme Court has deviated from the general requirement of intent and left room for greater weighing of effects.[65] This Part argues that courts should extend the approach to RASs given the similar interests underlying these cases.

A. Fit with Established Exceptions

An effects-plus framework is already embedded into the fabric of the discriminatory purpose requirement. As explained above, *Davis* and its progeny suggest that *something* more than sole evaluation of the effects of a facially neutral law is necessary to equal discriminatory purpose. In *Davis*, in the context of police officer exams, this something was intent. But current doctrine does not rigidly frame intent as the only element capable of fulfilling the requirement: *Davis* explicitly referenced the contexts of jury selection and voting, where the Court has accepted evidence of disparate impact without inquiry into motive when coupled with an additional element.[66] Of course, intent is the default second element and, as *Davis* and its progeny indicate, a departure must be justified by something other than the difficulty of bringing an equal protection claim. Professor Daniel Ortiz suggests that this justification

---

[63] *Id.* at 769.

[64] Sheila Foster, *Intent and Incoherence*, 72 TUL. L. REV. 1065, 1085 (1998) (noting that "adherence to the *Feeney* conception of intent has been selective").

[65] Prior to *Davis*, the Court accommodated consideration of solely effects in cases of overt discrimination such as *Yick Wo v. Hopkins*, 118 U.S. 356, 362–63 (1886), and *Gomillion v. Lightfoot*, 364 U.S. 339, 341–42 (1960). Although neither case has been overruled, the Court has expressly cabined their approach to exceedingly "stark" cases. *See, e.g.*, McCleskey v. Kemp, 481 U.S. 279, 293 & n.12 (1987) (describing *Gomillion* and *Yick Wo* as "rare cases in which a statistical pattern of discriminatory impact [alone] demonstrated a constitutional violation"); Village of Arlington Heights v. Metro. Hous. Dev. Corp., 429 U.S. 252, 266 (1977) (stating that cases such as *Yick Wo* and *Gomillion* showing "a clear pattern, unexplainable on grounds other than race" may support an inference of discriminatory purpose).

[66] Washington v. Davis, 426 U.S. 229, 239–45 (1976).

stems from the fundamental rights at stake in these domains.[67] This Part argues that the interests and challenges implicated by algorithmic RASs justify a different second element.

1. Jury Selection. — The concern of keeping racial bias out of jury selection has led the Court to minimize the requisite showing for a prima facie antidiscrimination claim. This is borne out through challenges to jury selection approaches and the prosecution's use of peremptory strikes. First, evidence of disparate impact in jury selection approaches has permitted an inference of discriminatory purpose based on a totality of the circumstances and without evaluation of intent.[68] *Davis* acknowledged that "systematic exclusion of eligible jurymen of the proscribed race" may prove discriminatory purpose but noted that this fact "does not in itself make out an invidious discrimination forbidden by the [Equal Protection] Clause."[69] A second element is needed. For example, in finding unconstitutional the use of the "key-man" system of selecting juries in *Castaneda v. Partida*,[70] the Court quoted *Davis* to explain that individuals could make out a prima facie claim with evidence that a racial group has been underrepresented in the jury selection process plus evidence that the process is susceptible to abuse.[71] Once a plaintiff has made out a prima facie case of discrimination, the burden shifts to the government to demonstrate a permissible, race-neutral justification beyond a lack of a discriminatory motive.[72]

Second, under the *Batson v. Kentucky*[73] framework, the Court has struck down states' use of peremptory challenges based on little more than a showing of disparate impact.[74] In *Batson*, the Court held that it was unconstitutional to exercise peremptory challenges to remove

---

[67] Ortiz, *supra* note 10, at 1136–37 (explaining that, in contexts implicating individual liberty interests or fundamental rights, the individual has a smaller burden of persuasion; at the same time, as the state's interest becomes more fundamental, its own burden shrinks in turn).

[68] *Davis*, 426 U.S. at 239–42; *see* Ortiz, *supra* note 10, at 1122–23.

[69] *Davis*, 426 U.S. at 239 (quoting Akins v. Texas, 325 U.S. 398, 403–04 (1945)).

[70] 430 U.S. 482 (1977). The case addressed a key-man system, under which "key" persons in good standing recommended others who would make responsible jurors. *Id.* at 484–85.

[71] *Id.* at 493–94 ("A prima facie case of discriminatory purpose may be proved . . . by the absence of [racial minorities] on a particular jury combined with the failure of the jury commissioners to be informed of eligible [minority] jurors in a community, . . . or with racially non-neutral selection procedures . . . ." (last two omissions in original) (quoting *Davis*, 426 U.S. at 241)). The Court overturned the sentence of a Mexican American defendant who had been convicted after indictment by a grand jury on which Hispanics had been underrepresented. *Id.* at 496, 501.

[72] *See id.* at 497–98; *Davis*, 426 U.S. at 241 (citing Alexander v. Louisiana, 405 U.S. 625, 632 (1972)).

[73] 476 U.S. 79 (1986).

[74] *See* Snyder v. Louisiana, 552 U.S. 472, 485 (2008) (finding that the default rule of showing discriminatory intent does not apply in the context of *Batson* challenges). *But cf.* Foster v. Chatman, 136 S. Ct. 1737, 1754 (2016) (rearticulating the standard for identifying discriminatory intent). The Court has struck down onerous requirements, holding that evidence permitting an inference that discrimination has occurred suffices. Johnson v. California, 545 U.S. 162, 170 (2005).

prospective jurors solely on the basis of race and that a defendant could establish purposeful discrimination "solely on the facts concerning . . . selection in his case."[75] It affirmed *Davis*'s notion that the "invidious quality" of government action claimed to be discriminatory "must ultimately be traced to a racially discriminatory purpose,"[76] but established a three-step burden-shifting framework that heavily weighed effects. First, a defendant must make a prima facie showing that relevant circumstances raise an inference that a peremptory challenge was exercised on the basis of race.[77] This requirement may be satisfied by demonstrating disproportionate impact.[78] Second, upon that showing, the burden shifts to the state to offer a race-neutral explanation beyond denial of a discriminatory motive.[79] Third, the court determines whether the defendant has shown "purposeful discrimination."[80] In its most recent case addressing *Batson* challenges, the Court reaffirmed its holding that clear statistical evidence of egregious disparate effect, coupled with little more, stands in for the intent requirement.[81]

Algorithmic RASs similarly justify departure from the element of intent. Both applications are paramount in the trial right — the fact-finding of juries complements the work of a neutral judge in determining a sentence. Yet, with both juries and algorithmic systems, concerns of bias, both conscious and unconscious, can infringe upon the individual liberty interest in a fair trial. Both represent contexts that may pose challenges in the discernment of pretext — our criminal justice system has retained peremptory strikes to protect government discretion, but it is often difficult to prove that the proffered reasons are inaccurate, implausible, or false.[82] These concerns, which justify a standard that is not solely based on notions of intent for jury selection, also play out in the case of algorithmic systems. The difficulty of proving the intent of prosecutors ratchets up to impossibility in the case of AI.

Moreover, the potential use of proxies in both jury selection and RASs suggests that assessment of impact is similarly vital in both contexts. Given prohibitions against the explicit consideration of race in jury selection, decisionmakers may rely on strongly correlated attributes

---

[75] *Batson*, 476 U.S. at 95 (emphasis omitted); *accord id.* at 96–98. The Court analogized to the context of Title VII, which is governed by a "disparate treatment" standard. *Id.* at 94 n.18.

[76] *Id.* at 93 (quoting *Davis*, 426 U.S. at 240).

[77] *See id.* at 96.

[78] *See id.* at 96–97.

[79] *Id.* at 97–98.

[80] *Id.* at 98.

[81] Flowers v. Mississippi, 139 S. Ct. 2228, 2235 (2019) (finding that clear statistical evidence of disparate racial impact — in this case, evidence that the state struck forty-one of forty-two Black prospective jurors — sufficed to establish the state was "motivated in substantial part by discriminatory intent" (quoting Foster v. Chatman, 136 S. Ct. 1737, 1754 (2016))).

[82] *See* Joshua C. Polster, *From Proving Pretext to Proving Discrimination: The Real Lesson of* Miller-El *and* Snyder, 81 MISS. L.J. 491, 521–34 (2012).

to arrive at the same outcomes.[83] This is all the more likely given the lack of meaningful constraints to limit the discretion of decisionmakers in their selection of jurors.[84] Similar concerns are present in the context of algorithmic RASs: with respect to sentencing, decisionmakers are already afforded wide latitude in determining factors to consider.[85] This discretion is amplified by the lack of meaningful oversight of RAS creators and the lack of rules governing some prediction processes.[86] The discretion of algorithmic designers is even more troubling than that of judges: whereas judges are subject to minimal measures of accountability,[87] the criminal justice system has not imposed a similar level of accountability on private companies designing RASs.[88] Thus, the interests of justice justify the omission of any requirement to prove discriminatory intent.

In the context of both jury selection and algorithmic RASs, the scope of the parties impacted extends beyond the persons whose rights are immediately at stake. Discriminatory jury selection impacts not only a defendant's equal protection rights but also those of the excluded juror.[89] Professor Brooks Holland notes that courts may show "greater ambivalence" when only a "guilty" defendant's rights are at issue — even though an equal protection claim does "not necessarily bear on the defendant's factual guilt or innocence" — than when the rights of "innocent" victims such as jurors are implicated.[90] However, as Holland notes, the harm of racial bias in jury selection reaches beyond excluded jurors to "society" and the "rule of law."[91] In the same way, racial bias in algorithmic RASs has great potential — in conjunction with uncertain

---

[83] *See* Ortiz, *supra* note 10, at 1125 (describing how the procedure used by one jury commissioner utilized attributes such as "neighborhood, prestige of profession, and homeowner status as proxies for civic responsibilities, . . . screen[ing] out whole classes of people — not only blacks, but also the poor").

[84] Russell D. Covey, *The Unbearable Lightness of* Batson*: Mixed Motives and Discrimination in Jury Selection*, 66 MD. L. REV. 279, 344 (2007) (noting the discretion of prosecutors to "remove virtually anyone from the jury").

[85] Federal sentencing courts face few constraints on the categories or sources of information considered. *See* Pepper v. United States, 562 U.S. 476, 488 (2011).

[86] *See* Andrea Nishi, Note, *Privatizing Sentencing: A Delegation Framework for Recidivism Risk Assessment*, 119 COLUM. L. REV. 1671, 1674 n.12, 1682–90 (2019) (noting the lack of meaningful oversight of RAS creators, given the practical inscrutability of source code to judges and defendants).

[87] Judges are appointed or elected and subject to sanction for improper acts. *See* Albert J. Krieger, *A Wave and a Wish*, CRIM. JUST., Summer 2003, at 1, 29; *cf.* John L. Warren III, *Holding the Bench Accountable: Judges qua Representatives*, 6 WASH. U. JURIS. REV. 299 (2014) (discussing strategies to increase judicial accountability).

[88] *See* Katyal, *supra* note 3, at 100 (describing how algorithms designed by private companies have evaded traditional methods of legal oversight and accountability).

[89] Brooks Holland, *Race and Ambivalent Criminal Procedure Remedies*, 47 GONZ. L. REV. 341, 358 (2011/2012).

[90] *Id.* at 358; *see also id.* at 359 (arguing that such rationales are "misguided").

[91] *Id.* at 358.

predictions of what a defendant may or may not do in the future — to further weaken confidence in sentencing.[92]

As demonstrated by the ease with which the government has been able to meet its burden of showing a race-neutral reason in jury selection cases,[93] allowing greater incorporation of effect may not in practice diminish the government's ability to evade searching review of equal protection challenges. However, reliance on impact is still beneficial because it allows plaintiffs to reach *Batson*'s step two — the state's burden. This allocation of the burden of persuasion more appropriately accounts for informational disparities and the greater role of the state in authorizing the use of algorithms. Individuals who have identified a disparate impact are less likely to be able to present concrete evidence regarding algorithmic operation than the state actors responsible for the use of such algorithmic systems. Under this framework, the onus will more appropriately fall on the state, which will be forced to reckon with the propriety of its prediction problem and the application to a defendant. Even if such a task does not necessarily force the state to cease use of a system creating a disparate impact, it will incentivize the state to understand the workings of its systems, avoid overly complicated approaches, and opt for explainable, transparent operations.

2. Political Apportionment. — In the context of vote dilution, the Court has accepted evidence of disparate impact as evincing discriminatory purpose when coupled with evidence of past discrimination, even when there is no clearly identifiable decisionmaker involved.[94] In rejecting a racial vote dilution challenge to at-large municipal elections in *City of Mobile v. Bolden*,[95] a plurality of the Justices invoked *Feeney* to emphasize that an Equal Protection Clause violation requires a showing of "purposeful" discrimination that cannot be met with evidence of disparate impact alone.[96] However, the Court swiftly changed course, as Congress amended the Voting Rights Act at least in part in response to criticism of *Bolden*.[97] Two years later, in a challenge to another at-large voting system that allegedly discriminated against Black residents in *Rogers v. Lodge*,[98]

---

[92] *See* Katyal, *supra* note 3, at 83–86.

[93] Alexis Straus, Note, *(Not) Mourning the Demise of the Peremptory Challenge: Twenty Years of* Batson v. Kentucky, 17 TEMP. POL. & C.R.L. REV. 309, 333–34 (2007).

[94] Ortiz, *supra* note 10, at 1127. However, Congress subsequently amended the Voting Rights Act to incorporate a results test. *See* Voting Rights Act Amendments of 1982, Pub. L. No. 97-205, § 2, 96 Stat. 131, 131 (codified as amended at 52 U.S.C. § 10303(a)(1)).

[95] 446 U.S. 55 (1980).

[96] *Id.* at 66–67, 71 n.17 (plurality opinion).

[97] *See* Michael Parsons, *Clearing the Political Thicket: Why Political Gerrymandering for Partisan Advantage Is Unconstitutional*, 24 WM. & MARY BILL RTS. J. 1107, 1116 (2016).

[98] 458 U.S. 613 (1982).

the Court reaffirmed intent as a primary focus.[99] However, it made clear that equal protection doctrine analyzes how the electoral system functions. The Court found that the intent element was satisfied by "an aggregate of . . . factors"[100] that more closely evidenced impact, including the fact that no Black people had ever been elected and the history of past discrimination against Blacks in the political process.[101] As the dissents noted, the Court did not identify the relevant decisionmakers or their intent[102] — a result seemingly inconsistent with *Davis*. The Court's opinion in fact quoted *Davis* in asserting that "an invidious discriminatory purpose may often be inferred from the totality of the relevant facts, including the fact, if it is true, that the law bears more heavily on one race than another."[103]

The Court has also been willing to infer a legislature's overreliance on a scrutinized classification, like race, based on resulting demographic impact or bizarre district shape. Professor Heather Gerken notes that the Supreme Court has referred to these cases as involving "facially neutral classifications" and yet declined to require proof of intent as necessitated by *Davis* and its progeny.[104] For example, in *Shaw v. Reno*,[105] the Court held in favor of White voters in their challenge to majority-minority districts without requiring proof of discriminatory purpose; the Court instead inferred impermissible reliance on race from the bizarre shape of the challenged district.[106] Similarly, while more modern redistricting cases have followed the rationale of *Davis* in requiring a showing that race was the predominant factor,[107] courts have embraced plaintiffs' use of statistical evidence of impact where states have used district-drawing software.[108] For example, in *Bush v. Vera*,[109] the Court struck down Texas's redistricting scheme, which was the product of the use of sophisticated software that drew lines in ways that suggested race was a proxy for political affiliation.[110]

---

[99] *Id.* at 617 (reaffirming that intent "has long been required in *all* types of equal protection cases charging racial discrimination" (first citing Village of Arlington Heights v. Metro. Hous. Dev. Corp., 429 U.S. 252, 265 (1977); and then citing Washington v. Davis, 426 U.S. 229, 240 (1976))).

[100] *Id.* at 620 (quoting Zimmer v. McKeithen, 485 F.2d 1297, 1305 (5th Cir. 1973)).

[101] *Id.* at 623–27.

[102] *See id.* at 647 (Stevens, J., dissenting); *id.* at 628–29 (Powell, J., dissenting).

[103] *Id.* at 618 (majority opinion) (quoting *Davis*, 426 U.S. at 242).

[104] Heather K. Gerken, *Understanding the Right to an Undiluted Vote*, 114 HARV. L. REV. 1663, 1695–96 (2001).

[105] 509 U.S. 630 (1993).

[106] *Id.* at 643–44. The Court's reasoning evoked the prior effects-friendly approach of *Yick Wo* and *Gomillion*. *Id.* at 644; *see supra* note 65.

[107] *See, e.g.*, Miller v. Johnson, 515 U.S. 900, 916 (1995).

[108] Huq, *supra* note 2, at 1281 (noting that statistical evidence in gerrymandering cases "helps tease out the correlations between districting and race, partisanship, and other relevant factors with precision" (citing Cooper v. Harris, 137 S. Ct. 1455, 1477–78 (2017))).

[109] 517 U.S. 952 (1996).

[110] *Id.* at 970–72 (plurality opinion). A plurality of Justices found the disparate racial impact troubling, with the inference of disparate impact supported by the fact that the districts in question

The use of algorithmic RASs resembles the voting context in the challenge of discerning improper motives. Implicit incorporation of race is expected in both: section 5 of the Voting Rights Act requires consideration of race, albeit in a limited manner;[111] similarly, algorithmic RASs implicitly incorporate factors highly correlated with race. In evaluating whether consideration of a factor by an algorithmic RAS is improper, courts face a doctrinal challenge similar to discerning the collective intent of a legislature. Both contexts lack one relevant human decisionmaker for evaluation of intent. Both also feature severe informational disparities between individuals challenging government action and the state itself — the proprietary nature of many algorithms creates a similar roadblock as do political factors insulating motives for the shaping of districts. The subtlety of these relationships counsels against entangling courts in the determination of the significance of a specific variable.

The difficulties of discerning intent justify evaluation of results such as the shape of a district or the ensuant marginalization of a racial group in sentencing. The Court's willingness to primarily consider effects should thus extend to critical consideration of RASs, which are replete with facially neutral practices that may work to deprive defendants of a fundamental liberty interest. As redistricting cases illustrate, the potential to impact a range of decisions, such as the need to correct representation flowing from unconstitutionally drawn district lines, is not fatal: an injunction against a redistricting scheme may carry at least as much impact as one against use of an RAS.[112]

B. Distinguishing Algorithms from Traditional Sentencing Challenges

Standing most directly in the way of plaintiffs seeking to make out a prima facie claim for racial discrimination in sentencing is the precedent of *McCleskey v. Kemp*.[113] There, the Court dismissed a defendant's challenge to Georgia's death penalty statute on the basis of a comprehensive statistical study conducted by Professors David C. Baldus, Charles Pulaski, and George Woodworth (the Baldus study),[114] which

---

were "bizarrely shaped and far from compact." *Id.* at 979. Similar disparate impact evidence justified the Court's recent invalidation of two North Carolina congressional districts as unconstitutional racial gerrymanders. *Cooper*, 137 S. Ct. at 1472, 1474. The Court upheld a district court finding that, even if there was no racial intent in the drawing of maps, the predominant use of race as a proxy for partisanship constitutes racial gerrymandering. *Id.*

[111] Jon Greenbaum et al., Shelby County v. Holder*: When the Rational Becomes Irrational*, 57 HOW. L.J. 811, 860 (2014).

[112] *See, e.g.*, Harris v. McCrory, 159 F. Supp. 3d 600, 627 (M.D.N.C. 2016), *aff'd sub nom. Cooper*, 137 S. Ct. 1455 (requiring that the North Carolina General Assembly redraw congressional district).

[113] 481 U.S. 279 (1987).

[114] *Id.* at 297.

showed that the imposition of capital punishment was strongly correlated with the race of a defendant and the race of a victim.[115] In so doing, the Court rejected the capacity of statistics to provide circumstantial evidence of a discriminatory individual sentencing decision.[116] Plaintiffs have accordingly struggled to overcome *McCleskey*'s bar in all but the starkest cases of discrimination.[117]

McCleskey was wrongly decided and should be overruled.[118] Commentators have opined that the Court's reasoning "misconstrued . . . the effectiveness of statistical analyses."[119] *McCleskey* stands out in that it involved a significant individual liberty interest — life itself, in the context of capital punishment, in which the Court has applied greater procedural protections and heightened liability relative to other sentencing cases[120] — and yet saddled the individual with the high burden of showing actual motivation.[121] This section demonstrates that the *McCleskey* decision does not necessarily bar development of a framework emphasizing impact for algorithmic RASs.[122] Although *McCleskey* addresses

---

[115] *Id.* at 286. The African American defendant had been convicted of killing a White police officer; the Baldus study showed that defendants charged with killing White victims were more likely to receive a death sentence than those charged with killing Black victims. *Id.* at 287.

[116] In this aspect, the *McCleskey* Court would seem to reject the application of RASs in sentencing, where they predict the likelihood of individual behavior from group factors.

[117] *See, e.g.*, United States v. Thurmond, 7 F.3d 947, 952–53 (10th Cir. 1993) (finding that statistical evidence of racial disparities in the prosecution of drug offenses was insufficient to establish that the challenged penal provision and Sentencing Guidelines were unconstitutional under the Equal Protection Clause). *See generally* David C. Baldus et al., *Race and Proportionality Since* McCleskey v. Kemp *(1987): Different Actors with Mixed Strategies of Denial and Avoidance*, 39 COLUM. HUM. RTS. L. REV. 143, 150–51 (2007); John H. Blume et al., *Post-*McCleskey *Racial Discrimination Claims in Capital Cases*, 83 CORNELL L. REV. 1771, 1780–98 (1998).

[118] *See generally* DAVID C. BALDUS ET AL., EQUAL JUSTICE AND THE DEATH PENALTY 370–87 (1990); Randall L. Kennedy, McCleskey v. Kemp*: Race, Capital Punishment, and the Supreme Court*, 101 HARV. L. REV. 1388, 1388 (1988); Vada Berger et al., Comment, *Too Much Justice: A Legislative Response to* McCleskey v. Kemp, 24 HARV. C.R.-C.L. L. REV. 437, 438 (1989). Justice Powell, who authored *McCleskey*, stated that he would change his deciding vote if he could. David Von Drehle, *Retired Justice Changes Stand on Death Penalty*, WASH. POST (June 10, 1994), https://www.washingtonpost.com/archive/politics/1994/06/10/retired-justice-changes-stand-on-death-penalty/9ccde42b-9de5-46bc-a32a-613ae29d55f3 [https://perma.cc/6NDB-62UL].

[119] *See, e.g.*, Foster, *supra* note 64, at 1146.

[120] *See, e.g.*, Kennedy v. Louisiana, 554 U.S. 407, 413 (2008) (holding capital punishment categorically unavailable for child rape cases in which the victim lives); Roper v. Simmons, 543 U.S. 551, 570–71 (2005) (invalidating death penalty for juvenile offenders); Ring v. Arizona, 536 U.S. 584, 609 (2002) (declaring it unconstitutional for "a sentencing judge, sitting without a jury, to find an aggravating circumstance necessary for imposition of the death penalty").

[121] Ortiz explains this discrepancy by noting that the state's interest in the administration of a "case touching the heart of the criminal process" provided for a correspondingly easier burden for the state overall relative to the individual. Ortiz, *supra* note 10, at 1148.

[122] *Cf.* Marc Price Wolf, Note, *Proving Race Discrimination in Criminal Cases Using Statistical Evidence*, 4 HASTINGS RACE & POVERTY L.J. 395, 405 (2007) ("In order for the Supreme Court to validate a racial discrimination equal protection claim based on a sophisticated statistical study, the Court does not have to overrule *McCleskey*.").

a plaintiff's burden in the sentencing context, the use of RASs is sufficiently distinct to justify a different approach.

The *McCleskey* Court undoubtedly was reluctant to accept disparate impact as the sole basis of an equal protection challenge. First, while the Court accepted the validity of the Baldus study,[123] it applied *Feeney*'s construction of discriminatory purpose in requiring a showing of intent by the legislature to enact or maintain the death penalty statute specifically "*because of* an anticipated racially discriminatory effect."[124] Statistical evidence yielding a generalized inference of class-based harm would not suffice.[125] The Court distinguished capital sentencing decisions from Title VII and jury selection in that, in the latter contexts, "the statistics relate to fewer entities, and fewer variables are relevant to the challenged decisions."[126] It emphasized that the discretion of prosecutors, juries, judges, and others involved in crafting a sentence is "essential" so as to demand "exceptionally clear proof" of abuse before escalation to strict scrutiny.[127] Second, the opinion stressed the need to provide the state an effective chance to rebut an inference of discriminatory intent: in jury selection cases, the decisionmaker may explain the disparity, whereas the state lacks the same opportunity to defend the prosecutor's and jury's decisions to seek and impose the death penalty, since their decisions may have been made years prior and the jury cannot generally be called to testify.[128] Third, the Court credited the presence of the "legitimate and unchallenged explanation" that Georgia law permitted capital punishment.[129]

Claims based on algorithmic RASs differ from McCleskey's claim in a few important ways. For example, McCleskey's claim "thr[ew] into serious question" the legitimacy of a broad range of sentences and convictions — "the principles that underlie our entire criminal justice system."[130] While a claim based solely on the Baldus study would impugn

---

[123] McCleskey v. Kemp, 481 U.S. 279, 291 n.7 (1987).

[124] *Id.* at 298.

[125] *Id.* at 294.

[126] *Id.* at 295 (footnote omitted). The Court explained that, "[i]n venire-selection cases, the factors that may be considered are limited, usually by state statute. . . . In contrast, a capital sentencing jury may consider *any* factor relevant to the defendant's background, character, and the offense." *Id.* at 295 n.14. Scholars argue that this statement is "misleading" given the lack of meaningful limits on considerations for selecting venire. *See* Ortiz, *supra* note 10, at 1144.

[127] *See McCleskey*, 481 U.S. at 297.

[128] *Id.* at 296.

[129] *Id.* at 297.

[130] *Id.* at 315. Motivating the Court were the concerns of unbridled liability that animated *Davis*. It cautioned that ruling for McCleskey might invite challenges to many "unexplained discrepancies that correlate to membership in other minority groups." *Id.* at 316. Given that his claim centered on the race of different parties, future claims could be based on the race of judges and attorneys or hinge on attributes such as physical attractiveness — "there is no limiting principle." *Id.* at 318; *see id.* at 317. The Court found that claims of racial stratification in criminal justice were best left to

a host of cases made by different decisionmakers — "every actor in the Georgia capital sentencing process"[131] — challenges to RASs implicate one uniform decisionmaker and thus involve only cases in which a particular algorithm was used.[132] RASs also offer the opportunity for explanations of factors. Though the proprietary nature of some RASs undermines the depth of their explanations, such opportunity is at least as present as in jury selection cases.[133] In sum, while *McCleskey* jeopardizes the fate of equal protection challenges in the sentencing context generally, the application of algorithmic RASs differs in important ways that justify departures from *McCleskey*'s approach.

III. APPLICATION OF AN EFFECTS-PLUS FRAMEWORK

It remains to be seen what an effects-plus framework may look like in a challenge to the use of algorithmic RASs. This Part contends, in line with other scholarly proposals,[134] that an effects-plus framework resolves the tension inherent in equal protection doctrine regarding algorithms. The framework would collapse current doctrine's two paths for a plaintiff to establish a prima facie claim. Under the first, a plaintiff may show a system's explicit classification based on race, which is presumptively invalid.[135] However, given the reliance of AI on proxies, such a showing should not be *conclusive* of invalidity, as it would dismiss or obscure the impact of inevitable proxies that feature in algorithmic prediction.

Thus, regardless of the presence of an explicit racial classification, a plaintiff must first demonstrate disparate impact, as evidenced through

---

legislatures. *Id.* at 319. Given the judiciary's history of addressing such challenges in contexts such as the jury right, this argument fails to meaningfully justify foreclosing judicial remedy.

[131] *Id.* at 292; *see* Blume et al., *supra* note 117, at 1778.

[132] *Cf.* Wolf, *supra* note 122, at 407 (arguing that a study that focuses on "repeat actors" will be more likely to overcome *McCleskey*); Andrew D. Leipold, *Objective Tests and Subjective Bias: Some Problems of Discriminatory Intent in the Criminal Law*, 73 CHI.-KENT L. REV. 559, 596 (1998) (noting that challenges to police departments or prosecutors' offices would be a "smaller leap" as those groups are "repeat players in the justice system[]"). In this way, challenges to the use of algorithmic RASs resemble challenges to the decisions of a specific judge who routinely considered an impermissible factor.

[133] This concern must be compared with judges' explanations, which also may be incomplete given unconscious biases and justifications based on life experience. *See supra* p. 1764.

[134] Aziz Z. Huq, *Constitutional Rights in the Machine Learning State*, 105 CORNELL L. REV. 1875, 1917 (2020) (proposing that equal protection analysis focus on the impact of such systems on "pernicious social stratification"); Hellman, *supra* note 48, at 820–34 (arguing that fairness ought to take into account algorithms' differing impacts on belief and action); Mayson, *supra* note 3, at 2282, 2287 (proposing that the nature of risk itself and the criminal justice system's response to it be reevaluated); Kleinberg et al., *supra* note 28, at 2 (suggesting that discrimination may best be avoided by "regulating the process through which algorithms are designed").

[135] *See* Pers. Adm'r of Mass. v. Feeney, 442 U.S. 256, 272 (1979). For a collection of scholarly arguments regarding the constitutionality of an algorithm's classification on a protected attribute, see sources cited *supra* note 49.

statistical evidence of racial disparity.[136] As the necessary second element, the plaintiff must demonstrate that the outcome predicted by an RAS is susceptible to racially biased analysis. This second element aligns with showing susceptibility to abuse in the jury selection cases and a history of discrimination in the voting cases because it addresses the likelihood of bias or manipulation in an action implicating fundamental liberty interests. This element acknowledges that predictions of an outcome that happens more readily among a certain class of people, based on past data that may reflect racial inequities, will increase racial disparities.[137] A plaintiff could challenge specific factors used in the RAS or critique the manner of use by a sentencing judge.

That brings us to the state's burden to provide a neutral (nondiscriminatory) justification.[138] This burden exists in both the *Davis* framework and cases arising in the jury selection and voting contexts. Relating to peremptory strikes, the Court has acknowledged that allowing any relevant reason to constitute a permissible, race-neutral justification would counteract the goal of *Batson*.[139] In practice, myriad reasons have been found to constitute non-pretextual rationales for dismissing a juror.[140] The promotion of accuracy, which drives statistical systems, ostensibly presents such a justification. Indeed, where the predicted outcome is the actual outcome of interest, the promotion of accuracy may suffice as a reasonable justification. However, where there is a mismatch, the burden of persuasion should require the state to provide a reason that isn't solely the general promotion of accuracy; accuracy with respect to an imperfect proxy that may be racially biased is not a useful aim for the criminal justice system. Requiring the government to articulate a reason for discriminatory effect in these cases may incentivize state actors and algorithmic developers to ensure they can understand why an algorithm

---

136 Such a showing could be made with evidence of disparity with respect to sentences for similar crimes and backgrounds, in line with considerations sentencing courts already must take into account. *See* 18 U.S.C. § 3553(a) (listing factors sentencing courts must consider).

137 *See* Mayson, *supra* note 3, at 2222; SARAH PICARD ET AL., CTR. FOR CT. INNOVATION, BEYOND THE ALGORITHM: PRETRIAL REFORM, RISK ASSESSMENT, AND RACIAL FAIRNESS 10 (2019), https://www.courtinnovation.org/sites/default/files/media/document/2019/Beyond_The_Algorithm.pdf [https://perma.cc/SE2P-RF7H] ("One by-product of risk algorithms is that the members of whichever groups have more frequent contact with the justice system will . . . be more frequently classified — and also *misclassified* — as high-risk.").

138 *See* Andrew D. Selbst, *Disparate Impact in Big Data Policing*, 52 GA. L. REV. 109, 193 (2017) (noting the option of a burden-shifting framework).

139 Miller-El v. Dretke, 545 U.S. 231, 240 (2005).

140 John P. Bringewatt, Note, Snyder v. Louisiana*: Continuing the Historical Trend Towards Increased Scrutiny of Peremptory Challenges*, 108 MICH. L. REV. 1283, 1285–86 (2010) (noting that peremptory strikes are still "routinely made on the basis of race" post-*Batson*, *id.* at 1286); Straus, *supra* note 93, at 333–34 (listing qualifying reasons including age, extensive criminal history, language barriers, and demeanor). *Snyder v. Louisiana*, 552 U.S. 472 (2008), partly resolved this issue by applying more probing analysis of potentially pretextual justifications. Bringewatt, *supra*, at 1302.

makes decisions ahead of time and limit the implementation of opaque sentencing algorithms where they cannot do so.

The approach can be illustrated by a race-based equal protection challenge to the system at issue in *Loomis*. The plaintiff would have to first show a disparate effect involving the use of an RAS — for example, that COMPAS yielded disparate racial impact in sentencing. The plaintiff would next have to show that the outcome predicted (future arrest) is susceptible to abuse through racial bias in the input data.[141] Given the racial distortions in arrest data, the state would have to justify its inclusion of the factor with a reason that does not amount solely to a desire to make accurate predictions of arrest, such as the relevance of an individual's age or maturity to their likelihood of recidivism or the value of particular neighborhood characteristics in demonstrating specific support systems. The mere retention of a human judge overseeing the process would not suffice to meet the state's burden if the decisionmaker in any way consulted the algorithmic outputs.

Evidence of discriminatory motive can also factor into this framework — for example, where a plaintiff seeks to show that creators have designed a program so as to make discrimination based on protected attributes probable. This motive may be evidenced through the inclusion of facially neutral factors known to cause disparate racial impact. For example, Judge Calabresi has argued that the legislature's enactment of a law while aware of the racial impact of a significant sentencing disparity would violate the Equal Protection Clause.[142] In a similar vein, where neighborhood characteristics have been shown to cause a discriminatory impact, their inclusion — even when framed in facially neutral ways —may evince a harmful bias in the system and a corresponding intent by human designers to effect racial discrimination. But this showing, which is difficult due to the proprietary nature of RASs, is not required.

The approach outlined above seeks to permit the use of algorithmic RASs in criminal justice only where their benefits in terms of uniformity and traceability outweigh potential harms. As an illustration, the use of risk assessment as a diagnostic tool to evaluate the effectiveness of interventions that weigh on the ability of individuals to reoffend or attend future court events is appropriate and adequately safeguards individual

---

[141] The plaintiff's claim would be strengthened by noting that the outcome predicted of future arrest differs from the outcome of interest of likelihood of recidivism. This divergence, which could be shown through studies of false arrests, is likely given the difficulty of observing and modeling recidivism directly.

[142] United States v. Then, 56 F.3d 464, 468 (2d Cir. 1995) (Calabresi, J., concurring).

autonomy.[143] But, in sentencing, which is inherently centered on punishment for past behavior,[144] prediction processes are based on uncontrollable factors and the costs of error are substantial. It is thus important to take into account the disconnect between outcomes of interest, which will often relate to the purposes of punishment, and the predicted proxies, which are often tainted by historic inequality. This approach is intended to work in tandem with proposed efforts to correct informational imbalances in sentencing, such as disclosure regimes that allow individuals to access the workings of programs utilized to make decisions.[145]

CONCLUSION

It is a truism widely acknowledged that the best predictor of future behavior is past behavior. However, the use of algorithmic RASs pushes us to weigh the costs of this principle in the face of biased predictions and illusory remedial schemes. It is increasingly challenging to raise an equal protection claim successfully given the discriminatory purpose requirement. Even in domains such as jury selection and voting, where an effects-plus framework does not require evidence of intent, litigants have faced an uphill battle as state actors are afforded deference in their explanations of facially neutral actions that produce discriminatory impact. A similar framework for algorithmic RASs would likely run into these hurdles.

Consequently, a doctrinal shift for the discriminatory purpose requirement — one that would incorporate a broader conception of intent or embrace primary evaluation of impact — is normatively beneficial.[146] But doctrinal shifts more often than not occur in increments. This Note showcases an incremental step in the development of a more robust equal protection framework that better responds to the workings of technology. Alongside transparency and disclosure proposals, it also aims to provide plaintiffs a wider opportunity to assert challenges in the face of discriminatory practices in the criminal justice system and thus realign the doctrine with its goal to protect against the harms of discrimination, whether conscious or not.

---

[143] *See* Chelsea Barabas et al., *Interventions over Predictions: Reframing the Ethical Debate for Actuarial Risk Assessment*, 81 PROCS. MACH. LEARNING RSCH. 62, 72–73 (2018). This approach aligns with Professor Sandra Mayson's proposal to provide a supportive response to risk. Mayson, *supra* note 3, at 2287.

[144] Sentencing relates to the goals of deterrence, rehabilitation, incapacitation, and retribution. The use of risk scores in sentencing must overall comport with a punishment strategy that takes into account some of these aims. *See* Ewing v. California, 538 U.S. 11, 25 (2003) (plurality opinion).

[145] *See* Selbst, *supra* note 138, at 190 ("[T]here is no reason, as a matter of policy, why trade secrets should have preferential status over something as important as fairness in criminal justice.").

[146] *See* sources cited *supra* note 40.