Constitutional Dimensions of Predictive Algorithms in Criminal Justice*

Michael Brenner[†]
Jeannie Suk Gersen[‡]
Michael Haley[**]
Matthew Lin[††]
Amil Merchant[‡‡]
Richard Jagdishwar Millett[***]
Suproteem K. Sarkar[†††]
Drew Wegner[‡‡‡]

This Article analyzes constitutional concerns presented by the use of risk-assessment technology in the criminal justice system, and how courts can best address them. Focusing on due process and equal protection, this Article explores avenues for constitutional challenges to risk-assessment technology at federal and state levels and outlines how instruments might be retooled to increase accuracy and accountability while satisfying constitutional standards.

Table of Contents

Introduction 268
I. Risk-Assessment Technology in Criminal Justice 269
A. How Risk-Assessment Technologies Are Used 269
B. The Risk-Assessment Tools 272

* The Authors wish to express their deep gratitude to Kyle Tramonte, April Xu, Bo Kim, and Jess Hui for making it possible to complete this Article, and to Andrew Hyun for research assistance.

†Michael Brenner is the Glover Professor of Applied Mathematics and Applied Physics, and Harvard College Professor, at Harvard University.

‡Jeannie Suk Gersen is the John H. Watson, Jr. Professor of Law at Harvard Law School.

** Michael Haley is a 2019 graduate of Harvard Law School and current term law clerk on the United States District Court for the District of New Hampshire.

††Matthew Lin is a 2019 graduate of Harvard Law School and an associate at Sheppard, Mullin, Richter & Hampton LLP.

‡‡Amil Merchant is a 2019 graduate of Harvard College.

*** Richard Jagdishwar Millett is a 2020 graduate of Harvard Law School and a current term law clerk on the United States District Court for the Northern District of California.

†††Suproteem K. Sarkar is a 2019 graduate of Harvard College and a current Ph.D. Candidate in Economics at Harvard University. Sarkar gratefully acknowledges the support of a National Science Foundation Graduate Research Fellowship.

‡‡‡Drew Wegner is a 2019 graduate of Harvard Law School and a current term law clerk on the United States Court of Appeals for the Ninth Circuit.

C. Existing Literature and Debate 275
II. Due Process 278
A. Opportunity to Challenge Algorithmic Risk Calculations 279
B. Due Process Implications of COMPAS 282
C. Due Process and Explainable Algorithms 283
III. Race Discrimination 287
A. Algorithms and Input Features 290
IV. Equal Protection 291
A. The Davis-McCleskey *Framework* 292
B. Malicious Designer 294
C. Intent By Continued Use 296
D. A Batson*-like Model of Equal Protection Analysis* 297
E. Risk-Assessment Instruments & Burden-Shifting Analysis 299
V. State Constitutional Claims 302
A. State Constitutions 302
B. Where to Bring Challenges to Risk-Assessment Technology 306
VI. Moving Forward 309

Introduction

Artificial intelligence and algorithmic tools are rapidly becoming embedded in our criminal justice system. These risk-assessment tools are used in key stages of the criminal process, from bail determinations to sentencing decisions. More than twenty states use some type of algorithmic modeling to calculate criminal defendants' recidivism risk.[1] Some states use algorithmic models that are developed by private companies and are thus proprietary. These proprietary tools' lack of transparency makes it impossible to determine precisely how their predictive assessments are generated.

This Article analyzes the constitutional issues presented by the use of risk-assessment technologies in our criminal justice system. Part I surveys their current use in criminal justice. Part II analyzes how the use of certain proprietary algorithmic models may violate due process. Part III details the discriminatory nature of risk-assessment instruments and argues that prevailing equal protection jurisprudence should be reexamined to address these harms. Part IV presents potential avenues for bringing constitutional challenges to the use of risk-assessment technologies at the state level. Finally, Part V outlines how these instruments might be retooled to increase accuracy and accountability while satisfying constitutional standards.

[1] *See infra* Part I.A.

I. Risk-Assessment Technology in Criminal Justice

A. *How Risk-Assessment Technologies Are Used*

A common context in which risk-assessment technologies are used in the criminal justice system is in setting the terms of parole and probation. These algorithmic models typically include a *risk component*, which measures the likelihood of reoffending, and a *needs component*, which measures the need for services addressing, for example, mental health and drug abuse.[2]

Some states have created their own tools for parole decisions.[3] Other states use transparent arithmetic scoring systems.[4] Yet other states use commercially available tools.[5] One such tool is the Correctional Offender Management Profiling for Alternative Sanctions ("COMPAS"), used, for

---

[2] For example, the Connecticut Salient Factor Score rates defendants on their likelihood of reoffending (risk) and their risk factors for mental health, drug abuse, and sex offenses necessitating additional services or monitoring (needs). Shamir Ratansi & Stephen M. Cox, Assessment and Validation of Connecticut's Salient Factor Score 3 (2007), http://portal.ct.gov/-/media/DOC/Pdf/RevalidationStudy2007pdf.pdf, *archived at* https://perma.cc/RM7E-Z96C.

[3] These states include: California, which uses the California Static Risk Assessment Instrument ("CSRA"), *see* Susan Turner, James Hess & Jesse Jannetta, *Development of the California Static Risk Assessment Instrument (CSRA)* 4 (Univ. of Cal., Irvine Ctr. for Evidence-Based Corr., Working Paper, 2009), http://ucicorrections.seweb.uci.edu/files/2009/11/CSRA-Working-Paper.pdf, *archived at* https://perma.cc/T58N-M4HS; Connecticut, which uses the Salient Factor Score ("SFS"), *see* Ratansi & Cox, *supra* note 2, at 3; Montana, which uses the Montana Offender and Reentry Risk Assessment Tool ("MORRA"), *see* Kevin Olson, Risk and Needs Assessments for Adult Case Management 1 (2018); and Texas, which uses the Texas Risk Assessment System ("TRAS"), *see The Texas Risk Assessment System: A New Direction in Supervision Planning*, 22 Crim. Just. Connections 1, 1 (2015), https://www.tdcj.texas.gov/connections/JanFeb2015/Images/JanFeb2015_agency_TRASS.pdf, *archived at* https://perma.cc/NJ8X-AWS3.

[4] These states include Nevada and South Dakota. *See* Nev. Bd. of Parole Comm'rs, Nevada Parole Recidivism Risk & Crime Severity Guidelines (2019), http://parole.nv.gov/uploadedFiles/parolenvgov/content/Information/ParoleRiskAssessmentValues.pdf, *archived at* https://perma.cc/3PMZ-8CHG; S.D. Dep't of Corr., 1.5. G.4 Parole- Community Risk Assessment and Supervision of Offenders 9 (2016), https://doc.sd.gov/documents/about/policies/Parole%20Services-Community%20Risk%20Assessment%20and%20Supervision%20of%20Offenders.pdf, *archived at* https://perma.cc/3XMB-X5DN.

[5] For example, the parole boards of Massachusetts, Nebraska, Oregon, and West Virginia use the Level of Service/Case Management Inventory ("LS/CMI"). *See Probation and Parole*, Crime & Just. Inst., http://www.crj.org/divisions/crime-justice-institute/our-work/probation-and-parole, *archived at* https://perma.cc/4BP3-MS6X; Richard L. Wiener et al., Validation Study of the LS/CMI Assessment Tool in Nebraska 1 (2014), https://supremecourt.nebraska.gov/sites/default/files/validation-study-ls-cmi-assessment-tool-ne.pdf, *archived at* https://perma.cc/U6KD-L4QF; Or. Admin. R. 291-078-0020 (2020); Leighann J. Davidson et al., Evidence-Based Offender Assessment: A Comparative Analysis of West Virginia and U.S. Risk Scores 1 (2015), http://cdn2.hubspot.net/hubfs/209455/LS_Blog_/Davidson_et_al_2015_EBP_Offender_Assessment_Comparative_Analysis.pdf, *archived at* https://perma.cc/34WC-J3NH.

example, in Florida,[6] Wisconsin,[7] Michigan,[8] most counties of New York,[9] and the most populous county of New Mexico.[10]

Algorithmic models are also frequently used during sentencing proceedings. During these proceedings, the focus lies primarily on recidivism risk, and the assessment contributes to a judge's determination of a specific offender's sentence. At least ten states use some form of risk-assessment instrument at sentencing, employing either their own tools[11] or commercially available tools.[12] Some states have made public studies on the accuracy of their risk-assessment systems.[13] COMPAS, a commercially available tool that is used for sentencing in multiple states, has generated significant controversy,[14] as we discuss below. This controversy has included recent legal challenges.[15]

---

[6] *See* Julia Angwin et al., *Machine Bias*, ProPublica (May 23, 2016), https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing, *archived at* https://perma.cc/ZQU7-4WQH.

[7] *See COMPAS*, Wis. Dep't Corr., https://doc.wi.gov/Pages/AboutDOC/COMPAS.aspx, *archived at* https://perma.cc/3M7C-YFZV.

[8] *See* Mich. Dep't of Corr., Administration and Use of COMPAS in the Presentence Investigation Report 3 (2017), https://www.michbar.org/file/news/releases/archives17/COMPAS-at-PSI-Manual-2-27-17-Combined.pdf, *archived at* https://perma.cc/3AVJ-FLVC.

[9] *See* Sharon Lansing, New York State COMPAS-Probation Risk and Need Assessment Study: Examining the Recidivism Scale's Effectiveness and Predictive Accuracy (2012), http://www.northpointeinc.com/downloads/research/DCJS_OPCA_COMPAS_Probation_Validity.pdf, *archived at* https://perma.cc/U7M7-QP47. The tool is not used in New York County.

[10] *See* N.M. Corr. Dep't, 2012–2013 Annual Report 21 (2013), https://cd.nm.gov/wp-content/uploads/2019/05/2012-2013_Annual_Report.pdf, *archived at* https://perma.cc/TN8B-VF5W; *see also Busy New Mexico Courts to Implement Risk Assessment Tool*, U.S. News & World Reports (June 1, 2017), https://www.usnews.com/news/best-states/new-mexico/articles/2017-06-01/busy-new-mexico-courts-to-implement-risk-assessment-tool, *archived at* https://perma.cc/T2GX-WNVM.

[11] For example, Ohio uses the Ohio Risk Assessment System ("ORAS"), *see* Ohio Risk Assessment System, Ohio Dep't Rehabilitation & Correction, http://www.drc.ohio.gov/oras, *archived at* https://perma.cc/U2CF-73YP; and Virginia has dependently developed its own system, *see* Va. Criminal Sentencing Comm'n, 2012 Annual Report 6 (2012), http://www.vcsc.virginia.gov/2012VCSCAnnualReport.pdf, *archived at* https://perma.cc/6P8W-57CE.

[12] For example, Colorado and Oklahoma use Level of Service Inventory-Revised ("LSI-R"). *See* Ctr. for Sentencing Initiatives, Use of Risk and Needs Assessment Information at Sentencing: Mesa County, Colorado 2–3 (2013), http://www.ncsc.org/~/media/Microsites/Files/CSI/RNA%20Brief%20-%20Mesa%20County%20CO%20csi.ashx, *archived at* https://perma.cc/A99K-EN2J; Okla. Stat. tit. 22, § 988.17 (2020).

[13] *See* Mo. Sentencing Advisory Comm'n, Recommended Sentencing: Biennial Report 50 (2009), https://www.mosac.mo.gov/file.jsp?id=45469, *archived at* https://perma.cc/8YE5-ECHW; Pa. Comm'n on Sentencing, Interim Report 7: Validation of Risk Scale 2 (2013), http://pcs.la.psu.edu/publications-and-research/research-and-evaluation-reports/risk-assessment/phase-i-reports/interim-report-7-validation-of-risk-scale/view, *archived at* https://perma.cc/9QMP-K865.

[14] *See* Angwin et al., *supra* note 6.

[15] *See, e.g.*, State v. Loomis, 881 N.W.2d 749, 753 (Wis. 2016). In *Loomis*, the Wisconsin Supreme Court held that the use of algorithms—even proprietary algorithms—in sentencing is constitutional, so long as the score generated is only one among many nondeterminative factors considered in the sentencing decision. This case will be discussed in greater detail in Part

Risk-assessment technologies are also used during pretrial release and bail proceedings, where algorithms attempt to forecast the likelihood of a defendant appearing for their next court date, as well as the risk to the community if the defendant is released. For example, Arizona mandates the use of tools to calculate "a pretrial defendant's risk of committing a new crime or failing to appear while on pretrial release for the purpose of assisting the court in determining release decisions and release conditions and to assist the pretrial services staff with supervision monitoring requirements."[16] New Jersey requires judges to justify release decisions that deviate from the algorithmic recommendation.[17]

II. Similarly, in 2019, a Michigan appellate court held that the use of COMPAS at sentencing did not violate defendants' due process rights, noting it served only as a nonbinding recommendation and that the court had discretion to weigh its value. People v. Younglove, No. 341901, 2019 WL 846117, at *3 (Mich. Ct. App. Feb. 21, 2019). In 2018, the Iowa Supreme Court vacated two decisions that had held that there was no statutory basis for utilizing sex offender risk assessment tools at sentencing; each case also declined to reach the defendants' due process claims for procedural reasons. *See* State v. Gordon, 921 N.W.2d 19, 26 (Iowa 2018); State v. Guise, 921 N.W.2d 26, 28 (Iowa 2018). In contrast to such due process claims, a complaint recently filed in the United States District Court for the Western District of Wisconsin alleges that the use of COMPAS in the parole process constitutes unconstitutional racial discrimination. *See* Henderson v. Stensberg, No. 18-cv-555-jdp, 2020 U.S. Dist. LEXIS 48386, at *1 (Mar. 20, 2020). The plaintiff argues that government officials violated the Fourteenth Amendment because they "supported using COMPAS for parole decisions despite knowing that the program is biased against African Americans, and they won't pay Northpointe for [a] corrective upgrade." *Id.* at *2. In a recent Opinion and Order, the court distinguished the plaintiff's equal protection claim from the due process claim considered in *Loomis* and denied Northpointe's motion to dismiss, converting it into a summary judgment motion that is yet to be decided. *See id.* at **11–12.

[16] ARIZ. C.J.A. § 5-201(A) (2014), https://www.azcourts.gov/Portals/0/admcode/pdfcurrentcode/5-201_New_December_2013.pdf, *archived at* https://perma.cc/6B96-N2VG. States using tools for this purpose include also include New Jersey, which uses the Public Safety Assessment ("PSA"), *see* Issie Lapowsky, *One State's Bail Reform Exposes the Promise and Pitfalls of Tech-Driven Justice*, WIRED (Sept. 5, 2017, 7:00 AM), https://www.wired.com/story/bail-reform-tech-justice, *archived at* https://perma.cc/D4S4-BPHH; and Indiana, which uses the Indiana Risk Assessment System ("IRAS") at multiple stages in the criminal justice process, *see* BD. OF DIRS. OF THE JUDICIAL CONFERENCE OF IND., POLICY FOR INDIANA RISK ASSESSMENT SYSTEM (2012), http://www.in.gov/judiciary/cadp/files/prob-risk-iras-2012.pdf, *archived at* https://perma.cc/AS6L-7EWU.

[17] *See* ALEXANDER SHALOM ET AL., THE NEW JERSEY PRETRIAL JUSTICE MANUAL 11 (2016), https://www.nacdl.org/getattachment/50e0c53b-6641-4a79-8b49-c733def39e37/the-new-jersey-pretrial-justice-manual.pdf, *archived at* https://perma.cc/AN2W-239G. This move was challenged in a products liability suit brought on behalf of a man killed by a defendant who was released on a PSA recommendation. *See* Rodgers v. Laura & John Arnold Found., No. 17-5556, 2019 WL 2429574, at *1 (D.N.J. June 11, 2019). The trial judge granted a motion to dismiss under Rule 12(b)(6) on the grounds that PSA is not a "product" for product liability purposes. *Id.* at *3. Failure to show proximate causation was a potential alternate ground for the decision. *Id.* The judge relied on Third Circuit precedent upholding New Jersey's criminal justice reform scheme as a whole. *Id.* (citing Holland v. Rosen, 895 F.3d 272, 278–79 (3d Cir. 2018) (holding that there is no "constitutional right to deposit money or obtain a corporate surety bond to ensure a criminal defendant's future appearance in court as an equal alternative to non-monetary conditions of pretrial release")).

Some states use algorithms to identify inmates' needs, such as mental health treatment, drug treatment, or special monitoring.[18] Other states employ scoring systems to assess recidivism risk of particular classes of offenders, such as sex offenders, in order to classify them.[19]

B. The Risk-Assessment Tools

COMPAS, which is owned by Equivant (formerly known as Northpointe), utilizes a proprietary algorithmic model that cannot be inspected.[20] Because the model is proprietary, independent researchers do not have access to the data upon which the model is based and cannot independently validate the process by which scores are calculated.[21] It is thus impossible to determine precisely how its inputs are weighted in generating recidivism predictions.

While the algorithms underlying COMPAS technology are not disclosed, the way that COMPAS administers its scoring assessments is fairly easy to understand. COMPAS offers both a risk-assessment tool and a risk-needs assessment tool, providing "risk scores" and "needs scores." The risk assessment uses six inputs to assess the likelihood of an individual recidivating between the time they are released and their eventual sentencing hearing.[22] The risk-needs assessment uses 137 inputs to produce numerical scores that are intended to assist in a decisionmaker's determination of con-

---

[18] For example, Alabama uses the Alabama Risk Assessment System ("ARAS"), *see* ALA. DEP'T OF CORR., MINIMUM STANDARDS FOR COMMUNITY PUNISHMENT AND CORRECTIONS PROGRAMS 8 (2016), http://www.doc.state.al.us/docs/AlabamaMinimumStandardsforCCP.pdf, *archived at* https://perma.cc/L7Q4-PCUM; Hawaii uses LSI-R, *see* TIMOTHY WONG, VALIDATION OF THE STATE OF HAWAII LSI-R PROXY 1 (2009), http://icis.hawaii.gov/wp-content/uploads/2013/07/copy2_of_copy_of_SARA-DVSI-Exploratory-Study-Oct-2008.pdf, *archived at* https://perma.cc/Q548-HSSH; and Louisiana uses the Louisiana Risk and Needs Assessment ("LARNA"), *see* LA. SENTENCING COMM'N, RECOMMENDATIONS OF THE LOUISIANA SENTENCING COMMISSION 14 (2012), http://www.lcle.state.la.us/sentencing_commission/2012_biannual_report_lsc_final.pdf, *archived at* https://perma.cc/Y5T9-BRG3. COMPAS provides such a function as well. *See Overwhelmed by Inmates with Special Needs? You're Not Alone*, EQUIVANT (Feb. 26, 2019), https://www.equivant.com/overwhelmed-by-inmates-with-special-needs-youre-not-alone, *archived at* https://perma.cc/Y7UY-VNA7.

[19] For example, Minnesota developed its own Minnesota Sex Offender Screening Tool-Revised for this purpose. DOUGLAS L. EPPERSON ET AL., MINNESOTA SEX OFFENDER SCREENING TOOL—REVISED (MNSOST-R) TECHNICAL PAPER: DEVELOPMENT, VALIDATION, AND RECOMMENDED RISK LEVEL CUT SCORES 2 (2003), https://rsoresearch.files.wordpress.com/2012/01/ia-state-study.pdf, *archived at* https://perma.cc/22BS-MFT8.

[20] *See* Angwin et al., *supra* note 6.

[21] Adam Liptak, *Sent to Prison by a Software Program's Secret Algorithms*, N.Y. TIMES (May 1, 2017), http://nyti.ms/2qoe8FC, *archived at* https://perma.cc/H8J5-65MY. "'The key to our product is the algorithms, and they're proprietary,' one of its executives said last year. 'We've created them, and we don't release them because it's certainly a core piece of our business.'" *Id.*

[22] *Official Response to Science Advances*, EQUIVANT (Jan. 18, 2018), https://www.equivant.com/official-response-to-science-advances, *archived at* https://perma.cc/E73P-PZNC.

ditions of release.[23] A defendant responds to a survey questionnaire—either on paper or in an interview conducted by an administrator—and the answers are sent to Equivant, which then processes the data through its proprietary model and returns a risk score and a needs score.[24]

COMPAS's accuracy is debated and contested.[25] Equivant has conducted internal validation studies,[26] and at least one third-party study has found satisfactory levels of predictive power.[27] However, other studies have called into question COMPAS's predictivity. One Dartmouth study found that COMPAS was no more predictive of recidivism than a linear regression accounting for only the defendant's age and number of prior offenses.[28] We address the constitutional issues posed by COMPAS's proprietary nature and algorithmic validity in Part II.

Another commonly used risk-assessment instrument is the Level of Service Inventory-Revised ("LSI-R"), which purports to be "[t]he most widely used and widely researched risk-need assessment in the world."[29] It is operated by MHS Assessments, a company that also offers a number of risk-assessment tools.[30] Several states that use some variation of LSI have studied

---

[23] Northpointe, Risk Assessment (2011), *archived at* https://perma.cc/4ZY7-U9UD. Inputs include responses to questions on, inter alia, the nature of the offense (e.g., Q2: "Which offense category represents the most serious current offense?"), criminal history (e.g., Q19: "How many prior drug possession/use offense arrests as an adult?"), family criminality (e.g., Q35: "Were your brothers or sisters ever arrested, that you know of?"), social environment (e.g., Q65: "Is there much crime in your neighborhood?"), and self-reported five-point Likert scale responses to prompts (e.g., Q107: "I feel very close to some of my friends."). *Id.*

[24] *See* Northpointe, COMPAS Risk & Need Assessment System (2012), http://www.northpointeinc.com/files/downloads/FAQ_Document.pdf, *archived at* https://perma.cc/V5VT-F7YS.

[25] On the topic of validating the risk assessment models, studies show mixed reports. Often, the model's designer will publish self-validation studies, as in the case of COMPAS. *Id.* However, very few of the recidivism prediction tools in use in the U.S. have been validated. *See* Sarah L. Desmarais & Jay P. Singh, Risk Assessment Instruments Validated and Implemented in Correctional Settings in the United States 2 (2013), https://csgjusticecenter.org/wp-content/uploads/2020/02/Risk-Assessment-Instruments-Validated-and-Implemented-in-Correctional-Settings-in-the-United-States.pdf, *archived at* https://perma.cc/Y4MM-6BFK.

[26] *See, e.g.*, Research & Dev. Dep't, COMPAS Scales and Risk Models Validity and Reliability: A Summary of Results from Internal and Independent Studies 1 (2010), https://epic.org/algorithmic-transparency/crim-justice/EPIC-16-06-23-WI-FOIA-201600805-COMPASSummaryResults.pdf, *archived at* https://perma.cc/SB4F-SUDP.

[27] *See* Thomas Bloomberg et al., Validation of the COMPAS Risk Assessment Classification Instrument 11 (2010), http://criminology.fsu.edu/wp-content/uploads/Validation-of-the-COMPAS-Risk-Assessment-Classification-Instrument.pdf, *archived at* https://perma.cc/K3Y8-UWSS.

[28] Julia Dressel & Hany Farid, *The Accuracy, Fairness, and Limits of Predicting Recidivism*, 4 Sci. Advances 1, 1 (2018).

[29] D.A. Andrews & James Bonta, *Level of Service Inventory-Revised*, MHS Assessments, https://storefront.mhs.com/collections/lsi-r, *archived at* https://perma.cc/4Y2R-72UY.

[30] *See* Press Release, LSI-R, LSI-R:SV, LS/CMI and YLS/CMI Now Available Through Assessments.com (Sept. 18, 2006), https://www.assessments.com/content/press_releases/Assessments.com%20and%20MHS.pdf, *archived at* https://perma.cc/TF6E-FK9N.

the model's validity and concluded that it was sufficiently valid for use in those states' courts.[31]

Some risk-assessment instruments are open-source and thus raise fewer transparency concerns. For example, the Ohio Risk Assessment System ("ORAS"), which is adapted for use in several states,[32] was originally created by the University of Cincinnati, and researchers there have studied its validity.[33] Because it is open-source, many of the criticisms about lack of transparency associated with COMPAS and LSI-R are less applicable to ORAS and other state tools based on it. A similar algorithmic model is the Public Safety Assessment ("PSA"), which was developed by the Laura and John Arnold Foundation and also uses an open-source scoring system.[34] The Foundation has found that the system increases pretrial defendants' attendance rates at subsequent court appearances.[35] Some states that use PSA have also studied its validity.[36]

A 2013 meta-analysis of fifty-three risk-assessment-tool validation studies found that nearly all tools in use had been validated by only one or two studies, and that in most cases, the studies were conducted by the group developing the tool, rather than an independent party.[37] Only two of the reviewed studies evaluated inter-rater reliability—the likelihood that two as-

---

[31] *See, e.g.*, Maureen L. O'Keefe, Kelli Klebe & Scott Hromas, Validation of the Level of Supervision Inventory (LSI) for the Community Based Offenders in Colorado: Phase II 9 (1998) (Colorado), http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.476.8457&rep=rep1&type=pdf, *archived at* https://perma.cc/D56B-6WE6; Wong, *supra* note 18, at 2 (Hawaii); Christopher T. Lowenkamp & Kristin Bechtel, *The Predictive Validity of the LSI-R on a Sample of Offenders Drawn from the Records of the Iowa Department of Corrections Data Management System*, 71 Fed. Prob 25, 30 (2007) (Iowa).

[32] *See, e.g.*, Edward Latessa, Brian Lovins & Matthew Makarios, Validation of the Indiana Risk Assessment System: Final Report ii, 18–19 (2013), https://www.in.gov/judiciary/iocs/files/prob-risk-iras-final.pdf, *archived at* https://perma.cc/AG7A-KQP7 (finding predictive validity r values of 0.27 for men and 0.20 for women).

[33] Edward Latessa et al., Creation and Validation of the Ohio Risk Assessment: Final Report 2 (2009), http://www.ocjs.ohio.gov/ORAS_FinalReport.pdf, *archived at* https://perma.cc/U5KG-7YTA (finding predictive validity r values between 0.22 and 0.44).

[34] *See generally* Laura & John Arnold Found., Public Safety Assessment: Risk Factors and Formula, https://craftmediabucket.s3.amazonaws.com/uploads/PDFs/PSA-Risk-Factors-and-Formula.pdf, *archived at* https://perma.cc/3LPZ-FAHC. For a study comparing the relative predictive validity of a range of risk assessment tools, see Desmarais & Singh, *supra* note 25, at 22.

[35] *New Data: Pretrial Risk Assessment Tool Works To Reduce Crime, Increase Court Appearances*, Arnold Ventures (Aug. 8, 2016), https://www.arnoldventures.org/newsroom/new-data-pretrial-risk-assessment-tool-works-reduce-crime-increase-court-appearances, *archived at* https://perma.cc/9K7F-QWYT (finding that implementing PSA in Lucas County, Ohio, led to a reduction in pretrial defendants skipping their court dates from 41% to 29%).

[36] *See, e.g.*, Matthew DeMichele et al., The Public Safety Assessment: A Re-Validation and Assessment of Predictive Utility and Differential Prediction by Race and Gender in Kentucky 48 (2018), https://craftmediabucket.s3.amazonaws.com/uploads/PDFs/3-Predictive-Utility-Study.pdf, *archived at* https://perma.cc/T7SC-7DT4 (finding PSA scores somewhat predictive of failure to appear ($r = 0.188$) and new criminal activity ($r = 0.171$) and weakly predictive of new violent criminal activity ($r = 0.067$)); *New Data from Ohio Validates PSA Impact*, Arnold Ventures (Aug. 10, 2016), https://www.arnoldventures.org/stories/new-data-ohio-validates-psa-impact, *archived at* https://perma.cc/8T2R-DKSN.

[37] Desmarais & Singh, *supra* note 25, at 1, 19.

sessments of the same subject conducted by different administrators will yield the same result—for these tools.[38]

C. Existing Literature and Debate

Some organizations that have decarceral goals see promise in the use of algorithms, especially in the pretrial detention context,[39] where algorithms may reveal that a large number of low-risk persons are needlessly detained.[40] However, some data scientists argue that these algorithms' predictive ability is so overstated that it would be more accurate to assume that *all* defendants awaiting trial posed no risk of reoffending.[41] Additionally, concerns about racially disparate outcomes of algorithms fuel skepticism about their use.[42] Specific concerns include the impact of racial biases on the input data itself and the limited legal means to address these biases.[43] Risk-assessment algorithms incorporate data points such as prior arrests and income. While facially race-neutral, in practice these data reflect discriminatory policing and the financial impact of structural racism.[44] Some researchers believe that no technical fixes can remedy these algorithms, and they call on states to discontinue the use of such technology.[45]

Other researchers have suggested that, rather than taking a formalistic approach that excludes racially correlated factors, algorithms should consciously take race into consideration to mitigate racially disparate outcomes.[46] Crystal Yang and Will Dobbie, for example, argue that this approach would better uphold the principles of the Equal Protection Clause

---

[38] *Id.* at 20.

[39] *See, e.g.*, *Pretrial Justice*, ARNOLD VENTURES, https://www.arnoldventures.org/work/pretrial-justice, *archived at* https://perma.cc/E2NB-CL7J.

[40] *See* Karen Hao & Jonathan Stray, *Can You Make AI Fairer than a Judge? Play our Courtroom Algorithm Game*, MIT TECH. REV. (Oct. 17, 2019), https://www.technologyreview.com/2019/10/17/75285/ai-fairer-than-judge-criminal-risk-assessment-algorithm, *archived at* https://perma.cc/9SBZ-MEKS.

[41] *See, e.g.*, Chelsea Barabas, Karthik Dinakar & Colin Doyle, *The Problems with Risk Assessment Tools*, N.Y. TIMES (July 17, 2019), https://nyti.ms/2SiUcT2, *archived at* https://perma.cc/ST5Z-CFGS.

[42] *See* VINCENT SOUTHERLAND & ANDREA WOODS, WHAT DOES FAIRNESS LOOK LIKE? CONVERSATIONS ON RACE, RISK ASSESSMENT TOOLS, AND PRETRIAL JUSTICE 10 (2018), http://www.law.nyu.edu/sites/default/files/Final%20Report—ACLU-NYU%20CRIL%20Convening%20on%20Race%20Risk%20Assessment%20%20Fairness.pdf, *archived at* https://perma.cc/58KB-QRQP.

[43] *Id.*

[44] Stephen Goldsmith & Chris Bousquet, *The Right Way to Regulate Algorithms*, CITYLAB (Mar. 20, 2018), https://www.citylab.com/equity/2018/03/the-right-way-toregulate-algorithms/555998, *archived at* https://perma.cc/8QZF-MFSG.

[45] *See, e.g.*, CHELSEA BARABAS ET AL., TECHNICAL FLAWS OF PRETRIAL RISK ASSESSMENTS RAISE GRAVE RISK 4 (2019), https://dam-prod.media.mit.edu/x/2019/07/16/TechnicalFlawsOfPretrial_ML%20site.pdf, *archived at* https://perma.cc/R25U-AZXE.

[46] *See, e.g.*, Crystal S. Yang & Will Dobbie, *Equal Protection Under Algorithms: A New Statistical and Legal Framework* 33–36 (Harvard Univ., Working Paper, 2009), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3462379, *archived at* https://perma.cc/N6PN-HMQN; Jon Kleinberg et al., *Discrimination in the Age of Algorithms*, 10 J. LEGAL ANALYSIS

and have proposed and tested potential algorithms that would implement this proposal.[47] Indeed, an accurate algorithmic prediction can be expected to have a disparate racial impact if the prevalence of recidivism varies across racial groups due to prevailing social inequalities.[48] At least one foreign jurisdiction has been open to the view that an algorithm's failure to take race into account is problematic: the Supreme Court of Canada ruled that a member of the Métis tribe could pursue an appeal on the grounds that the risk-assessment tool that had been used to evaluate him had not been sufficiently trained and validated on indigenous populations.[49]

Further, since scores are provided to a human decisionmaker, the decisionmaker's own biases may compound any disparate outcomes. One study presented participants with dummy defendant profiles that included the defendants' race and risk scores.[50] Even though they could see the risk scores, participants tended to rate Black defendants as more likely to skip bail or be re-arrested before trial than their risk scores predicted; by contrast, participants rated white defendants less likely to skip bail or be re-arrested than their risk scores predicted.[51] Judges have been found to harbor the same kinds of implicit biases as the general population.[52] In one survey of judicial attitudes toward risk assessments, judges "indicated a general belief among members of the judiciary that their judgment was more accurate than actuarial at-sentencing assessments,"[53] which suggests that judges are willing and may even prefer to deviate from algorithmic recommendations.

Scholars also argue that there is an inherent trade-off between predictive accuracy and racial fairness.[54] Reformers and scholars debate whether it is better to use a risk-assessment tool that treats everyone equally, regardless of racial group, or a tool that takes into account existing inequalities and ensures that no group bears the brunt of the tool's errors. Even for groups

---

113, 154–60 (2019); Cynthia Dwork et al., *Fairness Through Awareness*, arXiv 22 (Nov. 29, 2011), https://arxiv.org/pdf/1104.3913.pdf, *archived at* https://perma.cc/WW34-AJEZ.

[47] Yang & Dobbie, *supra* note 46, at 36–37.

[48] *See* Alexandra Chouldechova, *Fair Prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments*, 5 Big Data 153, 160–62 (2017).

[49] Ewert v. Canada, 2 S.C.R. 165, 166–67 (S.C.C. 2018) (finding that the cultural bias of a risk assessment tool that had not been validated on a native population could have either overestimated Ewert's risk and denied him his rights or underestimated his risk and undermined the state's interest in protecting society).

[50] Ben Green & Yiling Chen, Disparate Interactions: An Algorithm-in-the-Loop Analysis of Fairness in Risk Assessments 3 (Jan. 29–31, 2019) (unpublished conference paper), https://scholar.harvard.edu/files/19-fat.pdf, *archived at* https://perma.cc/U3WN-NPGY.

[51] *Id.* at 7.

[52] *See, e.g.*, Jeffrey J. Rachlinski et al., *Does Unconscious Racial Bias Affect Trial Judges?*, 84 Notre Dame L. Rev. 1195, 1221 (2009).

[53] Steven L. Chanenson & Jordan M. Hyatt, *The Use of Risk Assessment at Sentencing: Implications for Research and Policy* 10 (Villanova Univ. Charles Widger Sch. of Law, Working Paper, 2016), https://digitalcommons.law.villanova.edu/cgi/viewcontent.cgi?article=1201&context=wps, *archived at* https://perma.cc/CKB5-5N69.

[54] *See, e.g.*, Jon Kleinberg, Sendhil Mullainathan & Manish Raghavan, *Inherent Trade-Offs in the Fair Determination of Risk Scores*, arXiv 17 (Nov. 17, 2016), https://arxiv.org/pdf/1609.05807.pdf, *archived at* https://perma.cc/ZTR4-Y8Z7.

that see the potential for algorithms to improve the criminal justice system, the answer is unclear.[55] One proposed way of identifying the proper balance point is to evaluate the tool's overall impact on racial stratification, rather than its performance in discrete cases.[56] Under such an approach, false positives, where a defendant who would not reoffend is classified as "high risk," would be of less concern, provided that there was a positive net benefit to the racial group as a whole.[57]

An even sharper critique of risk-assessment algorithms asserts that they have no basis in the Anglo-American tradition of criminal sentencing because they subject the individual to punishment for characteristics over which the individual has no meaningful control.[58] The critique further takes issue with the presumption that whether a defendant will recidivate is essentially predetermined, which denies the defendant's agency.[59] Indeed, basing sentencing and bail determinations on the likelihood of recidivism arguably asks the wrong question. Perhaps the relevant inquiry is not who is most likely to recidivate, but whose likelihood of recidivism is most likely to be reduced by incarceration.[60]

Some critics have identified the focus on recidivism in sentencing as posing particular concerns. While the decision to detain a defendant pending trial is binary, the sentencing decision is not binary. Sentencing requires the judge to decide the severity of the sentence. Since there is little evidence that a longer sentence will impact a defendant's recidivism,[61] some scholars have questioned whether a higher likelihood of recidivism should necessarily result in a longer sentence.[62]

There have been numerous critiques of algorithms due to their lack of transparency, with some proposing a "technological due process" concept in light of artificial intelligence developments and advocating for transparency, accuracy, accountability, participation, and fairness.[63] But transparency in particular may conflict directly with the profit motives of companies that offer proprietary algorithms.[64]

---

[55] *See* SOUTHERLAND & WOODS, *supra* note 42, at 12.

[56] *See* Aziz Huq, *Racial Equity in Algorithmic Criminal Justice*, 68 DUKE L.J. 1043, 1128–29 (2019).

[57] *Id.* at 1127–28.

[58] *See* Dawinder S. Sidhu, *Moneyball Sentencing*, 56 B.C. L. REV. 671, 707–09 (2015).

[59] *See id.* at 713–14.

[60] Sonja B. Starr, *Evidence-Based Sentencing and the Scientific Rationalization of Discrimination*, 66 STAN. L. REV. 803, 855 (2014).

[61] *Id.* at 856.

[62] Danielle Kehl, Priscilla Guo & Samuel Kessler, Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing 14 (July 2017) (unpublished paper), http://nrs.harvard.edu/urn-3:HUL.InstRepos:33746041, *archived at* https://perma.cc/Q9YC-235Y.

[63] *See, e.g.*, Danielle Keats Citron & Frank Pasquale, *The Scored Society: Due Process for Automated Predictions*, 89 WASH. L. REV. 1, 19–20 (2014).

[64] Kehl, Guo & Kessler, *supra* note 62, at 33.

With complex algorithms, mere transparency may not suffice to allow a decisionmaker to understand them. In the context of algorithmic models, "explainability" refers to "the ability to explain or to present in understandable terms to a human."[65] In one view of explainability, presented by Finale Doshi-Velez and Been Kim, the relationship between an algorithm's inputs and outputs should be evaluated in the context of the ultimate task it seeks to perform.[66] A model that is more explainable is also a model that allows the human interpreter to more readily identify when the model has produced an incorrect result, to learn new information, and to discriminate less.[67]

II. Due Process

Embedded in due process are two core principles: 1) criminal defendants should be tried and punished as individuals, judged by their own actions and characteristics,[68] and 2) criminal defendants have the right to review and challenge the evidence used to determine guilt and punishment.[69] While all algorithms may be criticized for treating criminal defendants as statistics rather than as individuals, proprietary algorithms, such as COMPAS, are especially troubling because defendants cannot see, understand, or challenge the findings used against them.

Today, several states require that judges be provided with risk assessments and recidivism data at sentencing.[70] The training data and algorithmic

---

[65] Finale Doshi-Velez & Been Kim, *Towards a Rigorous Science of Interpretable Machine Learning*, arXiv 2 (Mar. 2, 2017), https://arxiv.org/pdf/1702.08608.pdf, *archived at* https://perma.cc/URR3-633R.

[66] *Id.* at 4–5.

[67] *Id.*

[68] *See* Williams v. New York, 337 U.S. 241, 247 (1949) ("Highly relevant—if not essential—to [a judge's] selection of an appropriate sentence is the possession of the fullest information possible concerning the defendant's life and characteristics. And modern concepts individualizing punishment have made it all the more necessary that a sentencing judge not be denied an opportunity to obtain pertinent information . . . . The belief no longer prevails that every offense in a like legal category calls for an identical punishment without regard to the past life and habits of a particular offender.").

[69] *See* In re Oliver, 333 U.S. 257, 273 (1948); *see also* Hamdi v. Rumsfeld, 542 U.S. 507, 533 (2004) ("For more than a century the central meaning of procedural due process has been clear: 'Parties whose rights are to be affected are entitled to be heard; and in order that they may enjoy that right they must first be notified.' It is equally fundamental that the right to notice and an opportunity to be heard 'must be granted at a meaningful time and in a meaningful manner.'") (citations omitted).

[70] *See, e.g.*, KY. Rev. Stat. Ann. § 532.007(3)(a) (West, Westlaw through Ch. 73 of 2020 Reg. Sess.) (requiring sentencing judges to consider "the results of a defendant's risk and needs assessment included in the presentence investigation"); Ohio Rev. Code Ann. § 5120.114(A)(1)–(3) (West, Westlaw through File 130 of 133rd General Assemb. (2019–2020)) (requiring the Ohio Department of Rehabilitation and Correction to adopt the use of a single risk assessment tool when risk assessments are ordered by a court in sentencing or other purposes); 42 Pa. Cons. and Stat. Ann. § 2154.7(a) (West, Westlaw through 2020 Reg. Sess. Act 13) (requiring the adoption of "a sentence risk assessment instrument for the sentencing court to use to help determine the appropriate sentence within the limits established by law"); *see also* Ariz. C.J.A. § 6– 201.01(J)(3) (2020), https://www.azcourts.gov/Portals/0/admcode/pdfcurrentcode/6-201.01_Amended_01-15-20.pdf?ver=2020-01-17-081823-580,

modeling that underlie certain risk-assessment tools, such as COMPAS, are held as trade secrets by the technology's developers. The question is whether this "black box" methodology violates the due process rights of criminal defendants by denying them the opportunity to challenge their output risk scores, or the means by which those scores were calculated. At least one court—the Wisconsin Supreme Court—has held that the use of algorithmic risk assessments does not violate due process.[71] In this section, we will discuss why the Due Process Clause and other policy justifications do require courts to give expanded opportunities for defendants to challenge the validity of risk assessments. We will then explain why the Wisconsin Supreme Court failed to properly address the due process concerns. Finally, we will discuss how developments in explainable artificial intelligence can help address procedural deficiencies of current risk-assessment tools.

A. Opportunity to Challenge Algorithmic Risk Calculations

In 1977, the Supreme Court in *Gardner v. Florida*[72] found a violation of the Due Process Clause when a sentencing judge sentenced a defendant to death after reviewing a presentence report that contained a "confidential" portion.[73] This portion of the report was not disclosed to the defendant or his counsel, who thus could not contest the accuracy of the information contained therein.[74] By sentencing the defendant without disclosing or explaining the confidential information contained in the presentence report, the sentencing judge had violated the defendant's due process rights by preventing him from challenging the factual findings that were used to justify his sentence.[75]

There is no such occasion for a defendant to challenge the accuracy or materiality of the conclusions presented in proprietary algorithmic risk assessments. Defendants, defense counsel, and even sentencing judges cannot review the method by which algorithmic risk-assessment tools reach conclusions. While considering a defendant's background is constitutionally permitted, risk-assessment algorithms weigh and process details about a defendant's background in ways that the defendant can neither analyze nor challenge. Processing information by means of a proprietary algorithm may thus violate due process because a defendant cannot effectively challenge

archived at https://perma.cc/7EUN-MR3C ("For all probation eligible cases, presentence reports shall [ ] contain case information related to criminogenic risk and needs as documented by the standardized risk assessment and other file and collateral information"); OKLA. STAT. ANN. tit. 22, § 988.18(B) (West, Westlaw through Ch. 5 of Second Reg. Sess. of 57th Leg. (2020)) (requiring the use of the Level of Services Inventory (LSI) or other risk assessment to evaluate defendants prior to sentencing in order to determine defendant's pro-social needs, recidivism risk, and appropriateness of "various community punishments").

[71] *See* State v. Loomis, 881 N.W.2d 749, 767 (Wis. 2016).

[72] 430 U.S. 349 (1977).

[73] *See id.* at 351.

[74] *Id.*

[75] *Id.* at 362.

the accuracy of the process through which the algorithm reaches its conclusions. Just as judges must provide legally valid explanations for their decisions,[76] defendants should be allowed to challenge the method by which an algorithm reaches its conclusion.

To better understand how the use of risk-assessment instruments such as COMPAS may implicate the Due Process Clause, it is helpful to compare them to presentence reports commonly used in federal sentencing determinations. Algorithmic risk assessments and federal presentencing reports feature similar information about the defendant. Federal probation officers file presentence investigation reports[77] containing information about a defendant's criminal record, financial condition, and other "circumstances affecting [the defendant's] behavior that may be helpful in imposing [a] sentence or in correctional treatment."[78] These reports contain calculations of the defendant's applicable guideline range according to the Sentencing Guidelines, based on the defendant's personal characteristics, much like the calculations underlying the recidivism predictions offered in algorithmic risk assessments.[79]

Congress amended the Federal Rules of Criminal Procedure in 1974 to require disclosure of presentence reports to defendants.[80] Before 1974, federal courts widely held that trial courts retained discretion over whether or not to disclose the contents of presentence reports to defendants.[81] Since the 1974 amendments, federal criminal defendants are entitled to review the contents of their presentence reports and to challenge information contained

---

[76] As a basic procedural requirement, federal sentencing courts must explain the sentences they impose in order to give appellate courts a basis for reviewing the sentence, regardless of whether the sentence is within the discretionary guideline range. *See* Gall v. United States, 552 U.S. 38, 51 (2007); United States v. Carty, 520 F.3d 984, 992 (9th Cir. 2008) (en banc). However, this requirement is derived from statutory, not constitutional, law. *See* 18 U.S.C.A. § 3553(c) (West 2018). *But see* United States v. Collins, 684 F.3d 873, 887 (9th Cir. 2012) (quoting United States v. Rudd, 662 F.3d 1257, 1260 (9th Cir. 2011)) ("As a matter of procedural due process, 'a sentencing judge must explain a sentence sufficiently' to communicate 'that a reasoned decision has been made' and 'permit meaningful appellate review.'"). No Supreme Court decision has incorporated the explanation requirement to state courts. *See* 16C C.J.S. *Constitutional Law* § 1744 (2020). Some state courts do not require sentencing courts to explain their decisions. *See*, *e.g.*, Riley v. State, 480 So. 2d 32, 34 (Ala. Crim. App. 1985); State v. Welsh, 245 N.W.2d 290, 296–97 (Iowa 1976); People ex rel. Dubinsky v. Conboy, 337 N.Y.S.2d 876, 878 (N.Y. App. Div. 1972); Sullivan v. Cupp, 634 P.2d 288, 289 (Or. Ct. App. 1981).

[77] FED. R. CRIM. P. 32(c)(1)(A).

[78] FED. R. CRIM. P. 32(d)(2).

[79] *Id.*

[80] The disclosure requirements are now codified in the Federal Rules. *See* FED. R. CRIM. P. 32(e).

[81] *See* United States v. Stidham, 459 F.2d 297, 299 (10th Cir. 1972) ("We have held that it is not a violation of due process to deny a defendant's request to see the presentence report."); *see also* United States v. Lowe, 482 F.2d 1357, 1358–59 (6th Cir. 1973); United States v. McKinney, 450 F.2d 943, 943 (4th Cir. 1971); United States v. Virga, 426 F.2d 1320, 1323 (2d Cir. 1970); Cook v. Willingham, 400 F.2d 885, 885 (10th Cir. 1968). *But see* United States v. Picard, 464 F.2d 215, 220 (1st Cir. 1972) (ruling that presentence reports must be disclosed to defendants at sentencing as a matter of policy, but not as a constitutional requirement).

therein.[82] Such challenges may target the veracity of the report's factual assertions and result in the defendant's entitlement to further evidentiary review by the sentencing court.[83]

Defendants' legal right to challenge a federal presentence report is instructive in the present context. Because information considered by predictive algorithmic tools is similar, if not identical, to that in federal presentencing reports, state and federal courts should fashion procedural rules similar to the federal rules to allow defendants to challenge the accuracy of algorithmic risk assessments.

Challenging the use of risk-assessment tools is also analogous to challenging sentencing-range determinations. Once aggravating and mitigating factors are considered and a sentencing range has been calculated, a defendant may object to their prescribed offense level.[84] Consider a hypothetical defendant who has pleaded guilty to abusive sexual contact. The Sentencing Guidelines prescribe a certain offense level for abusive sexual contact, depending on the specific crime for which the defendant is sentenced.[85] That offense level may be increased two levels if the court finds that the defendant was entrusted with the custody, care, or supervisory control of the victim.[86] The defendant pleading guilty to abusive sexual contact may object to a two-level increase in the defendant's offense level by challenging the factual finding that the defendant was entrusted with "custody, care, or supervisory control" of the victim. If successful, the challenge has the potential ultimately to lower the sentence under the Sentencing Guidelines.[87]

Challenging an algorithmic risk assessment should include an opportunity for the defendant to examine the underlying algorithm. Specifically, the defendant must be able to assess whether the algorithm properly calculated

---

[82] The Federal Rules establish the procedure for objecting to the contents of a presentence report, including "objections to material information, sentencing guideline ranges, and policy statements contained in or omitted from the report." FED. R. CRIM. P. 32(f).

[83] *See* United States v. Heckel, 570 F.3d 791, 795 (7th Cir. 2009); United States v. Curran, 926 F.2d 59, 62 (1st Cir. 1991); United States v. Fernandez-Angulo, 897 F.2d 1514, 1516 (9th Cir. 1990) ("[W]hen the defendant challenges the factual accuracy of any matters contained in the presentence report, the district court must, at the time of sentencing, make the findings or determinations required by Rule 32."); United States v. Romano, 825 F.2d 725, 729 (2d Cir. 1987) (holding that however broad the district court's discretion may be in determining the appropriate procedure for affording the defendant an opportunity to challenge information in a presentence report, "some process was due by which [defendant] could challenge the accuracy of pre-sentence information presented to the district court"); United States v. Duerksen, 782 F.2d 132, 132 (8th Cir. 1986). We conceptualize the policy motivating the Federal Rules—benefits of disclosure to defendants—to inform the way that courts should think about the Due Process Clause. Since there is a large gap in the case law about how to deal with algorithmic decisions, the Federal Rules of Criminal Procedure can be seen as persuasive authority in how courts should shape due process requirements in both state and federal criminal cases.

[84] *See* FED. R. CRIM. P. 32(i).

[85] U.S. SENTENCING GUIDELINES MANUAL § 2A3.4 (U.S. SENTENCING COMM'N 2018), https://www.ussc.gov/guidelines/2018-guidelines-manual/2018-chapter-2-c#NaN, *archived at* https://perma.cc/6HB5-QCVS.

[86] *Id.*

[87] *See, e.g.*, United States v. Swank, 676 F.3d 919, 921 (9th Cir. 2012).

recidivism risk using both accurate factual inputs about the defendant and legally permitted manipulations of those inputs.[88] To illustrate, consider the following mathematical example: 2_2 = 4. We know the inputs: 2 and 2. We know the output: 4. What we do not know is the function of the computation, "_". Could it be that the equation is 2+2? 2×2? 2^? This may not be of mathematical significance because the output, 4, remains constant no matter the computation. But it could be of legal significance if, for example, multiplication was constitutionally impermissible. Just as defendants have an interest in understanding and challenging the factors that determined their sentencing offense level, they also have an interest in understanding and challenging the inputs and calculations that an algorithm uses to estimate risk. In both situations, a defendant's personal circumstances and characteristics are factored into the ultimate calculation. And in both, the defendant is in the best position to correct the record in the event of a miscalculation. Thus, the reasons for allowing the defendant to challenge the calculation of a sentencing range[89] should similarly allow the defendant to challenge an algorithmic risk assessment.

B. Due Process Implications of COMPAS

In *State v. Loomis*,[90] the Wisconsin Supreme Court addressed a due process challenge to the use of COMPAS assessments at sentencing, ruling that these assessments could be considered as long as they were not dispositive of the ultimate sentencing decision.[91] However, the *Loomis* Court did not substantively engage with the arguments we articulated above; rather, the court dismissed the due process concerns presented by the COMPAS model with little meaningful analysis.

The court reasoned that the use of COMPAS assessments does not deny defendants individualized sentencing, because judges have discretionary authority to override the COMPAS assessments.[92] According to the court, rather than reducing the defendant to a set of factors, the presentence report (which included the COMPAS assessment) helped to provide the sentencing judge with pertinent information.[93] As we have argued above, however, this misses the point: if criminal defendants do not know how COMPAS assessments are made, they lack a meaningful opportunity to challenge their accuracy. More fundamentally, algorithms trained with group data cannot provide *individualized* assessments. While appearing to provide individual-

---

[88] Although COMPAS uses inputs provided by the defendant, other tools such as the PSA use data from administrative sources that may not be available to defendants. *See* LAURA & JOHN ARNOLD FOUND., *supra* note 34.

[89] *See, e.g.*, FED. R. CRIM. P. 32.

[90] 881 N.W.2d 749 (Wis. 2016).

[91] *See id.* at 765.

[92] *See id.*

[93] *See id.*

ized assessments, risk assessments may improperly influence a judge's decision because they provide only generalized conclusions based on group data. Judicial discretion in sentencing is meant to serve individualization; including non-individualized assessments in presentence reports does the opposite.

Responding to this argument, the Wisconsin Supreme Court reasoned that a defendant's ability to verify the accuracy of facts about his criminal history and of his answers to the questionnaire analyzed by COMPAS was a sufficient means of protesting the veracity of the algorithmic assessment; thus, there was no need to challenge what parts of the questionnaire the algorithm could consider.[94] While the court properly emphasized the importance of accurate information, it left unaddressed the core issue of the case: the inability to challenge the *inputs* the algorithm considered and the *methodology* of its analysis.

The *Loomis* court did not meaningfully engage with the due process issues surrounding COMPAS. As the concurrence noted, the court was offered amicus briefing from COMPAS developers that could have shed light on the nature of the COMPAS algorithm, but chose not to allow this briefing.[95] By punting the due process implications of this developing technology, the Wisconsin Supreme Court did not settle the issue—it merely postponed the debate.

C. Due Process and Explainable Algorithms

Due process requires that criminal defendants be given notice and an opportunity to be heard.[96] These requirements protect defendants from arbitrary government actions by forcing the government to explain the bases of its actions and by allowing defendants to challenge them.[97] However, these protections are only meaningful if those bases can be understood by defendants, their counsel, and the court. Not being privy to the reasoning behind their sentencing or bail determination means that defendants cannot meaningfully challenge such decisions.

A possible solution to providing adequate notice might be to disclose to defendants the decision rules powering the algorithms. However, decision-rule disclosure may still fail to provide an adequate explanation for an algorithm's conclusion. Andrew D. Selbst and Solon Barocas describe a number of ways that algorithmic models fail to provide adequate explanations for their decisions.[98] First, interpretation of a model's code may require special-

---

[94] *See id.* at 761.

[95] *See id.* at 774 (Abrahamson, J., concurring).

[96] *See* Cleveland Bd. of Educ. v. Loudermill, 470 U.S. 532, 542 (1985).

[97] *See* Ponte v. Real, 471 U.S. 491, 495 (1985) ("The touchstone of due process is freedom from arbitrary governmental action . . . .").

[98] *See generally* Andrew D. Selbst & Solon Barocas, *The Intuitive Appeal of Explainable Machines*, 87 FORDHAM L. REV. 1085 (2018). Selbst and Barocas also discuss how the existence of automated decisionmaking processes are often kept secret. *Id.* at 1093. This Arti-

ized knowledge of computer science and statistics,[99] which many defendants and defense attorneys lack. Understanding these models may require attorneys to develop specialized knowledge of algorithms or to consult computer experts. Both of these options may be unavailable.

Second, even with computer expertise, the patterns incorporated within the model may be so complex that a human would not be able to understand them.[100] While simple models may be more readily interpretable, more complex machine-learning models cannot easily be understood by examining the decision rules underlying the algorithms. Especially complex machine-learning models may require millions of calculations to reach their decisions, a process that would be impossibly difficult for a human fully to comprehend.[101] In the case of these more complicated models, a typical defendant or defense counsel might be unable easily to identify the factors that are most significant in reaching a predictive output. Furthermore, the specific data and decision labels used to train a model may not be representative of the population to which its predictions are applied. Machine-learning algorithms do not universally provide simple insight into their inner workings and may require the power of a computer to interpret.

While providing the raw code for an algorithm may fail to provide adequate notice, risk-assessment tools can potentially generate explanations for their decisions that satisfy the notice requirement of the Due Process Clause. Machine-learning scholars are working to develop methods for interpreting machine-learning models that render the models more "explainable," so as to enhance "the degree to which a human can understand the cause of [an algorithmic] decision."[102] Explainable models would help defendants understand and challenge the bases for their risk classifications in bail and sentencing. Explanations may take various forms, including decision trees[103] and local interpretable model-agnostic models.[104]

Machine-learning scholars argue that the level of interpretability necessary for a certain task varies with the context of the task.[105] Due Process protections make the need for interpretability clear: defendants are entitled to explanations for the bases of a decision affecting them, necessary for ade-

---

cle assumes that the defendant knows of the existence of the algorithm. Failing to disclose the existence of an algorithm may implicate other constitutional issues not explored here.

[99] *Id.* at 1093–94.

[100] *See id.* at 1094–96 (describing the difficulty inherent in forming an intuitive understanding of complex models as "inscrutability"); *see also* Jenna Burrell, *How the Machine 'Thinks': Understanding Opacity in Algorithms,* 3 Big Data & Society 1, 4–5 (2016).

[101] *See* Christopher Molnar, Interpretable Machine Learning: A Guide for Making Black Box Models Explainable § 7 (2019) (ebook).

[102] *See id.* § 2.

[103] Alex A. Freitas, *Comprehensible Classification Models: A Position Paper*, 15 SIGKIDD Explorations 1, 1 (2014).

[104] Marco Tulio Ribiero et al., *Why Should I Trust You?: Explaining the Predictions of Any Classifier*, *in* Proceedings Of The 22nd ACM SIGKDD International Conference On Knowledge Discovery And Data Mining (2016).

[105] Doshi-Velez & Kim, *supra* note 65, at 1.

quate notice and ability to challenge the decision. Explainable algorithms may serve to provide notice of the bases for an algorithmic decision in a manner that is understandable to a human judge.

Doshi-Velez and Kim provide some general considerations that should be used when evaluating the need for explainability of algorithms in performing different tasks.[106] First, explanations can be global or local.[107] Global explanations describe the algorithm's approach to a task in general, whereas local explanations focus on explanations for individual decision.[108] For defendants who are analyzed by risk-assessment tools, an algorithm should provide both global and local explanations. Global explanations would describe factors that affected the algorithm's approach to classifying defendants in general, such as the number of prior convictions or age. Local explanations would indicate which of the factors were most significant for an individual defendant's classification. Some tools already provide a limited explanation for their decisions by listing significant factors considered by the algorithm.[109] However, many tools currently lack local explanations for classifications of individual defendants. Without local explanations, defendants would find it difficult to challenge the validity of the decision in their unique cases.

The quality of an explanation given by an algorithm also depends on the expertise and resources available to the reader.[110] In the sentencing context, a defense attorney would have the same amount of time to challenge an algorithm's decision as to challenge a human judge's bail or sentencing decisions—as little as a few minutes for some bail arguments. Criminal defense attorneys must therefore have a strong ability to understand and challenge the validity of criminogenic factors in bail and sentencing, even if they lack specialized knowledge in computer science and statistics. If risk-assessment tools are to meet constitutional due process requirements, they should provide explanations that are simple enough for criminal defense attorneys to understand without the need for expertise in statistics.

Companies developing risk-assessment tools have a strong incentive to prevent publication of the code underlying their programs as proprietary trade secrets. Explainable machine-learning algorithms may alleviate due process concerns about lack of transparency and notice while allowing companies to protect their trade secrets.[111] A court may rule that the notice requirement of the Due Process Clause is satisfied by a self-generated explanation for an algorithm's risk classification of an individual. Companies

---

[106] *Id.* at 7–8.
[107] *Id.*
[108] *Id.*
[109] *See* Laura & John Arnold Found., *supra* note 34, at 2.
[110] Doshi-Velez & Kim, *supra* note 65, at 8.
[111] *See* 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, 64–65 (2017).

like Equivant could facilitate the state's compliance with due process requirements by programming such explanations into the tools without disclosing their proprietary code.

However, depending on an algorithm's self-generated explanation, failure to disclose the rules powering a risk-assessment tool may still deny defendants the opportunity to challenge the validity of the risk assessment. While a self-generated explanation for a risk assessment would provide more notice to defendants, due process may also require that defendants be able to challenge the validity of the bases used to reach the decision. Selbst and Barocas argue that automated decisions should not be based on factors that seem unintuitive to human beings.[112] Even if defendants knew every factor that determined their risk assessment, the explanation would be unsatisfactory if the determining factors were illogical or seemingly nonsensical. For example, imagine that COMPAS labeled a defendant as high risk, but its self-generated explanation stated that the most significant factor in its decision was the last digit of the defendant's Social Security number. While disclosing that factor would identify the basis for the decision, the defendant would be unable to understand the decision because the leading factor is so counterintuitive. Thus, due process may require that defendants be allowed to see the rules and further analyses powering algorithmic risk-assessment tools because simply identifying leading factors of a model might fail to adequately explain its decision.

Both human judges and algorithms powered by artificial intelligence form judgments based on their prior experiences.[113] However, it can be difficult for a judge to precisely explain how the judge formed a judgment based on those experiences. Explainable artificial intelligence allows algorithm designers to provide robust explanations for the algorithm's outputs, which empower third parties to analyze the statistical processes. Thus, while the explanations given by algorithms may fail to provide due process for defendants today, explainable algorithms may allow for greater due process protections in the future. Specifically, it is possible to improve procedural justice for defendants by requiring explanations for conclusions arrived at through algorithmic processes, because such explanations could prove more detailed and rigorous than those provided by human judges.

---

[112] Selbst & Barocas, *supra* note 98, at 1096–99.

[113] *See, e.g.*, Justice Michael B. Hyman, *Implicit Bias in the Courts*, 102 ILL. B.J. 40, 43–44 (2014); Jeffrey J. Rachlinski et al., *Does Unconscious Racial Bias Affect Trial Judges?*, 84 NOTRE DAME L. REV. 1195, 1221 (2009); Donald C. Nugent, *Judicial Bias*, 42 CLEV. ST. L. REV. 1, 19–20 (1994) (noting that literature "amply supports" the proposition that "judges' early lives, their experiences both on and off the bench, and their professional careers instill in them certain ideas, beliefs and attitudes about issues and people (including oneself)," and that "while this dynamic often results in reasonable judgments, it also leads to many distorted and systematically biased decisions").

III. Race Discrimination

It is well-documented that algorithmic tools statistically discriminate against minority defendants.[114] Studies show bias both against Black defendants and in favor of white defendants: while Black defendants are more likely to be incorrectly flagged as being at a high risk of recidivism, white defendants are more likely to be mislabeled as low risk.[115] As Supreme Court Justice Lewis Powell noted in *Batson v. Kentucky*,[116] sometimes the "result bespeaks discrimination."[117] As a result of these disparities in outcomes, in 2014, then-Attorney General Eric Holder called on the U.S. Sentencing Commission to study the use of risk-assessment tools, warning that they "may exacerbate unwarranted and unjust disparities that are already far too common in our criminal justice system and in our society."[118]

To illustrate the discriminatory nature of these risk-assessment instruments, we highlight certain statistical disparities arising from the COMPAS predictive model. The following analysis is based on more than 7,000 individual COMPAS assessments obtained through state Right-to-Know Law requests.

Julia Dressel and Hany Farid have found that COMPAS predicts the recidivism risk for white and Black defendants with similar levels of accu-

---

[114] *See, e.g.*, Angwin et al., *supra* note 6 (finding that, even when controlling for prior crimes, future recidivism, age, and gender, Black defendants were 45% more likely to be assigned higher risk scores than white defendants); Bernard H. Harcourt, *Risk as a Proxy for Race,* 4–10 (Univ. of Chi. Pub. Law & Legal Theory, Working Paper No. 323, 2010) (concluding that as sentencing has grown to focus more and more on prior criminal records to the exclusion of other factors, incarceration rates for non-whites has increased, and thus criminal history is a proxy for race and should be avoided in sentencing algorithms), https://chicagounbound.uchicago.edu/cgi/viewcontent.cgi?article=1265&context=public_law_and_legal_theory, *archived at* https://perma.cc/9ULC-YMQE; Mark E. Olver et al., *Thirty Years of Research on the Level of Service Scales: A Meta-Analytic Examination of Predictive Accuracy and Sources of Variability,* 26 Psychol. Assessment., 156–76 (2014) (finding that ethnic minorities receive higher LSI-R scores than non-minorities) ("One possibility may be that systematic bias within the justice system may distort the measurement of 'true' recidivism."); Kevin W. Whiteacre, *Testing the Level of Service Inventory–Revised (LSI-R) for Racial/Ethnic Bias*, 17 Crim. Just. Pol'y Rev. 330, 336 (2006) (finding, in a study of 532 male residents of a work-release program using the LSI-R, that 42.7% of Black defendants were incorrectly overclassified as high-risk compared with 27.7% of white defendants and 25% of Hispanic defendants). *But see* Jennifer L. Skeem & Christopher T. Lowenkamp, *Risk, Race, & Recidivism: Predictive Bias and Disparate Impact*, 54 Criminology 680, 680 (2016) (finding little evidence of bias); John Monahan & Jennifer L. Skeem, *Risk Assessment in Criminal Sentencing*, 12 Ann. Rev. Clinical Psychol., 489, 499 (2016) (resisting the notion that recidivism factors are a proxy for race).

[115] *See* Angwin et al., *supra* note 6.

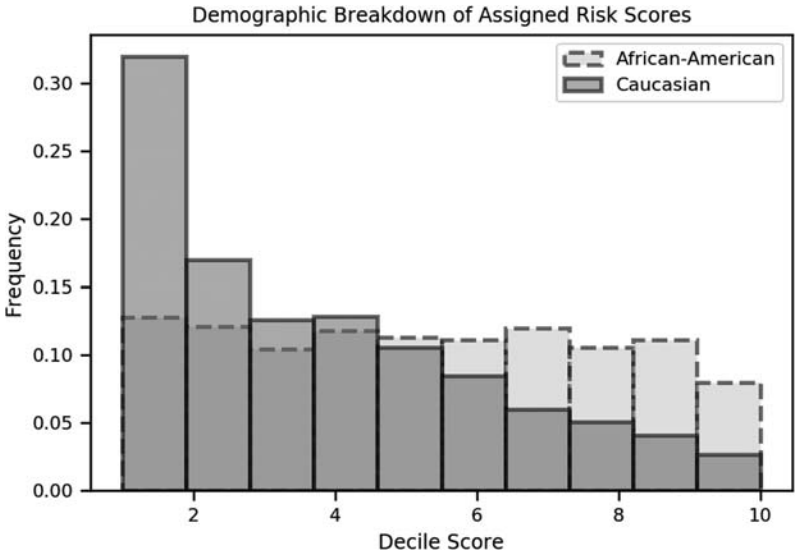
[116] 476 U.S. 79 (1986).

[117] *Id.* at 94.

[118] *Id.*; *see also* State v. Loomis, 881 N.W.2d 749, 775 (Wis. 2016) ("Attorney General Holder warned that using 'static factors and immutable characteristics, like the defendant's education level, socioeconomic background or neighborhood' in sentencing could have unintended consequences, including undermining our goal of 'individualized justice, with charges, convictions, and sentences befitting the conduct of each defendant and the particular crime he or she commits.'").

racy, 67.0% and 63.8% respectively.[119] Using a measure of accuracy known as the Area Under the Curve of the Receiver Operating Characteristic curve, Equivant has also found that the difference in outputs of its General Recidivism Risk Scale was not statistically significant between white and Black defendants, at 0.693 and 0.704 respectively.[120]

In recent years, numerous measures of fairness have been presented within the computer science and statistics literature.[121] Two of the most prominent means of measuring fairness are demographic parity and equality of odds. Demographic parity strives for equality in the distribution of model outputs—in other words, for risk-assessment algorithms, the predicted risk scores—across different racial groups. To achieve fairness under this measure, approximately the same percentage of white and Black defendants should have very low risk scores, approximately the same percentage should have very high scores, and so on. However, as illustrated in Figure A, the output distributions of COMPAS scores for Black and white defendants are markedly different.



[119] *See* Dressel & Farid, *supra* note 28, at 4.

[120] William Dieterich et al., Northpointe Inc. Research Department, Compas Risk Scales: Demonstrating Accuracy Equity And Predictive Parity 15 (2016). It is interesting to note just how inaccurate these models actually are. In fact, one study found that COMPAS was no more accurate or fair than the aggregate predictions of twenty individuals with little to no criminal justice expertise who responded to an online survey. *See id.* at 4. We will return to the topic of accuracy in Part VI, where we discuss potential methods to improve the algorithmic models that underlie COMPAS and other risk-assessment tools.

[121] Arvind Narayanan, *Tutorial: 21 Fairness Definitions and Their Politics*, YouTube (Feb. 23, 2018), https://www.youtube.com/watch?v=JIXIuYdnyyk, *archived at* https://perma.cc/5Z7D-TVH6.

Figure A: COMPAS Risk Score distributions among white and Black defendants. The relative frequency of white defendants who receive lower (less risky) decile scores from COMPAS is higher than the frequency of Black defendants who do, while the frequency of Black defendants who receive higher ("riskier") decile scores from COMPAS is higher than the frequency of white defendants who do. Data from ProPublica.[122]

COMPAS predictions are often stratified along a "High/Low" continuum: on a scale between one and ten, any score above five is considered "high" while any score below is considered "low." In this context of a binary predictor, demographic parity means that the same percentage of defendants of each race should receive low scores, and the same should be true for high scores. Table A highlights the disparity between Black and white defendants along these lines:

Table A: White defendants receive a larger share of low scores than do Black defendants.

	% Low Scores	% High Scores
White Defendants	66.9	33.1
Black Defendants	42.4	57.6

A second common measure of statistical fairness—equality of odds—compares the false positive rate ("FPR") and false negative rate ("FNR") of each racial group. In contrast to demographic parity, equity of odds considers true recidivism outcomes. FPR refers to the percentage of defendants who *did not* recidivate that were marked as high-risk, whereas FNR considers defendants who *did* recidivate despite being labeled as low-risk. Table B illustrates that white defendants are much more likely to be falsely labeled low-risk, whereas Black defendants are more likely to be falsely labeled high-risk:

[122] Jeff Larson et al., *How We Analyzed the COMPAS Recidivism Algorithm*, ProPublica (May 23, 2016), https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm, *archived at* https://perma.cc/R6GD-WPBX.

Table B: The false positive and false negative rates are not balanced between Black and white defendants.

White Defendants

	% Low Scores	% High Scores
Recidivated	49.6 (FNR)	50.4
Did Not	78.0	22.0 (FPR)

Black Defendants

	% Low Scores	% High Scores
Recidivated	28.5 (FNR)	71.5
Did Not	57.7	42.3 (FPR)

These disparities parallel findings of previous work.[123] Both from the perspectives of demographic parity and equality of odds, the algorithmic model used in COMPAS systematically overestimates risk for Black defendants.[124]

A. Algorithms and Input Features

Although race need not be explicitly included as an input to decision rules, recent literature suggests that input features incorporated in risk-assessment analysis coalesce to create proxies for race.[125] Inputs often correlate with race, and machine-learning models can make use of these features to effectively predict race. Therefore, a model may proxy for race even if it does not explicitly consider race as an input.

COMPAS poses more than 130 questions to defendants on topics ranging from the defendant's criminal history to subjective topics including personality traits.[126] Criminal history alone (as measured by prior arrests) varies among racial groups,[127] such that this one feature provides information about

---

[123] *See, e.g.*, Dressel & Farid, *supra* note 28, at 1.

[124] *See* Angwin et al., *supra* note 6.

[125] *See* Skeem & Lowenkemp, *supra* note 114, at 680–81 ("Although race is omitted from these instruments, critics assert that risk factors that are sometimes included (e.g., marital history and employment status) are 'proxies' for minority race and poverty.") (citations omitted).

[126] Questions from the COMPAS include:

(100) Do you often become bored with your usual activities?

(117) I'm really good at talking my way out of problems.

(125) I have never intensely disliked anyone.

See Northpointe, *supra* note 23.

[127] *See, e.g.*, The Sentencing Project, Report to the United Nations on Racial Disparities in the U.S. Criminal Justice System (2018), https://www.sentencingproject.org/publications/

the race of a particular defendant in a majority of cases. Other features, such as socio-economic status and the individual's personal perspectives, may also correlate with race. In high dimensions (that is, when a large number of features are considered), the inputs of individuals of different races may potentially be more easily aggregated to form a proxy for race.[128] These proxies for race appear even in simple models.[129]

These kinds of effects support the possibility that proxies for race operate in a discriminatory fashion within algorithms like COMPAS. To be sure, such proxies may sometimes not meaningfully affect an algorithm's final outputs, in which case the existence of these proxies may be legally insignificant. But in other cases, racial proxies arising in risk-assessment algorithms may result in harmful disparate impact.

IV. Equal Protection

The increasing popularity of risk-assessment instruments in the criminal justice system raises serious concerns regarding the equal protection rights of criminal defendants,[130] particularly racial minorities.[131] Current equal protection jurisprudence is ill-equipped to address the discrimination brought about by risk-assessment technology. The Supreme Court's equal protection decisions in *Washington v. Davis*[132] and *McCleskey v. Kemp*[133] appear to foreclose the argument that the use of risk-assessment technology may violate the Equal Protection Clause, as the doctrine stands today. The Court should reassess the law in this context to better equip courts to account for this technology. Specifically, we propose that the increasing use of risk-as-

---

un-report-on-racial-disparities/, *archived at* https://perma.cc/BJ4R-MACC ("African Americans are more likely than white Americans to be arrested; once arrested, they are more likely to be convicted; and once convicted, and they are more likely to experience lengthy prison sentences. African-American adults are 5.9 times as likely to be incarcerated than whites and Hispanics are 3.1 times as likely. As of 2001, one of every three Black boys born in that year could expect to go to prison in his lifetime, as could one of every six Latinos—compared to one of every seventeen white boys.") (internal citations omitted).

[128] The Authors do not yet take a position on whether this is a bad thing in creating a fair model. Recent studies have suggested that actually creating individual models for each class is the best way to create an accurate model. However, this raises thorny ethical issues and is beyond the scope of this Article.

[129] *See* Moritz Hard, Eric Price, & Nathan Srebro, *Equality of Opportunity in Supervised Learning* 16–19, *in* Proceedings of the 30th Conference on Neural Information Processing Systems (2016) (conference paper), https://arxiv.org/abs/1610.02413, *archived at* https://perma.cc/UD7C-UC4G.

[130] Many of the arguments made in this article are relevant to gender classifications as well. Due to the substantial differences between equal protection doctrine as it relates to race and gender, *see, e.g.*, United States v. Virginia, 518 U.S. 515, 531–32 (1996), we restrict our examination in this Article to the impact of risk-assessment tools on racial minorities.

[131] *See* Bernard E. Harcourt, *Risk as a Proxy for Race* 4 (John M. Olin Program in Law and Econ., Working Paper No. 535, 2010), https://chicagounbound.uchicago.edu/cgi/viewcontent.cgi?article=1432&context=law_and_economics, *archived at* https://perma.cc/2U8B-KA4C.

[132] 426 U.S. 229, 270 (1976).

[133] 481 U.S. 279, 319 (1987).

sessment technology demands that the Court reexamine the *Davis-McCleskey* framework and adopt a burden-shifting model resembling that which the Court has long used in the context of jury selection.[134] We argue that a model of discrimination based on the reasoning in *Batson v. Kentucky*[135] is a more appropriate lens through which courts can understand and address these issues.

A. The Davis-McCleskey *Framework*

In this section, we analyze how courts have traditionally understood discriminatory intent, and then argue that "algorithmic intent" cannot be adequately addressed by prevailing equal protection doctrine.

The Supreme Court first addressed equal protection in the context of criminal procedure in *Strauder v. West Virginia*.[136] In *Strauder*, a Black petitioner, Taylor Strauder, was convicted of murder by an all-white jury. At the Supreme Court, Strauder prevailed on the theory that West Virginia's exclusion of Black jurors violated the Equal Protection Clause.[137] Nearly a hundred years after *Strauder*, the Court's decision in *Washington v. Davis* held that a showing of discriminatory impact, standing alone, is insufficient to trigger strict scrutiny: discriminatory intent is also required.[138] This has made proving equal protection violations remarkably difficult.

Claims brought by criminal defendants under the Equal Protection Clause tend to be fact-intensive inquiries.[139] Courts scrutinize a variety of evidence, including the facial neutrality (or non-neutrality) of the law, the discriminatory impact of the law's administration, and case-specific evidence presented by the claimant. Today, equal protection claims are governed by the Supreme Court's landmark decisions in *Washington v. Davis* and *McCleskey v. Kemp*. Together, *Davis* and *McCleskey* mandate a dual showing of both discriminatory *effect* and *intent* for a petitioner to prevail on equal protection grounds.[140] The intent standard was perhaps most famously articulated by Justice Stewart, who noted that a law must have been enacted

---

[134] *See* Eubanks v. Louisiana, 356 U.S. 584, 587 (1958); *see also* Alexander v. Louisiana, 405 U.S. 625, 631–32 (1972) ("Once a prima facie case of invidious discrimination is established, the burden of proof shifts to the State to rebut the presumption of unconstitutional action by showing that permissible racially neutral selection criteria and procedures have produced the monochromatic result.").

[135] 476 U.S. 79 (1986).

[136] 100 U.S. 303, 306 (1879). *See generally* Michael J. Klarman, *The Racial Origins of Modern Criminal Procedure*, 99 MICH. L. REV. 48 (2000).

[137] *See Strauder*, 100 U.S. at 311.

[138] *Id.*

[139] *See, e.g.*, Yick Wo v. Hopkins, 118 U.S. 356, 374 (1886); Village of Arlington Heights v. Metro. Hous. Dev. Corp., 429 U.S. 252, 266 (1977).

[140] *See* Washington v. Davis, 426 U.S. 229, 240–41 (1976); McCleskey v. Kemp, 481 U.S. 279, 298 (1987); *see also* Reva B. Siegel, *Blind Justice: Why the Court Refused to Accept Statistical Evidence of Discriminatory Purpose in* McCleskey v. Kemp*—And Some Pathways for Change*, 112 NW. U.L. REV. 1269, 1271 (2018), https://ssrn.com/abstract=3181190, *archived at* https://perma.cc/W3LW-RVLD.

"because of, not merely in spite of" its discriminatory impact in order to be invalidated on equal protection grounds.[141] This intent standard controls most equal protection claims today,[142] and presents a burden that is nearly impossible for criminal defendants to carry.

Whether the "intent" framework of equal protection doctrine could ever be applied to algorithms (absent a showing of a malicious designer) is an open question.[143] Traditionally, an "actor is presumed to have intended the natural consequences of his deeds."[144] Yet this conception of intent does not hold true in the case of algorithmic decisionmaking, as algorithms do not have autonomous decisionmaking capabilities. The intent standard thus proves ill-equipped to respond to discriminatory risk-assessment technology.

As previously discussed, there are two fundamental problems with proprietary machine-learning models as they are used today: the inability to access model parameters and a lack of explanations for results. In many cases, the only information available is predictive outputs. Though some studies indicate a difference among the score distributions of different races,[145] without model parameters, one cannot assess whether there is a specific relationship in the inputs that is likely to produce discriminatory scores based on an individual defendant's race. Without access to the training data and procedures, one cannot determine if an algorithm is programmed to intentionally discriminate on the basis of race.

Even if the model parameters were available, explaining the output of the model would remain difficult. Whereas judges explain the combination of factors leading to a sentencing or bail decision, the only explanation for predictive outputs would be the internal weights of the model. Often, there can be hundreds or thousands of parameters that specify how the input features interact. In complex models, these values can approximate arbitrary interactions between the parameters and make it difficult to pinpoint what exactly contributes to a high recidivism score.

---

[141] Personnel Adm'r of Massachusetts v. Feeney, 442 U.S. 256, 279 (1979).

[142] To prevail on equal protection grounds, the Supreme Court demands a showing of invidious intent in almost all contexts. *See, e.g.*, Wright v. Rockefeller, 376 U.S. 52, 60 (1964) (gerrymandering); Keyes v. Sch. Dist., 413 U.S. 189, 205 (1973) (school desegregation); Jefferson v. Hackney, 406 U.S. 535, 548 (1972) (Social Security Act).

[143] *See, e.g.*, Aziz Z. Huq, *Racial Equity in Algorithmic Criminal Justice*, 68 DUKE L.J. 1043, 1134 (2019), https://ssrn.com/abstract=3144831, *archived at* https://perma.cc/QW8P-KELC.

[144] *Davis*, 426 U.S. at 253 (Stevens, J., concurring).

[145] Angwin et al., *supra* note 6. *But see generally* Anthony W. Flores, Kristin Bechtel, & Christopher T. Lowenkamp, *False Positives, False Negatives, and False Analyses: A Rejoinder to "Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And It's Biased against Blacks"*, 80 FED. PROBATION 38 (2016) (arguing that ProPublica's report was based on faulty statistics and data analysis and failed to show that the COMPAS is racially biased).

B. Malicious Designer

One scenario that the current intent-based framework would readily address is that of a malicious designer who intentionally creates an algorithm to target certain racial groups. The malicious designer could effectuate that intent simply by adding a certain value to the output risk scores of defendants of a particular race, which would increase the risk scores for individuals of that race in accordance with the imputed bias. This method would be easily detectable in analysis of the algorithm, but more complex methods could also be used. Data from particular races could be selectively emphasized in training, input features from particular races could be altered, or specific weights could be set to discriminate against specific racial groups. Due to the complexity of these approaches and of the machine-learning models themselves, determining the existence of these forms of bias from output predictions would be difficult. More likely, a determination of bias would require finding a "smoking gun" that proves that the designers took these steps in a malicious, intentional fashion in creating the model. This is unlikely to happen.

Of course, discriminatory impact is possible without a malicious design. During training, algorithms may learn to pick up on select features, such as race, simply based on the available data inputs or the applied statistical procedures.

Consider an example. Imagine two groups, A and B, that differ in terms of a protected attribute such as race. Suppose members of group A recidivate at a rate of seventy percent, while those in group B recidivate at a rate of thirty percent. With no additional information, a machine-learning model might use group identity as a proxy for recidivism. Given that race is a protected class, use of such modeling is suspect.

Though the above example illustrates a direct incorporation of race in a dataset, in reality, the influence of race may be much subtler due to data imbalance and systematic biases in measurement. Data imbalances arise when the training data do not have a proportionate number of cases from each racial group. For example, facial-recognition software that is trained using a dataset comprising mostly white faces may have trouble recognizing the faces of members of other racial groups.[146] In these cases, the models may only learn to fit the patterns for the majority class, leading to high accuracy for that group, but an inability to capture the nuances and differences among the tested racial groups. Features that correspond to recidivism

[146] *See, e.g.*, Ali Breland, *How White Engineers Built Racist Code—and Why it's Dangerous for Black People*, GUARDIAN (Dec. 4, 2017), https://www.theguardian.com/technology/2017/dec/04/racist-facial-recognition-white-coders-black-people-police, *archived at* https://perma.cc/63NW-PEFF.

for the majority racial group may not be the same as those for a minority group, leading to disparate outcomes when the algorithm is used.[147]

Systematic biases in inputs may also impact results. Consider the following questions from the COMPAS questionnaire: "Based on the screener's observation, is this person a suspected or admitted gang member?"; and "How often did you feel you have nothing to do in your spare time?" Questions such as these open the survey to error. The first relies on the screener's intuition and could extend the screener's bias. The second may lead to culturally specific responses.[148] And questions related to criminal history, such as prior arrests and convictions, can extend biases from biased enforcement.

All of these biases in data can impact the fairness of machine-learning models. In cases involving systematic error, these problems cannot be avoided by simply making "better" statistical models in the sense of using more complex mathematical tools. These models will extend the biases from the *data* and lead to discriminatory predictions along racial lines.

Consider again the example of the two groups, A and B, where A recidivates at a rate of seventy percent and B recidivates at a rate of thirty percent. Each feature imputed into the model may highlight and exacerbate the distributional differences between racial groups. That may impact even simple models. For example, if income correlates to recidivism, a disparate impact may result from entrenched wealth-related differences across racial groups. Inclusion of other similar variables in the model may interact to coincide with race.

In a sufficiently complex model, the aggregate of these features could act as a "proxy" for race in the model. With more than 100 questions as inputs, the COMPAS risk-needs assessment is much more likely to compound these variables into a proxy for race than it would be if it used only a handful of inputs. COMPAS uses cultural and socio-economic factors that often have strong ties to race or other protected characteristics.[149] Without access to the data or model details, it is difficult to determine the exact impact on predictive outputs. The data suggest that there is some type of discrepancy between the predictive outputs of defendants based on race. But it is impossible to determine the exact cause of the disparate effect, whether it be unbalanced training data, systematic biases, or something else altogether.

---

[147] Dwork et al., *supra* note 46, at 11–15, 20.

[148] For example, a person's overall happiness or life satisfaction is often measured in surveys using a 10-point scale, with those claiming to be the happiest giving a score of 9 or 10. *See* Gaël Brulé & Ruut Veenhoven, *The '10 Excess' Phenomenon in Responses to Survey Questions on Happiness*, 131 Soc. Indic. Res. 853, 855 (2016). Within this group of the happiest respondents, there is evidence that those from certain countries, particularly many Latin American or Middle Eastern countries, are more likely to report a score of 10 than a score of 9. *Id.* at 857–59. Among other factors, cultural links between happiness, prestige, and social acceptance, as well as cultural norms around grading or tendencies to select "extreme" survey responses may play a role in whether a score of 9 or 10 is more frequent in different countries. *Id.* at 866–68.

[149] *See e.g.*, David R. Williams et al., *Understanding Associations Among Race, Socioeconomic Status, and Health: Patterns and Prospects*, 35 Health Psychol. 407 (2016).

If professional statisticians and legal professionals cannot adequately discern algorithmic intent to this end, it is hard to see how a criminal defendant might use it successfully to argue along traditional equal protection lines, as these algorithmic models cannot be said to discriminate intentionally as needed to establish an equal protection violation.

Without an explicit explanation from the algorithm's designers or some opportunity to examine the model's formulation, there is unlikely to be any "smoking gun" evidence that the COMPAS model treats members of different races in an intentionally discriminatory fashion. And yet we know such disparate impact occurs because there is a clear disparity in how the model handles individuals from different races. These differences demand inquiry into how the model was crafted and trained. Absent such analysis, algorithmic intent cannot serve as a surrogate of discriminatory intent in an equal protection argument.

C. Intent By Continued Use

Plainly, the algorithmic and "malicious designer" understandings of discriminatory intent are unsuited to the *Davis-McCleskey* equal protection doctrine. But there is one final possible conception of discriminatory intent: intent by continued use. If a judge believed that risk-assessment technology discriminated against defendants of a particular race, the assessment could be analyzed as follows:

(1) Risk-assessment technology discriminates on the basis of race;
(2) Continued use of such technology will perpetuate discrimination;
(3) Perpetuation of discrimination is de facto endorsement of it;
(4) De facto endorsement of discrimination amounts to intentional discrimination;
(5) Intentional discrimination offends the Equal Protection Clause;
(6) Therefore, the judge must strike down the use of risk-assessment technology or else offend equal protection.

Courts have previously been willing to endorse this view. In 2010, Congress passed the Fair Sentencing Act ("FSA") to "restore fairness to federal cocaine sentencing."[150] Prior to the passage of the FSA, federal law more harshly penalized crack cocaine offenses than similar offenses involving powder cocaine, resulting in systematically higher sentences for minority drug offenders. In passing the FSA, Congress sought to correct the sentencing regime's discriminatory impact.[151]

---

[150] United States v. Blewett, 719 F.3d 482, 484 (6th Cir. 2013), *rev'd en banc*.

[151] *See, e.g.*, Andrew Cockroft, *Congress Blewett by Not Explicitly Making the Fair Sentencing Act of 2010 Retroactive*, 107 J. Crim. L. & Criminology 325, 326 (2017) ("Congress passed the [FSA] in an attempt to remedy the discriminatory effects of the Anti-Drug Abuse Act of 1986.").

One issue unresolved by the FSA's passage was whether it applied retroactively to those who had previously been sentenced under the Anti-Drug Abuse Act of 1986, which had been the law until 2010. Not long after the FSA's passage, this question reached the Sixth Circuit in *United States v. Blewett*.[152] Reasoning that a failure to read the FSA retroactively would perpetuate the preexisting discriminatory system, a Sixth Circuit panel ruled that the Equal Protection Clause demanded that defendants be resentenced. The decision was quickly reversed en banc seven months later.[153]

Had it been allowed to stand, the *Blewett* panel decision could have had remarkable implications: conceivably, thousands of drug offenders would have been resentenced under the new regime. But the panel's reasoning—that "continued use" of the preexisting sentences amounted to state-sanctioned discrimination—had been rejected by the Supreme Court decades earlier. The Supreme Court had said that a state violates the Equal Protection Clause when it chooses a course of action *because of* its racially discriminatory effects, not simply when discrimination is inadvertently permitted to continue.[154] In other words, the Court had said that an equal protection claim demands proof of intent that is more than awareness of consequences. Thus, the framework of intent by continued use, like the algorithmic intent and malicious designer frameworks, proves unhelpful to a claimant seeking to vindicate the equal protection right.

D. A Batson*-like Model of Equal Protection Analysis*

In the absence of a showing of discriminatory intent, proving an equal protection violation under the *Davis-McCleskey* framework is impossible. Algorithmic intent is such an uncomfortable fit with the traditional equal protection doctrine that it raises the question of the doctrine's obsolescence in the face of developing technology. An alternative constitutional approach to discrimination here is needed. As we argue here, one exists within the Supreme Court's equal-protection analysis.

The Supreme Court has reevaluated equal protection doctrine in the past to address unreasonable legal hurdles. In 1981, an all-white jury convicted a Black man, James Kirkland Batson, of burglary. A local prosecutor in Louisville, Kentucky who was assigned to Batson's case used the state's peremptory challenges to remove all four Black veniremen from the jury pool. Batson moved to discharge the jury, arguing that its all-white composition deprived him of his Sixth Amendment right to a trial by a jury of one's peers and his Fourteenth Amendment right to equal protection. The trial judge dismissed the motion, and Batson was convicted.

---

[152] *Blewett*, 719 F.3d at 487.
[153] United States v. Blewett, 746 F.3d 647, 649 (6th Cir. 2013) (en banc).
[154] *See, e.g.*, Personnel Adm'r of Massachusetts v. Feeney, 442 U.S. 256, 279 (1979).

In affirming Batson's conviction, the Kentucky Supreme Court concluded, under the law as it stood at that time, that Batson had failed to meet his burden of proof to establish an equal protection violation. The then-prevailing standard, established in *Swain v. Alabama*,[155] placed an enormous burden on the criminal defendant. As the Supreme Court of the United States summarized in *Batson*:

> [Under *Swain*], a black defendant could make out a prima facie case of purposeful discrimination on proof that the peremptory challenge system was "being perverted" in that manner . . . . For example, an inference of purposeful discrimination would be raised on evidence that a prosecutor, "in case after case, whatever the circumstances, whatever the crime and whoever the defendant or the victim may be, is responsible for the removal of [Black jurors] who have been selected as qualified jurors by the jury commissioners and who have survived challenges for cause, with the result that no [Black jurors] ever serve on petit juries." . . . While the defendant showed that prosecutors in the jurisdiction had exercised their strikes to exclude blacks from the jury, he offered no proof of the circumstances under which prosecutors were responsible for striking black jurors beyond the facts of his own case.[156]

Swain demanded that a defendant show *systemic* racial discrimination in "case after case"—not merely in the defendant's own individual case. Naturally, Batson was unable to prove systemic discrimination in the jury-selection procedures of the county as a whole.

As the Supreme Court recognized, the standard set forth in *Swain* unduly burdened defendants seeking relief for violations of their equal protection rights. Such a burden was, according to the Court, so "crippling" that "peremptory challenges [became] largely immune from constitutional scrutiny."[157] In response, the Court adjusted and tempered the standard by which criminal defendants can prove equal protection violations in jury selection, and in so doing created the now famous "*Batson* challenge," setting the stage for the invalidation of jury selection procedures around the country.[158]

Functionally, *Batson* created a burden-shifting framework that is triggered by a finding of disparate impact. Under *Batson*, a defendant can advance an equal protection claim by drawing a connection between the irregular racial composition of a jury and discrimination in the jury selection process. Thus, under *Batson*, a defendant need not show racial animus or

---

[155] 380 U.S. 202, 208–09 (1965).

[156] Batson v. Kentucky, 476 U.S. 79, 91–92 (1986) (internal citations and quotations omitted).

[157] *Id.* at 92–93.

[158] *See, e.g.*, Flowers v. Mississippi, 139 S. Ct. 2228 (2019).

discriminatory intent in order to establish a prima facie case of an equal protection violation. Instead, a defendant need only demonstrate membership in a targeted racial class and present evidence that discrimination *may* have occurred. By doing so, the defendant raises an inference of intentional discrimination. Upon this showing, the burden of proof shifts to the state to provide a neutral explanation for the alleged discrimination. If the state can successfully carry this burden to the satisfaction of the judge, the defendant's equal protection claim will fail. However, if the state cannot do so, the defendant's claim will prevail.

E. Risk-Assessment Instruments & Burden-Shifting Analysis

With respect to algorithmic intent, the Court should adopt a model for analyzing equal protection violations similar to that used in *Batson v. Kentucky*. Adopting a *Batson*-like model, defendants subjected to discriminatory risk-assessment technology should be allowed to establish a prima facie case of an equal protection violation if they can posit facts that permit an inference of intentional discrimination. This could be accomplished in two steps. First, the defendant could introduce the questionnaire associated with the defendant's risk-assessment score containing questions that correlate with race. The defendant could also present data showing a systematic overestimation of recidivism risk in minority defendants. Importantly, this data would not operate in the same way that the Baldus study did in *McCleskey*. The social-science data in *McCleskey* was offered to *prove* the equal protection violation and was rejected as such. Here, the defendant would offer the data to establish an *inference* of racial discrimination, shifting the burden of proof onto the state in much the same way that a lack of Black jurors on a jury did in *Batson*.

Thus, the presumption of racial discrimination would shift the burden to the state to offer a race-neutral explanation. Per the Supreme Court's own doctrine, once a prima facie case of invidious discrimination is established, the state bears the burden of showing that "permissible racially neutral selection criteria and procedures have produced" the result caused by the COMPAS system, and/or other similarly designed algorithms.[159] As in *Batson*, if the state is unable to offer a racially neutral explanation, or if the court deems the explanation inadequate, the defendant should prevail on an equal protection claim.

Two questions arise. First, why would the Court accept a defendant's proffered statistical evidence of racial discrimination in risk-assessment technology, considering that it rejected the statistical evidence in *McCleskey*? Critical to the equal protection argument in *McCleskey* was the presentation of empirical evidence—the Baldus study—which suggested systemic

[159] Alexander v. Louisiana, 405 U.S. 625, 632 (1972).

racial bias in Georgia juries' imposition of the death penalty.[160] The study analyzed over 2,500 murder cases and concluded, after accounting for dozens of nonracial variables, that criminal defendants charged with murdering white victims were substantially more likely to receive a capital sentence than those charged with murdering Black victims.[161] The *McCleskey* Court rejected the Baldus study because it believed that statistical analysis was an inappropriate means of establishing bias in capital punishment decisions due to the "nature" of those decisions:

> But the *nature* of the capital sentencing decision, and the relationship of the statistics to that decision, are fundamentally different from the corresponding elements in the venire-selection or Title VII cases. Most importantly, each particular decision to impose the death penalty is made by a petit jury selected from a properly constituted venire. Each jury is unique in its composition, and the Constitution requires that its decision rest on consideration of innumerable factors that vary according to the characteristics of the individual defendant and the facts of the particular capital offense . . . . Thus, *the application of an inference drawn* from the general statistics to a specific decision in a trial and sentencing simply is not comparable to the application of an inference drawn from general statistics to a specific venire-selection or Title VII case. In those cases, the statistics relate to fewer entities, and fewer variables are relevant to the challenged decisions.[162]

In other words, the Court rejected the Baldus study due to the nature of the state's *sentencing regime*, not the nature of the *statistical analysis*. That the state could not rebut the statistical conclusions drawn from the Baldus study was of critical importance to the *McCleskey* Court.[163] But unlike the state of Georgia, jurisdictions using proprietary risk-assessment technology would have both the opportunity and the ability to challenge the inferences drawn from the statistical evidence by providing transparency in the underlying algorithmic calculations. Importantly, the Supreme Court does not bar the use of statistical evidence in equal protection claims across the board.[164] This, combined with the general impossibility of demonstrating invidious

---

[160] *See* David C. Baldus et al., *Comparative Review of Death Sentences: An Empirical Study of the Georgia Experience*, 74 J. Crim. L. & Criminology 661, 661–753.

[161] *See id.*

[162] *McCleskey*, 481 U.S. at 294–95 (emphases added).

[163] *See id.* at 296 ("Another important difference between the cases in which we have accepted statistics as proof of discriminatory intent and this case is that, in the venire-selection and Title VII contexts, the decisionmaker has an opportunity to explain the statistical disparity . . . . Here, the State has no practical opportunity to rebut the Baldus study." (internal citations omitted)).

[164] Batson v. Kentucky, 476 U.S. 79, 93 (1986) ("We have observed that under some circumstances proof of discriminatory impact may for all practical purposes demonstrate unconstitutionality because in various circumstances the discrimination is very difficult to explain on nonracial grounds . . . . For example, total or seriously disproportionate exclusion of

discriminatory algorithmic intent, is a sound reason to consider the *Batson* model in this context.

Second, what should we make of *Batson*'s emphasis on the uniqueness of the jury and on the relationship between the guarantees of the Equal Protection Clause and the Sixth Amendment? Are jury selection procedures unique? Yes and no. In *Batson*, the Court focused on the Sixth Amendment right to a trial by jury. Interestingly, the Court articulated the relationship between equal protection and the right to trial by jury in concluding that the former protected the latter: "[p]urposeful racial discrimination in selection of the venire violates a defendant's right to equal protection because it denies him the protection that a trial by jury is intended to secure[.]"[165] In other words, the *Batson* Court read the Sixth and Fourteenth Amendments as operating in tandem to guarantee a fair jury selection procedure.

We have already suggested that the use of the COMPAS system offends due process.[166] Why should equal protection with respect to risk assessment instruments not operate in tandem with due process, as it does with respect to the right to a jury trial? The Equal Protection Clause should be construed to safeguard due process rights in the context of algorithmic decisions, just as it did the Sixth Amendment jury-trial right in *Batson*.

In order for courts to address the discriminatory harms caused by risk-assessment instruments, the *Batson* model proves more compatible than the *Davis-McCleskey* framework, because machine bias is not susceptible to a traditional intent-based inquiry.[167] As the criminal justice system increasingly incorporates artificial intelligence, continued use of an anachronistic discriminatory intent rule will prevent defendants from being able to seek meaningful relief from discriminatory harms. Consistent with the Court's own thinking on equal protection and discrimination in criminal justice, an adjusted doctrinal framework for addressing discriminatory algorithms should be adopted. The use of discriminatory risk-assessment technology undermines faith in the fairness of the judicial process.[168] So long as courts do not address this problem, equal protection in criminal justice may prove to be "but a vain and illusory requirement."[169]

[Black jurors] from jury venires . . . is itself such an unequal application of the law . . . as to show intentional discrimination . . . (internal quotations and citations omitted)).

[165] *Id.* at 86 (internal quotations omitted).

[166] *See supra* Part II.

[167] We are not the first to suggest breaking *Batson* free from its voir dire silo and applying the reasoning to other areas of law. For example, David Baldus and his co-authors suggested that *Batson* and its progeny share much in common with the disparate treatment and pattern-and-practice models of Title VII litigation. *See* David C. Baldus et al., *Statistical Proof of Racial Discrimination in the Use of Peremptory Challenges: The Impact and Promise of the* Miller-El *Line of Cases As Reflected in the Experience of One Philadelphia Capital Case*, 97 IOWA L. REV. 1425, 1430 (2012).

[168] *See Batson*, 476 U.S. at 87; *see also* Ballard v. United States, 329 U.S. 187, 195 (1946); McCray v. New York, 461 U.S. 961, 968 (1983) (Marshall, J., dissenting from denial of certiorari).

[169] Norris v. Alabama, 294 U.S. 587, 598 (1935).

V. State Constitutional Claims

The vast majority of criminal adjudication takes place in state courts.[170] States such as Florida, New Jersey, and Wisconsin already use risk-assessment instruments in various stages of their criminal justice systems, and the debate regarding the constitutionality of those technologies is under way.[171] Recognizing that legal challenges at the state level may well pave the way for challenges in the federal courts, this section discusses how state courts might grapple with the constitutional issues surrounding the use of risk-assessment technology and identifies potential state-level legal challenges that may be most fruitful.

A. State Constitutions

State constitutions provide viable means for practitioners to challenge risk-assessment instruments. Notwithstanding the Supremacy Clause,[172] a state may "adopt in its own Constitution individual liberties more expansive than those conferred by the Federal Constitution."[173] States typically do this, not only by adopting express constitutional provisions protecting rights not recognized by the U.S. Constitution, but also by interpreting provisions of their state constitutions more broadly than federal courts do the U.S. Constitution.[174] Specifically, many state constitutions have their own due process[175] and equal protection[176] clauses. Even if an equal protection, due process, or other challenge is foreclosed at the federal level, it may be viable under state

---

[170] *Compare* United States Courts, Table JCI–U.S. Federal Courts Federal Judicial Caseload Statistics (Mar. 31, 2010) (showing 77,287 federal criminal cases filed from April 1, 2009 to March 31, 2010), https://www.uscourts.gov/statistics/table/jci/federal-judicial-caseload-statistics/2010/03/31, *archived at* https://perma.cc/YC9K-QCZX, *with* R. LaFountain et al., Nat'l Ctr. for State Courts, Examining the Work of State Courts: An Analysis of 2010 State Court Caseloads 3 (2012) (reporting 20.4 million incoming non-traffic criminal cases in state courts in 2010), http://www.courtstatistics.org/~/media/Microsites/Files/CSP/DATA%20PDF/CSP_DEC.ashx, *archived at* https://perma.cc/64WF-C7VG.

[171] *See supra* notes 6–7, 15–17.

[172] U.S. Const., art. VI, cl. 2.

[173] PruneYard Shopping Ctr. v. Robins, 447 U.S. 74, 81 (1980) (citing Cooper v. Cal., 386 U.S. 58, 62 (1967)).

[174] *See* Randy J. Holland et al., State Constitutional Law: The Modern Experience 2–5 (2010).

[175] *See, e.g.*, N.M. Const., art. II, § 18; Va. Const., art. I, § 11; Wis. Const., art. I, § 8.

[176] *See, e.g.*, Cal. Const., art. I, §§ 7–8; Fla. Const., art. I, § 2; Tex. Const., art. I, § 3a.

law.[177] In fact, developments in state constitutional law often influence later changes at the federal level.[178]

Many states construe their constitutions to achieve conformity with the federal Constitution. However, other states take a more independent approach. The opportunity for adopting a more expansive interpretation of equal protection or due process increases in cases where a state constitutional provision is worded differently from the federal analogue[179] or has a distinct legislative history.[180] Some state constitutions even contain express provisions that unmoor them from the federal Constitution.[181] Along these lines, several states that use risk-assessment tools have already recognized more expansive individual rights under their own constitutions, opening the door to challenging the use of risk-assessment algorithms in these states.

In Florida, where many counties use COMPAS for pretrial assessment,[182] the state constitution contains provisions that parallel the Due Process and Equal Protection Clauses of the federal Constitution, but with broader scope.[183] Though the language of Florida's due process clause is al-

---

[177] *See* William J. Brennan, Jr., *State Constitutions and the Protection of Individual Rights*, 90 HARV. L. REV. 489, 498–501 (1977) (giving numerous examples where a state constitution is more protective of individual rights than the federal constitution); William B. Rubenstein, *The Myth of Superiority*, 16 CONST. COMMENT. 599, 606–08 (1999) (discussing successful challenges to laws banning sodomy under state constitutions following a Supreme Court ruling that such laws were allowable under the federal constitution).

[178] *See, e.g.*, James A. Gardner, *State Constitutional Rights as Resistance to National Power: Toward a Functional Theory of State Constitutions*, 91 GEO. L. J. 1003, 1039 (2003) (noting the influence of state constitutional law cases on the Supreme Court's decision to incorporate the exclusionary rule against the states); *see also* Joseph Blocher, *What State Constitutional Law Can Tell Us About the Federal Constitution*, 115 PENN. ST. L. REV. 1035, 1036–37 (2011) ("State-level rights guarantees served as the model for many of the most familiar features of the Bill of Rights and of American constitutional law."). Conversely, the successful setting of a precedent recognizing greater protections for criminal defendants on state law grounds regarding the use of risk-assessment tools would be insulated from Supreme Court review and possible reversal. *Cf.* Dana Walsh, Note, *The Dangers of Eyewitness Identification: A Call for Greater State Involvement to Ensure Fundamental Fairness*, 54 B.C. L. REV. 1415, 1447 (2013).

[179] *See* HOLLAND ET AL., *supra* note 174, at 153.

[180] *See, e.g.*, State v. Hunt, 450 A.2d 952, 965 (N.J. 1982) (Pashman, J., concurring); *cf.* Jeffrey S. Sutton, *What Does—and Does Not—Ail State Constitutional Law*, 59 U. KAN. L. REV. 687, 711 (2011) (critiquing approaches premised only on authority to interpret state constitutions differently rather than distinct historical backgrounds of state constitutions).

[181] *See, e.g.*, CAL. CONST. art. I, § 24 ("Rights guaranteed by this Constitution are not dependent on those guaranteed by the United States Constitution."). On the other hand, some states have adopted express constitutional provisions with the opposite effect. *See, e.g.*, FLA. CONST. art. I, §§ 12, 17 (mandating that protections against unreasonable searches and seizures and against excessive punishments be "construed in conformity with" the Supreme Court's interpretations of the Fourth and Eighth Amendments to the federal constitution, respectively).

[182] *See* Angwin et al., *supra* note 6.

[183] FLA. CONST. art. I, § 2 ("All natural persons, female and male alike, are equal before the law and have inalienable rights, among which are the right to enjoy and defend life and liberty, to pursue happiness, to be rewarded for industry, and to acquire, possess and protect property. No person shall be deprived of any right because of race, religion, national origin, or physical disability.").

most identical to the language of the federal Due Process Clause,[184] Florida's clause has been interpreted more broadly.[185]

Florida's constitution also contains auxiliary procedural protections that lack direct federal equivalents. In particular, Article 1, Section 14 of the Florida constitution grants criminal defendants the right to pretrial release on reasonable conditions in most cases.[186] Because the Florida Supreme Court often reads different state constitutional provisions liberally and in tandem,[187] it might entertain the argument that using proprietary or discriminatory algorithms burdens a defendant's presumptive right to pretrial release.[188] Florida and similarly situated states could therefore provide fertile soil in which to pursue constitutional challenges to risk-assessment instruments.

Compared to Florida, the state of Wisconsin, which uses COMPAS at sentencing,[189] has recognized a much narrower role for its constitution. Wisconsin's constitution has an equal protection clause[190] couched in the natural-rights language of the Declaration of Independence.[191] The Wisconsin Supreme Court has restricted the breadth of the provision to parallel the Equal Protection Clause of the Fourteenth Amendment.[192] Likewise, Wisconsin generally treats its due process clause[193] as "the substantial equivalent[ ]" of the federal Due Process Clause.[194]

However, state constitutional challenges may still be viable in Wisconsin and other states that have ostensibly limited the reach of their constitutions. In *State v. Dubose*,[195] a victim of an armed robbery identified the defendant during two showups[196]—a procedure in which a suspect is presented in isolation to a witness for identification.[197] The defendant was convicted after the trial judge denied his motion to suppress the showup

---

[184] FLA. CONST. art. I, § 9 ("No person shall be deprived of life, liberty or property without due process of law.").

[185] *See* Traylor v. State, 596 So. 2d 957, 961 (Fla. 1992). For example, Florida appellate courts have invoked their state constitution's due process clause in rejecting the federal case-by-case approach to determining whether an indigent parent has a right to counsel in cases potentially involving permanent deprivation of parental rights. *See* M.E.K. v. R.L.K., 921 So. 2d 787, 790 (Fla. Dist. Ct. App. 2006) (citing *Traylor*, 596 So. 2d at 961).

[186] FLA. CONST. art. I, § 14.

[187] *See, e.g.*, *Traylor*, 596 So. 2d at 966–70.

[188] Such arguments could draw inspiration from *Batson*, which similarly read the Sixth and Fourteenth Amendments in tandem. *See supra* Section IV.D.

[189] *See* State v. Loomis, 881 N.W.2d 749, 767 (Wis. 2016); *see also* WIS. DEP'T CORR., *supra* note 7.

[190] WIS. CONST. art. I, § 1.

[191] *See* Reginald D. v. State, 533 N.W.2d 181, 184–85 (Wis. 1995); WIS. CONST. art. I, § 1 ("All people are born equally free and independent, and have certain inherent rights; among these are life, liberty and the pursuit of happiness; to secure these rights, governments are instituted, deriving their just powers from the consent of the governed.").

[192] *See, e.g.*, *Reginald*, 533 N.W.2d at 184–85.

[193] WIS. CONST. art. I, § 8.

[194] McManus v. State, 447 N.W.2d 654, 660 (Wis. 1989).

[195] 699 N.W.2d 582 (Wis. 2005).

[196] *Id.* at 586.

[197] *Id.* at 584 n.1.

identification, and his conviction was upheld on appeal.[198] The Wisconsin Supreme Court reversed on due process grounds, holding that the showup identifications "were unnecessarily suggestive."[199] In so doing, it departed from the more permissive federal requirement to assess whether a procedure is unnecessarily suggestive before shifting the burden of proof onto the state to demonstrate how the showup was reliable under the totality of the circumstances.[200] The Wisconsin Supreme Court found the federal standard inadequate under the state's due process clause, holding instead that showups are inherently suggestive and thus only admissible in cases of necessity.[201] Importantly, the court's adoption of the more stringent procedural standard rested on a series of empirical studies regarding the unreliability of eyewitness identification.[202]

Although the Wisconsin Supreme Court has not extended the state's due process clause further,[203] it also has not called into question one of *Dubose*'s underlying principles: empirical studies disputing the reliability of evidentiary procedures may provide a "compelling justification" for recognizing greater protections under Wisconsin's constitution.[204] Thus, empirical challenges premised on COMPAS's low predictive value or racial bias may be viable under Wisconsin's due process clause. Such claims are markedly different from the federal claims rejected in *Loomis*, which centered on individualization in sentencing and the accuracy of inputs in Wisconsin's sentencing procedure.[205] As in *Dubose*, a potential challenge might focus on COMPAS's lack of overall methodological reliability or fundamental unfairness.[206]

As a final example, New Jersey's state constitution may also provide opportunities to challenge risk-assessment technology. Like Florida, New Jersey employs risk-assessment tools for pretrial purposes[207] and has taken a strikingly independent approach to constitutional interpretation.[208] New Jersey uses PSA,[209] which has a transparent and relatively straightforward

---

[198] *Id.* at 586–87.

[199] *Id.* at 585.

[200] *See id.* at 590–91.

[201] *Dubose*, 699 N.W.2d at 593–94, 596–97.

[202] *Id.* at 591–92.

[203] *See* State v. Luedtke, 863 N.W.2d 592, 606–08 (Wis. 2015).

[204] *Id.* at 606.

[205] *See* State v. Loomis, 881 N.W.2d 749, 760–67 (Wis. 2016).

[206] *Cf.* Walsh, *supra* note 178, at 1437–38 ("The great unreliability of eyewitness identifications, in addition to their great influence on a criminal proceeding, suggest that a defendant's right to a fundamentally fair proceeding is violated by the admission of unreliable eyewitness testimony at trial.").

[207] *See* Lapowsky, *supra* note 16; *see also* SHALOM ET AL., *supra* note 17.

[208] For examples of the New Jersey Supreme Court interpreting similar provisions of the state constitution more expansively than the federal constitution, see, e.g., State v. Alston, 440 A.2d 1311, 1319–21 (N.J. 1981) (standing to challenge search and seizures); *In re* Grady, 426 A.2d 467, 474 (N.J. 1981) (the right to sterilization); State v. Schmid, 423 A.2d 615, 627 (N.J. 1980) (free speech protected in some instances against private interference); State v. Baker, 405 A.2d 368, 374–76 (N.J. 1979) (the right of unrelated persons to live as a single unit); State v. Johnson, 346 A.2d 66, 68 (N.J. 1975) (consent to search).

[209] SHALOM ET AL., *supra* note 17, at 7.

methodology,[210] thereby not implicating many of the due process concerns associated with other tools (such as COMPAS), while still providing space to bring challenges on equal protection grounds.[211] Such challenges could be particularly effective in the state because it has adopted a more rigorous version of the *Batson* model for jury selection.[212] Under the New Jersey constitution, the state's facially valid race-neutral explanation for a peremptory strike is insufficient to defeat a defendant's prima facie showing of discrimination, unless that explanation also includes a "genuine and reasonable ground for believing that a prospective juror might have an individual or personal bias that would make excusing him or her rational and desirable."[213]

Given that New Jersey has been willing to impose these more rigorous requirements on the state in defendants' equal protection challenges to the jury selection process, it is conceivable that similar challenges to biased risk-assessment tools may have more success under the New Jersey constitution than under the federal Constitution. In the context of pretrial risk assessment, an analogously rigorous test could mandate that the state provide a race-neutral explanation that includes a genuine and reasonable individualized ground for believing that the specific individual defendant poses a flight or safety risk. While this burden may not be all that much more difficult to satisfy in practice—PSA questions tend to directly capture these considerations[214]—it would at least give the defendant a greater opportunity to discuss personal circumstances as a means of challenging pretrial detention or potentially onerous conditions of release.

B. Where to Bring Challenges to Risk-Assessment Technology

The viability of challenges to risk-assessment technology based on state constitutional laws will depend on a number of state-specific factors. Two especially important factors are the degree to which a state permits interpretations of provisions of its own constitution that diverge from the interpretation of parallel federal constitutional provisions, and whether the risk-assessment tool favored in that state utilizes transparent or proprietary algorithms.

With regard to the first factor, the most common approach is the "lockstep" approach, whereby states generally attempt to construe their constitu-

---

[210] *See* LAURA & JOHN ARNOLD FOUND., *supra* note 34.

[211] *See* Madeleine Carlisle, *The Bail-Reform Tool That Activists Want Abolished*, ATLANTIC (Sept. 21, 2018), https://www.theatlantic.com/politics/archive/2018/09/the-bail-reform-tool-that-activists-want-abolished/570913/, *archived at* https://perma.cc/ZNY4-AJTJ.

[212] *See* James H. Coleman, Jr., *The Evolution of Race in the Jury Selection Process,* 48 RUTGERS L. REV. 1105, 1132–33 (1996).

[213] State v. Chevalier, 774 A.2d 597, 602 (N.J. Super. Ct. App. Div. 2001) (quoting State v. McDougald, 577 A.2d 419, 435 (N.J. 1990)).

[214] *See* Carlisle, *supra* note 211.

tions in conformity with the federal Constitution.[215] In these jurisdictions, state constitutional claims are unlikely to be independently viable, as they will fall or rise based on the state's application of analogous federal law. In practice, however, even so-called lockstep states do occasionally interpret their constitutions in ways that diverge from federal constitutional doctrine. For example, Illinois has adopted a "limited lockstep approach" whereby it recognizes the federal Constitution as dominant, but chooses to identify or bolster additional rights if "a specific criterion—for example, unique state history or state experience—justifies departure from federal precedent."[216] Other states reject the lockstep approach, viewing their own constitutions either as primary shields of civil liberties, as secondary protections if the federal Constitution does not apply, or as coequal and independent sources to complement federal constitutional rights.[217] Some of these states are relatively cautious in departing from federal constitutional doctrine, while others diverge more frequently.[218]

The accessibility and transparency of the algorithm used in a state's risk-assessment program are also relevant in determining the viability of due process and similar claims in that state. This interaction is illustrated in Table C.

---

[215] *See* Joseph Blocher, *Reverse Incorporation of State Constitutional Law*, 84 S. CAL. L. REV. 323, 339 (2011).

[216] People v. Caballes, 851 N.E.2d 26, 42–43 (Ill. 2006) (quoting Lawrence Friedman, *The Constitutional Value of Dialogue and the New Judicial Federalism*, 28 HASTINGS CONST. L.Q. 93, 104 (2000)).

[217] Respectively, these approaches are commonly referred to as the primacy, interstitial, and dual sovereignty models of state constitutional analysis. Robert F. Utter, *Swimming in the Jaws of the Crocodile: State Court Comment on Federal Constitutional Issues when Disposing of Cases on State Constitutional Grounds*, 63 TEX. L. REV. 1025, 1027–30 (1985).

[218] *See, e.g.*, Shane R. Heskin, *Florida's State Constitutional Adjudication: A Significant Shift As Three New Members Take Seats on the State's Highest Court?*, 62 ALB. L. REV. 1547, 1583 (1999) ("Some state courts seem less comfortable in reaching a different result than the Supreme Court, and appear almost apologetic for reaching a different conclusion. Florida's high court makes no apologies." (footnote omitted)).

Table C: Likely viability of constitutional challenges to algorithms under different states' approaches to interpreting their own constitutions.

<table>
<tr><th>Type of Algorithm Used</th><th>Lockstep States</th><th>Cautious States</th><th>Independent States</th></tr>
<tr><td rowspan="2">Transparent (e.g., PSA)</td><td rowspan="4">Likely no separately viable claims.
Example: Wyoming[219]</td><td colspan="2">Separately viable claims must focus on the discriminatory aspects of the algorithm (i.e. equal protection challenges).</td></tr>
<tr><td>Example: Missouri[220]</td><td>Example: New Jersey[221]</td></tr>
<tr><td rowspan="2">Opaque (e.g., COMPAS)</td><td colspan="2">Separately viable claims may focus on either the discriminatory aspects of the algorithm (i.e. equal protection claims), or the lack of procedural safeguards to challenge the accuracy of the algorithm due to its secrecy (i.e. due process claims).</td></tr>
<tr><td>Example: Wisconsin[222]</td><td>Example: Florida[223]</td></tr>
</table>

The manner in which state constitutions afford procedural and substantive protections additional to federal protections for criminal defendants varies. The viability of state-level challenges to the use of risk-assessment technology will depend on state-specific factors, including the type of risk-assessment instrument used and the state's approach to constitutional inter-

[219] *See* Saldana v. State, 846 P.2d 604, 624 (Wyo. 1993) (Urbigkit, J., dissenting).

[220] *See* State v. Parker, 836 S.W.2d 930, 942 (Mo. 1992) (Price, J., concurring) ("Interestingly, the Missouri Constitution *may* require greater protection of the right of an individual to serve on a petit jury than does the United States Constitution." (emphasis added)).

[221] *See, e.g.*, State v. Schmid, 423 A.2d 615, 628 (N.J. 1980) (holding New Jersey constitution sometimes protects free speech against private interference, unlike federal constitution); State v. Chevalier, 774 A.2d 597, 602 (N.J. Super. Ct. App. Div. 2001) (explaining New Jersey's heightened requirement for showing that a peremptory challenge was made for race-neutral reasons upon a contrary prima facie showing to the contrary).

[222] *Compare* State v. Dubose, 699 N.W.2d 582, 596–97 (Wis. 2005) (holding Wisconsin constitution contains broader due process rights than the federal Constitution with respect to criminal identification procedures), *with* State v. Luedtke, 863 N.W.2d 592, 606–07 (Wis. 2015) (explaining that the "broader right" recognized in *Dubose* is "restricted . . . to the specific identification procedure known as a 'showup'" and that *Dubose* "did not create a precedential sea change with respect to the recognition of a broader due process protection under the Wisconsin Constitution than under the United States Constitution").

[223] *See* Traylor v. State, 596 So. 2d 957, 961–62 (Fla.1992); *see also* Heskin, *supra* note 218, at 1583–84 ("The Florida Supreme Court recognizes its right and duty to construe its constitution independently of the U.S. Supreme Court's interpretation of federal law . . . Florida's fundamental law and constitutional adjudication [is] increasingly independent of U.S. Supreme Court case law.").

pretation. As distinct legal regimes, state-level jurisdictions allow practitioners additional opportunities to invite judicial scrutiny of biased and opaque risk-assessment technology.

VI. Moving Forward

There is no doubt that risk-assessment technology will play a role in the criminal justice system going forward. But many risk-assessment tools in use today raise serious constitutional concerns. With fundamental liberties at stake, properly balancing the value of the technology with the constitutional questions is an urgent and critical matter for the criminal justice system. So, what *should* these algorithms look like? We believe that there are three benchmarks whereby we can analyze these tools to ensure that they operate within the bounds of our constitutional system: accuracy, simplicity, and fairness.

First, these tools should make accurate predictions. Eliminating input features may not necessarily serve defendants well, as some studies suggest.[224] Instead, incorporating additional features related to recidivism, while accounting for known racial disparities, would improve the accuracy of predictive outputs and allow the scores also to achieve greater distributional fairness.

There is a tension between accuracy and fairness as it relates to notice. Theoretically, a subject's score today should remain the same tomorrow, assuming no changes in the individual's circumstances. Adaptive models that evolve by incorporating additional data external to the defendant are thus concerning, as risk scores may change, not because a defendant's characteristics change, but because something about the algorithm changed as it processed more data. This poses a notice problem, in that a defendant cannot know *how* the defendant will be classified if the model is continuously adapting. On the other hand, we want the algorithm to improve to compensate for initial inaccuracies. One solution for this is to mandate reassessments of risk scores whenever the algorithmic model adapts to achieve a desired level of accuracy. An additional solution might be to impose transparency requirements, including disclosure of major adaptations or changes to the algorithm.

Second, risk-assessment technology should strive for simplicity, not necessarily at a programmatic level, but in terms of explainability, accessibility, and functionality. A defendant and counsel should receive an understandable explanation, in the simplest terms, of how the defendant's risk score was calculated and which factors contributed significantly to the score. For instance, the defendant should have notice that the defendant's level of

---

[224] *See* Jon Kleinberg & Sendhil Mullainathan, *Simplicity Creates Inequity: Implications for Fairness, Stereotypes, and Interpretability* 39, arXiv (June 2, 2019), https://arxiv.org/abs/1809.04578, *archived at* https://perma.cc/BB6K-EM7P.

education contributed to the score being higher or lower and the degree to which that affected the outcome. A useful analogue would be a credit report, which often enumerates in explicit detail which factors most impact one's score—for example, age of credit, credit card utilization, debt-to-income ratio. As for risk-assessment technology, this could look as simple as a report that reads: *"The following two factors most greatly contributed to your risk-assessment score: (1) your number of prior offenses; (2) your age being under 25. The following factors may help you decrease your risk-assessment score: (1) finding stable employment; (2) passing a drug test."* Alternatively, the report may include counterfactual explanations, such as: *"You were labeled high risk because you have been arrested twice within the past five years. If you had only been arrested once in the past five years, you would have been labeled medium risk."*[225]

Finally, the models must be operationally fair. An initial step would be to audit current models to see if various fairness criteria are met and to assess how protected characteristics such as race are incorporated or weighed. Critically, courts should insist on using only those risk-assessment tools that operate transparently. Many such models exist, including PSA. A further step would be to incorporate causal methodologies that account for the effects of being in a particular protected class (for example, counterfactual fairness).[226] For example, the model might be programmed to consider the impact of race on wealth, and then account for this difference before generating a prediction. By adjusting for the confluence of race and wealth, the remaining features might be more reflective of individual characteristics. These approaches could help mitigate potential biases in predictive models and provide an improved understanding of what patterns lead to the risks on which criminal justice decisionmaking purports to rely.

---

[225] *Cf.* Sandra Wachter et al., *Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR*, 31 HARV. J.L. & TECH. 841, 844 (2018) ("You were denied a loan because your annual income was £30,000. If your income had been £45,000, you would have been offered a loan."). Such explanations are intended to communicate "the smallest change to the world that can be made to achieve a desirable outcome." *Id.* at 845.

[226] *See* Doaa Abu-Elyounes, *Contextual Fairness: A Legal and Policy Analysis of Algorithmic Fairness*, 2020 U. ILL. J.L. TECH. & POL'Y 1, 31 (2020). For a formal overview of counterfactual fairness and examples of its possible applications, see generally Matt J. Kusner et al., *Counterfactual Fairness*, ARXIV (Mar. 8, 2018), https://arxiv.org/pdf/1703.06856.pdf, *archived at* https://perma.cc/ZJ5X-NH5R.