

Mechanical Jurisprudence and Domain Distortion: How Predictive Algorithms Warp the Law

Dasha Pruss

History and Philosophy of Science, University of Pittsburgh
PSA 2021 Proceedings Preprint

Abstract: The value-ladenness of computer algorithms is typically framed around issues of epistemic risk. In this paper, I examine a deeper sense of value-ladenness: algorithmic methods are not only themselves value-laden, but also introduce value into how we reason about their domain of application. I call this *domain distortion*. In particular, using insights from jurisprudence, I show that the use of recidivism risk assessment algorithms (1) presupposes legal formalism and (2) blurs the distinction between liability assessment and sentencing, which distorts how the domain of criminal punishment is conceived and provides a distinctive avenue for values to enter the legal process.

1 Introduction

In the discourse on evidence-based sentencing, a movement which advocates grounding sentencing decisions in scientific and empirical methods, recidivism risk assessment algorithms have taken on central importance (Monahan and Skeem, 2016). Proponents of recidivism risk assessment algorithms, which estimate an individual’s risk of rearrest for a future crime, offer a ‘progressive argument’ for their adoption: using risk assessment algorithms to inform sentences could reduce judge bias in decision-making and direct resources toward high-risk offenders. Evidence-based sentencing promotes such algorithms

I am grateful to the Pitt HPS and Philosophy graduate student community for their feedback on a work-in-progress presentation of this paper, and especially grateful to Marina DiMarco and Jack Samuel for their helpful written comments. Thanks also to David Danks for his invaluable comments on a later draft of the paper. Finally, I am deeply grateful to Jessie Allen for planting the seed for the ideas in this paper, for her incisive comments, and for our many stimulating discussions of jurisprudence.

as “rational, objective, and empirically sound technology for improving decisionmaking” (Hannah-Moffat, 2013, 271) while the developers of the tools claim that “objective statistical assessments are, in fact, superior to human judgment” (NorthPointe, 2015, 15).

The objectivity associated with computer algorithms is subject to familiar critiques of the value-free ideal in science, the idea that scientific reasoning should strive to be free of non-epistemic values (Douglas, 2009). Much like other scientific methods, algorithmic decision-making contends with non-epistemic values introduced by dealing with epistemic risk.¹ Moreover, there is now overwhelming evidence that algorithms can perpetuate and exacerbate the biases that plague human judgment – harmful social values can get ‘baked in’ (Danks and London, 2017).

In the context of risk assessment, critics stress that the algorithms are racially biased (Angwin et al., 2016; Harcourt, 2010), unreliable (Dressel and Farid, 2018) and that their use “amounts to overt discrimination based on demographics and socioeconomic status” (Starr, 2014, 806). Indeed, following one particularly high-profile audit (Angwin et al., 2016), recidivism risk assessment algorithms have become the poster-child for ethically problematic algorithms in the rapidly growing fairness-aware machine learning (Fair ML) literature.²

To date, most of the concern about the value-ladenness of risk assessment algorithms

¹I follow Biddle and Kukla (2017) in using the term ‘epistemic risk’ to refer to the risk of error at any stage of knowledge production, including inductive risk.

²As of January 2021, the Angwin et al. article has been cited at least 1200 times and is referenced in the introduction of virtually every Fairness, Accountability and Transparency (formerly FAT*, now called FAccT) paper about algorithmic fairness.

has centered around ‘algorithmic fairness’ and the right way to measure and prevent algorithmic bias. This focus tacitly assumes the following conditional: if risk assessment algorithms can be made free from values, they should be adopted in criminal sentencing. In other words, as long as algorithms come as close as possible to satisfying the value-free ideal, their use is preferable to biased judgment. Among other problems, this perspective neglects two problematic jurisprudential commitments of risk assessment algorithms, which illustrate an unrecognized avenue by which algorithms can be value-laden: by influencing the concepts, assumptions, and normative aims that are taken for granted in its context of application. I call this phenomenon *domain distortion*.

First, insofar as risk assessment algorithms are intended to remove judge discretion and produce consistent sentencing results, their application presupposes a formalist interpretation of legal principles – namely, that laws have one correct, mechanically discoverable meaning. Formalism, sometimes disparagingly referred to as ‘mechanical jurisprudence’, sustained heavy criticism from 20th century legal realists; it is rejected by many contemporary legal scholars for failing to capture, descriptively, what judges actually do and, normatively, what judges ought to do. It is, in essence, the value-free ideal of the legal world. Risk assessment algorithms distort the domain of criminal sentencing by reifying a widely disparaged jurisprudential presupposition and neglecting the essential interpretive component of judging. In practice, risk assessments are selectively considered by judges to *augment* judgment, sometimes amplifying existing racial biases in human judgment.³

Second, the use of risk assessment algorithms blurs the line between the domain of liability assessment (choosing a verdict) and the domain of sentencing (given a verdict,

³See, e.g., Stevenson (2018).

choosing a punishment). Jurisprudence – the philosophy of law – has traditionally been concerned with the former domain, while the latter is up to the personal discretion of judges. Risk assessment algorithms explicitly take future liability assessments into consideration when deciding sentences for current liability assessments, which I argue effectively dissolves the separation between these domains. One consequence of this blurring of domains concerns the implicit purpose of criminal sentences: deciding criminal sentences based on predictive features that have nothing to do with prior criminal conduct, such as demographic and socioeconomic information, presupposes that the purpose of punishment is consequentialist (crime control) rather than deontological (retribution).⁴ My aim here is not to advocate for either of these positions, but rather to point out that, in blurring the domains of liability assessment and sentencing, the use of risk assessment algorithms in sentencing means an implicit normative commitment to a consequentialist view of sentencing.

I begin with some brief background on risk assessment algorithms. For the bulk of the paper I defend, in turn, the claims that the use of risk assessment algorithms in sentencing (1) presupposes formalist reasoning and (2) blurs the line between liability assessment and sentencing. These are both routes by which algorithmic decision-making distorts how we reason about their domain of application, introducing value in a deeper sense than mere epistemic risk.

⁴Monahan and Skeem (2016) have also pointed out this issue.

2 Risk Assessment Algorithms

The racial disparities in the US criminal justice system are deeply troubling and well-documented. Blacks are often given harsher, longer sentences than whites for the same crimes, and this disparity has grown worse over time (Lopez, 2017). The US also incarcerates more people and at higher rates than any other country, and disproportionately incarcerates blacks (Western and Wildeman, 2009). Risk assessment algorithms are often presented as a progressive reform – a way to abolish cash bail, reduce mass incarceration, reduce bias in judgment and sentencing, and make sentencing “smart” and “evidence-based” (Estelle and Phillips, 2018; Starr, 2014).

Like actuarial algorithms, risk assessment algorithms assign risk scores to individuals based on features (e.g., age, gender, criminal history) which correlate with a certain probability of an outcome (e.g., rearrest within two years) in population samples. For instance, if a person shares characteristics with a group of individuals, 60/100 of whom were found to reoffend, then a risk assessment algorithm could predict that an individual has a 60% risk of recidivism. Decisions about individuals can then be made based on a numerical threshold – individuals classified as ‘high risk’ for recidivism may get longer prison sentences than ‘low risk’ individuals.⁵

COMPAS is one of the commonly used risk assessment algorithms in US state criminal courts. By comparing 137 factors, like answers to a questionnaire and defendant demographics (excluding information about race), to those of previous offenders, COMPAS calculates a recidivism risk score between 1 and 10 (NorthPointe, 2015). This score is

⁵Risk assessment also informs other penal decisions, such as parole, bail amount, and resource allocation. The focus in this paper is specifically on sentencing.

included in a defendant’s presentence investigation report, which is presented to a judge at the time of sentencing (Forward, 2017). Some courts are beginning to use machine learning algorithms, such as random forests, that serve a similar function to actuarial risk assessment algorithms like COMPAS (Berk, 2017).

The value-ladenness of recidivism risk assessment algorithms is now standard fare in the Fair ML literature.⁶ In 2016, journalists at ProPublica showed that COMPAS tends to make different types of classification errors for blacks and whites – blacks are more likely to be falsely classified by COMPAS as ‘high risk’ for recidivism, while whites are more likely to be falsely classified as ‘low risk’. Equivant (formerly Northpointe), the company that makes COMPAS, responded to ProPublica’s accusation (that blacks are likely to be wrongly classified as future criminals) by arguing that because COMPAS makes equally accurate predictions for both groups (whites and blacks with the same score reoffend at similar rates) the algorithm is therefore *not* racially biased (Dieterich et al., 2016). ProPublica, in turn, rebutted this rebuttal, arguing that from the perspective of someone who is part of the group more likely to be wrongly classified, simply sorting blacks and whites correctly at the same rate is not enough to make the algorithm unbiased (Angwin and Larson, 2016).

This dispute captured the imagination of the Fair ML community, which over the past three years has churned out a buffet of competing formal definitions of fairness.⁷ Tabling

⁶The focus in Fair ML is typically on value-ladenness in the sense of algorithmic bias, rather than inductive risk (the relative social costs of falsely imprisoning versus falsely releasing a defendant).

⁷See Corbett-Davies and Goel (2018) for a review.

the problems with formalizing fairness, the working assumption in these efforts matches the evidence-based sentencing movement: so long as risk assessment algorithms are free from harmful values, they should be adopted in criminal courts to reduce judge bias. A closer look at two jurisprudential problems not only calls this assumption into doubt, but also shows that the values introduced by risk assessment algorithms run deeper than mere biased predictions and epistemic risk.

3 What Is It That Judges Do?

A longstanding debate within jurisprudence concerns what it is that judges do when they interpret laws or deliver judicial decisions. Legal formalism is the view that laws are rules derived from the linguistic meaning of legal texts, and as such have a determinate, discoverable meaning that is applicable to facts (Solum, 2005). With respect to judicial reasoning, formalism holds that judges should (and do) decide cases based on this linguistic meaning of ‘black letter law’ and consistent with earlier precedent. As such, formalism implies that there is one correct way to decide cases. This adherence to rules thus restricts discretion in legal decision-making (Schauer, 1988).

Once a mainstream legal philosophy, formalism met heavy criticism from early 20th century scholars from a jurisprudential school of thought known as legal realism. In contrast to formalists, legal realists hold that jurisprudential reasoning does – and should – depend on factors outside of the strict textual meaning of a law.⁸ Law, legal realists argue, is found not in the meaning of legal statute and precedent, but rather in the behavior

⁸Note that legal realism, as the term is used in jurisprudence, has the opposite connotation of scientific realism.

of judges and legal actors – “law in action,” rather than “law in the books” (Kruse, 2011; Pound, 1910). Legal realism is thus a negative claim about formalism: single, objective interpretations of legal rules are impossible, undesirable, or fail to capture what judges really do in practice.

The realist critique take many forms. One modest realist argument is that, even if legal formalist reasoning is in principle possible, it is nevertheless undesirable. For one, laws tend to outlive the worlds of their creators, and mechanically applying laws in our current context can have unanticipated harmful consequences contrary to the drafters’ intentions. Hence, formalism is disparagingly referred to by its critics as “mechanical jurisprudence.”⁹

Other realist critiques question the very coherence of formalism. Singer, for instance, argues that legal rules often lack the certainty demanded by formalism, and further that there are different (and sometimes contradictory) ways of reading legal precedents (Singer, 1988). Similarly, Llewellyn argues that there are always multiple “correct” ways to interpret cases. A case’s interpretation depends in part on context and the “sense of the situation” of the court – in other words, an element of ineffable judicial expertise is a part of law itself (Llewellyn, 1950, 397). Other realists, like Cohen, go farther and question the coherence of legal concepts, like ‘corporation’ or ‘person’. These concepts, Cohen writes, depend on the very questions they are used to ask, like ‘is entity x subject to suit’; they are thus viciously circular and empty, an illusion covering up the true social forces that drive judicial decisions (Cohen, 1944, 816).

Even proponents of legal realism, however, tend to agree that certain factors ought not influence judges’ determination of guilt, such as a criminal defendant’s race, socioeconomic

⁹Pound, 1908 first coined this term.

background and the like. Nevertheless, jurisprudential decisions seem, in practice, to be influenced by such factors. Recent empirical studies on judges, though such studies are still quite rare, consistently lend support to legal realism as a descriptive thesis – judges’ decisions are influenced not only by political leanings of judges and social climate, but also by factors like defendant characteristics (Rachlinski and Wistrich, 2017). In one such study, Spamann and Klöhn presented four fictitious scenarios to US federal judges; in each case, caselaw either strongly or weakly supported the defendant, and the defendant was described as having either favorable or unfavorable personal characteristics. These legally irrelevant defendant characteristics were stronger predictors of the judgment outcome than caselaw, even though the judges’ written reasons appealed exclusively to legal principles for their decision (Spamann and Klöhn, 2016).

In sum, legal realists hold that jurisprudential reasoning necessarily depends on factors not contained in the text of the law, such as public good, popular sentiment, political climate, and the like – that there is an ineliminable human component to jurisprudence.

4 Mechanical Jurisprudence, Realized

The dialectic about the merits and value-ladenness of risk assessment algorithms shares a structural similarity with debates about legal formalism and realism.¹⁰ A standard for-

¹⁰Green and Viljoen (2020) recently analogized algorithmic reasoning to legal formalism to critique of the former. They argue for “algorithmic realism,” a call to recognize “the internal limits of algorithms and to the social concerns that fall beyond the bounds of algorithmic formalism” (1). By contrast, I am arguing that using algorithms for legal decision-making necessarily casts legal interpretation as a formalist enterprise.

malist response to realist critiques of biased judges is that, even if judges are not formalists in practice – that is, they do not make decisions based strictly on legal rules – they still *should* be making decisions as formalists. Legal rules may not be unbiased, but following them to the letter, warts and all, is still more justified than idiosyncratic judgment. After all, if legal reasoning is not constrained in the formalist sense, then it is unclear what distinguishes it from mere politics and opinion. Realist claims about the untenability of formalism does not justify its absence; at best, realism calls for greater transparency about the real nature of decisions, without providing grounds for their justification. Similarly, we might think that algorithmic decision-making in sentencing, even if it has its own sources of bias, is still preferable to idiosyncratic bias that pervades human decision-making.

Legal scholars like Ronald Dworkin have offered some middle-of-the-road responses to this issue from the perspective of jurisprudence. On Dworkin’s account, legal principles do constrain judges, but not in the formalist sense – decisions cannot be mechanically derived from laws because there is an ineliminable interpretive component to jurisprudence. What judges do, on Dworkin’s law-as-interpretation account, is a combination of finding and making law: much like literary interpreters, judges interpret the law to make it the best it can be while remaining consistent with what has come before (Dworkin, 1986). In particular, judges should interpret law in such a way as to maximize certain desirable features of a legal system, including justice, fairness, and due process, as well as the system’s ‘integrity’ (in essence, its moral coherence). This, Dworkin argues, not only descriptively captures what judges claim to be doing, but also provides satisfactory *grounds* for law, i.e., justification for the use of force to enforce laws.

We need not agree with every aspect of Dworkin’s story to derive a broader moral from it: the dichotomy between exclusively mechanical and idiosyncratic decisions is a

false one. Law is a human enterprise and requires dynamic interpretation, but judgment is nevertheless undergirded by legal principles.

Risk assessment algorithms, however, are not dynamic or interpretive in this way; they provide the same recommendation given the same demographic information, precluding the possibility to reinterpret legal rules as the world changes and a defendant's context shifts. The presumption that it is possible to generate correct mechanical recommendations from legal principles and the facts of a case is formalist, and must contend with the critical reasons realists have given against legal formalism. This means that the use of risk assessment algorithms comes with a normative presumption about jurisprudence, even if the algorithms could be made value-free in a superficial sense.

The extent to which risk assessment algorithms instantiate formalist reasoning in practice depends on an empirical question, namely, how much the judge's ultimate decision is influenced by the risk score. This question – whether risk assessment algorithms effectively automate judgment – was at the core of *State v. Loomis*, a 2016 Wisconsin supreme court dismissal of an appeal against the use of COMPAS in sentencing decisions. Loomis, a man who received a high risk score and a correspondingly harsh sentence, appealed on the basis that his due process was violated by the use of COMPAS, since the algorithm is proprietary and the details of its function are not up for dispute (*State v. Loomis*, 2016). The court ruled that because the output of such algorithms is merely supplementary information and is not the sole basis for a judge's decision, their use does not violate due process. The judge who sentenced Loomis even insisted that he “would have imposed the same sentence regardless of whether it considered the COMPAS risk scores” (Forward, 2017).

Here it is worth considering the prevalence of cognitive biases in human reasoning.

Relevantly, automation bias refers the human tendency to assign higher levels of authority and trust to automated sources relative to non-automated sources, like other people (Park, 2019). Related is the issue of complacency, which refers to the tendency to rely uncritically on automated systems that require human oversight – people become complacent when an automated system appears to be performing its job well (Parasuraman and Manzey, 2010). Complacency is sometimes blamed for easily preventable accidents involving machines and human operators, such as recent deaths of drivers of semi-automatic Tesla cars (Boudette, 2016) or accidents involving airplane pilots relying uncritically on faulty data outputs from cockpit machinery (Parasuraman and Manzey, 2010). Considering that the US criminal justice system is overloaded and decision fatigue among judges appears to be a pervasive problem,¹¹ automation bias plausibly jeopardizes the legitimate use of sentencing algorithms assumed by the Wisconsin supreme court. Empirical evidence is still limited, but early studies on recidivism risk assessment algorithms in Kentucky showed that judges are more likely to override a low risk assessment in favor of harsher bond conditions for black defendants than for white defendants, suggesting that the real story is more complicated (and more troubling) than simple automation bias (Stevenson, 2018; Albright, 2019).

In short, the use of risk assessment algorithms distorts the domain of criminal sentencing because it requires a problematic view of jurisprudence, which in turn could shape judge behavior. This demonstrates one striking way in which the use of algorithmic decision-making can introduce value to the legal process.

¹¹For one, judges’ decisions are influenced by how recently they have had a break (Danziger et al., 2011).

4.1 What’s Special About This Case?

At this point, one might object that domain distortion, even if present in this case, is not specific to risk assessment algorithms. Efforts to reduce bias and discretion in sentencing are not unique to the current move toward algorithmic decision-making – similar motivations underpinned the 1984 introduction of federal sentencing guidelines to limit “unwarranted disparity” of sentences for similar crimes, in part by establishing a system of mandatory sentencing guidelines (98th Congress, 1984). Among the changes introduced by the guidelines was a 258-box grid called the “Sentencing Table,” which through a complicated series of rules mechanically determines the severity of a sentence based on a defendant’s criminal history (Stith and Cabranes, 1998, 3). The guidelines were introduced at a moment of draconian crackdown on crime in the heyday of the drug war in the US. Today, the federal sentencing guidelines are perhaps most notorious for requiring longer sentences for the possession of crack cocaine compared to powder (Murphy, 2002), a recognized race proxy that resulted in harsher sentences for blacks for the crime of drug possession.

At first, the domain distortion introduced by risk assessment algorithms may seem different in degree, not in kind, from that of federal sentencing guidelines: both impose formalism, with poor consequences. Critics of federal sentencing guidelines even make reference to an issue similar to automation bias, pointing out that the system of rules in the federal sentencing guidelines “lends an appearance of having been constructed on the basis of science and technocratic expertise, giving it a threshold plausibility to a general public not familiar with its actual contours and operation” (Stith and Cabranes, 1998, xi).

To this I respond that, though risk assessment algorithms and federal sentencing guide-

lines share a similar goal and exacerbate racial disparities in practice, sentencing guidelines do not shift how the domain of criminal sentencing is reasoned about. This is because sentencing guidelines do not fall into the purview of jurisprudence and thus are not subject to critiques of formalism, whereas risk assessment algorithms do and are. To show why, it is necessary to introduce a second form of domain distortion due to risk assessment algorithms, namely, the shift in how liability assessment and sentencing are treated in relation to each other.

5 Blurred Lines

Traditionally, jurisprudence has considered sentencing and liability assessment (i.e., determination of guilt) as distinct enterprises, except in unusual circumstances like capital punishment cases, which can be decided by juries. The separation of these domains is reflected in courtroom practices – juries are instructed not to consider the punishment when making liability assessments; facts are held to a different standard in sentencing than in liability; and even back when federal sentencing guidelines were mandatory, judges had far more discretion about sentencing than they do about liability assessment (Ross, 2002). I argue, however, that the line between these domains is blurred by the use of risk assessment in sentencing. This is because risk assessment algorithms are *predictive* algorithms: they explicitly take future liability assessments into consideration when deciding sentences for current liability assessments. Federal sentencing guidelines, on the other hand, belong to the domain of sentencing; as such, they remain comfortably insulated from jurisprudential critiques, though they can of course be criticized on other grounds.

Presuming that sentencing and liability assessment are separate domains (or not) carries important normative baggage. When the Federal Sentencing Commission set out to

draft sentencing guidelines in 1984, it confronted what it referred to as the “philosophical problem” of determining “the purposes of criminal punishment”: is the purpose of punishment to serve retribution proportional to an offender’s culpability for a crime (“just desert”), or is it to lessen the likelihood of future crime, either by deterring others or incapacitating the defendant (“crime control”)? Rather than dealing with this difficult issue, the commission simply assumed that following the former will help with the latter (Monahan, 2006). Ultimately, it was decided that information about criminal history could be used in determining sentences, but that defendant characteristics like age or race, which have “little moral significance” (Moore, 1986, 317) cannot be used in sentencing, even if they are statistically predictive of recidivism (Monahan, 2006).

Conversely, risk assessment algorithms like COMPAS do take ‘morally insignificant’ variables – including socioeconomic information, education history, and familial relationships – into account. This, in effect, presupposes that the purpose of punishment is consequentialist (crime control) rather than deontological (retributive), and breaks down the separation between liability and sentencing. My purpose here is not to advocate for a particular position on sentencing, but to point out that the consequentialist values implicit in risk assessment algorithms distort how the domain of criminal sentencing is reasoned about when using other methods, like sentencing guidelines.

There is, however, important nuance here. Notably, even before the advent of risk assessment algorithms, judges were permitted to consider recidivism risk, historically based on clinical judgment, when deciding sentences. This suggests that the boundary between liability and sentencing may not have been particularly sharp to begin with. Risk assessment algorithms make the role of future liability assessment in current liability assessment more explicit, but how much further they dissolve the separation between these domains

in practice depends on how much judges considered recidivism in the first place, which is an empirical question.

6 Conclusion

The value-ladenness of algorithmic methods is typically discussed in the context of epistemic risk and algorithmic bias. In this paper, I examined a deeper sense of value introduced by algorithmic methods: *domain distortion*, changes in the way their domain of application is reasoned about. I illustrated how domain distortion can occur through an analysis of the use of risk assessment algorithms in criminal sentencing. Using insights from jurisprudence, I argued that risk assessment algorithms presuppose legal formalism, which distorts the domain of criminal sentencing, and blur the line between liability and sentencing, which presumes that the purpose of punishment is consequentialist. Empirical work remains to be done to assess how strong these distortion effects are in practice. This case study shows how domain distortion provides a distinctive avenue for values to enter the domain that algorithms are applied to, a value entry-point that is neglected by a focus on epistemic risk.

References

- 98th Congress (1984). Sentencing Reform Act of 1984. *H.R. 5773*.
- Albright, A. (2019). If You Give a Judge a Risk Score: Evidence from Kentucky Bail Decisions.
- Angwin, J. and J. Larson (2016, July). ProPublica Responds to Company’s Critique of Machine Bias Story. *ProPublica*. <https://www.propublica.org/article/propublica-responds-to-companys-critique-of-machine-bias-story>.
- Angwin, J., J. Larson, S. Mattu, and L. Kirchner (2016). Machine Bias: There’s software used across the country to predict future criminals. And it’s biased against blacks. *ProPublica*. <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>.
- Berk, R. (2017). An impact assessment of machine learning risk forecasts on parole board decisions and recidivism. *Journal of Experimental Criminology* 13(2), 193–216.
- Biddle, J. B. and R. Kukla (2017, June). The Geography of Epistemic Risk. In *Exploring Inductive Risk: Case Studies of Values in Science*, pp. 215–237. Oxford University Press.
- Boudette, N. E. (2016, September). Autopilot Cited in Death of Chinese Tesla Driver. *New York Times*.
- Cohen, F. S. (1944). Transcendental Nonsense and the Functional Approach. *ETC: A Review of General Semantics* 2(2), 82–115.

- Corbett-Davies, S. and S. Goel (2018). The Measure and Mismeasure of Fairness: A Critical Review of Fair Machine Learning. *arXiv:1808.00023 [cs]*. arXiv: 1808.00023.
- Danks, D. and A. J. London (2017). Algorithmic Bias in Autonomous Systems. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence, IJCAI'17*, pp. 4691–4697. AAAI Press.
- Danziger, S., J. Levav, and L. Avnaim-Pesso (2011). Extraneous factors in judicial decisions. *Proceedings of the National Academy of Sciences* 108(17), 6889–6892. Publisher: National Academy of Sciences Section: Social Sciences.
- Dieterich, W., C. Mendoza, and T. Brennan (2016). COMPAS Risk Scales: Demonstrating Accuracy Equity and Predictive Parity: Performance of the COMPAS risk scales in Broward County. *Northpointe Inc.*
- Douglas, H. E. (2009). *Science, Policy, and the Value-Free Ideal*. University of Pittsburgh Press.
- Dressel, J. and H. Farid (2018). The accuracy, fairness, and limits of predicting recidivism. *Science Advances* 4(1).
- Dworkin, R. (1986). *Law's Empire*. Harvard University Press.
- Estelle, S. M. and D. C. Phillips (2018). Smart sentencing guidelines: The effect of marginal policy changes on recidivism. *Journal of Public Economics* 164, 270–293.
- Forward, J. (2017). The Loomis Case: The Use of Proprietary Algorithms at Sentencing. *InsideTrack, State Bar of Wisconsin* 9(14).

- Green, B. and S. Viljoen (2020). Algorithmic realism: expanding the boundaries of algorithmic thought. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, FAT* '20, Barcelona, Spain, pp. 19–31. Association for Computing Machinery.
- Hannah-Moffat, K. (2013, April). Actuarial Sentencing: An “Unsettled” Proposition. *Justice Quarterly* 30(2), 270–296.
- Harcourt, B. E. (2010). Risk as a Proxy for Race. SSRN Scholarly Paper ID 1677654, Social Science Research Network, Rochester, NY.
- Kruse, K. R. (2011). Getting Real About Legal Realism, New Legal Realism and Clinical Legal Education. *New York Law School Law Review* 56, 26.
- Llewellyn, K. N. (1950). Remarks on the Theory of Appellate Decision and the Rules or Canons about How Statutes Are to Be Construed. *Vanderbilt Law Review* 3, 395.
- Lopez, G. (2017, November). Report: black men get longer sentences for the same federal crime as white men. *Vox*.
- Monahan, J. (2006). A Jurisprudence of Risk Assessment:. *Virginia Law Review* 92, 45.
- Monahan, J. and J. L. Skeem (2016). Risk Assessment in Criminal Sentencing. *Annual Review of Clinical Psychology* 12(1), 489–513.
- Moore, M. (1986). Purblind Justice: Normative Issues in the Use of Prediction in the Criminal Justice System. In *Criminal Careers and "Career Criminals"*, Volume 2. Washing, D.C.: National Academy Press.

- Murphy, D. E. (2002). Cocaine and Federal Sentencing Policy. *United States Sentencing Commission Commission*.
- NorthPointe (2015). *Practitioners Guide to COMPAS*. NorthPointe.
http://www.northpointeinc.com/files/technical_documents/FieldGuide2_081412.pdf.
- Parasuraman, R. and D. H. Manzey (2010, June). Complacency and bias in human use of automation: an attentional integration. *Human Factors* 52(3), 381–410.
- Park, A. L. (2019, February). Injustice Ex Machina: Predictive Algorithms in Criminal Sentencing. *UCLA Law Review*.
- Pound, R. (1908). *Mechanical Jurisprudence*. Columbia University Press.
- Pound, R. (1910). Law in Books and Law in Action. *American Law Review* 44, 12–36.
- Rachlinski, J. J. and A. J. Wistrich (2017). Judging the Judiciary by the Numbers: Empirical Research on Judges. SSRN Scholarly Paper ID 2979342, Social Science Research Network, Rochester, NY.
- Ross, J. E. (2002). What Makes Sentencing Facts Controversial - Four Problems Obscured by One Solution. *Villanova Law Review* 47, 25.
- Schauer, F. (1988, January). Formalism. *Yale Law Journal* 97(4).
- Singer, J. (1988, March). Legal Realism Now. *California Law Review* 76(2), 465.
- Solum, L. B. (2005). Legal Theory Lexicon: Formalism & Instrumentalism.
- Spamann, H. and L. Klöhn (2016). Justice is Less Blind, and Less Legalistic, Than We Thought: Evidence from an Experiment with Real Judges. *Journal of Legal Studies* 45.

- Starr, S. B. (2014). Evidence-based Sentencing and The Scientific Rationalization of Discrimination. *Stanford Law Review* 66, 803–872.
- State v. Loomis (2016). 881 N.W.2d 749. *Wisconsin Supreme Court*.
- Stevenson, M. T. (2018). Assessing Risk Assessment in Action. *Minnesota Law Review* 103, 303.
- Stith, K. and J. A. Cabranes (1998, October). *Fear of Judging: Sentencing Guidelines in the Federal Courts*. University of Chicago Press.
- Western, B. and C. Wildeman (2009). The Black Family and Mass Incarceration. *The ANNALS of the American Academy of Political and Social Science* 621(1), 221–242.